Fisseha Mekuria
Ethiopia Nigussie
Tesfa Tegegne (Eds.)

Communications in Computer and Information Science 1026

Information and Communication Technology for Development for Africa

Second International Conference, ICT4DA 2019
Bahir Dar, Ethiopia, May 28–30, 2019
Revised Selected Papers

Springer

Communications in Computer and Information Science 1026

More information about this series at http://www.springer.com/series/7899

Fisseha Mekuria · Ethiopia Nigussie ·
Tesfa Tegegne (Eds.)

Information and Communication Technology for Development for Africa

Second International Conference, ICT4DA 2019
Bahir Dar, Ethiopia, May 28–30, 2019
Revised Selected Papers

Springer

Editors
Fisseha Mekuria
Council for Scientific and Industrial Research
Meraka ICT Institute
Pretoria, South Africa

Tesfa Tegegne
ICT4D Research Center
Bahir Dar University
Bahir Dar, Ethiopia

Ethiopia Nigussie
Department of Future Technologies
University of Turku
Turku, Finland

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-030-26629-5 ISBN 978-3-030-26630-1 (eBook)
https://doi.org/10.1007/978-3-030-26630-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The proceedings of the second edition of the International Conference on ICT for Development for Africa (ICT4DA), which was organized by the Bahir Dar Institute of Technology in Ethiopia, is hereby presented. The proceedings of ICT4DA 2019 conference was compiled in collaboration with international experts. The conference brings together researchers, developers, and practitioners of ICT for development from around the world. The theme of this year's ICT4DA was "Enabling Sustained & Inclusive Economic Development through ICT Innovations."

The conference program consisted of oral presentations of 29 technical papers. These technical papers were accepted out of the 69 papers received during a paper submission period. The papers were accepted after a technical peer review process. The technical papers were categorized into three tracks, Track 1: Intelligent Systems and Data Science; Track 2: Wireless Communication and Emerging Networking; and Track 3: Natural Language Processing. Besides the technical paper presentations, the technical program also featured three keynote speeches. The three keynote speakers along with the title of the talks were, Keynote 1: "Internet-of-Things for Development in Africa: opportunities and challenges," Adj. Prof. Ethiopia Nigussie from University of Turku, Finland; Keynote 2: "Impact of Noise on Big Data Analytics," Prof. Bhekisipho Twala from University of Johannesburg, South Africa; and Keynote 3: "International collaboration in ICT4D research and development: opportunities, challenges and best practices for sustainability," Prof. Judy van Biljon from University of South Africa. In addition, two panel discussions under the following themes were included, Panel 1: "The role of Emerging ICT Technologies as a Solutions Revolution for the UN-SDGs," chaired by Prof. Fisseha Mekuria and Dr. N. Bekele-Thomas; and Panel 2: "Emerging ICT technologies in Support of Sustainable Bilateral & Institutional Collaboration in Africa," chaired by Dr. Ethiopia Nigussie.

The coordination of the ICT4DA 2019 conference was led by the General Chair of ICT4DA, Dr. Tesfa Tegegne, and the contribution of the Organizing Committee was instrumental for the success of the conference. It was also a great pleasure to work with such an excellent Organizing Committee, and we thank them for their hard work in organizing and supporting the conference. In particular, the Technical Program Committee (TPC) members, led by the TPC chair, Prof. Fisseha Mekuria (CSIR, South Africa), and co-chair, Adj. Prof. Ethiopia Nigussie (University of Turku), who completed the peer-review process of the technical papers and created a high-quality technical program relevant to the conference's theme. We are also grateful to all the authors who submitted their papers, contributing to the success of the ICT4DA 2019 conference. We strongly believe that the ICT4DA 2019 conference provided a good forum for all involved researchers, developers, practitioners, and public/private industry players to discuss ICT technology trends and research aspects that are relevant for socioeconomic development. We also expect that future ICT4DA conferences will be as successful and stimulating, as the conference series continues to develop and deliver relevant and innovative contributions to the global knowledge in ICT4D as presented in this volume.

April 2019

Tesfa Tegegne
Fisseha Mekuria
Ethiopia Nigussie

Organization

Organizing Committee

Proceedings Chair

Tesfa Tegegne	ICT4D Research Center, BiT, Bahir Dar University, Ethiopia

Program Chairs

Fisseha Mekuria	CSIR, South Africa
Ethiopia Nigussie	University of Turku, Finland

Program Committee

Surafel Lemma Abebe	Ababa Institute of Technology, AAU, Ethiopia
Habtamu Abie	Norwegian Computing Centre, Norway
Esubalew Alemneh	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia
Andres Arcia Moret	University of Cambridge, UK
Dereje Yohannes Ashenafi	Addis Ababa Science and Technology University, Ethiopia
Yaregal Assabie	Addis Ababa University, Ethiopia
Hailu Beshada Balcha	ICT4D Research Center, BiT, Bahir Dar University, Ethiopia
Elefelious Getachew Belay	University of Milan, Italy
Imed Ben Dhaou	University of Monadtir, Tunisia
Desta Berihu	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia
Mekuanint Agegnehu Bitew	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia
Taye Girma Debelee	Addis Ababa Science and Technology University, Ethiopia
Sigurd Eskeland	Norwegian Computing Centre, Norway
Ali Farooq	University of Turku, Finland
Gebeyehu Belay Gebremeskel	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia
Samson Gejibo	University of Bergen, Norway
Fekade Getahun	Addis Ababa University, Ethiopia
Liang Guang	Huawei Technologies, China
Antti Hakkala	University of Turku, Finland
Mesfin Kifle	Addis Ababa University, Ethiopia
Swaib Kyanda Kaawaase	Makerere University, Uganda
Khalid Latif Dentsu	Aegis Network, Finland

Adane Mamuye	University of Gondar, Ethiopia
Fisseha Mekuria	CSIR, South Africa
Fikreselam Gared Mengistu	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia
Luzango Mfupe	CSIR, South Africa
Drake Patrick Mirembe	Uganda Technology and Management University (UTAMU), Uganda
George Musumba Dedan	Kimathi University of Technology, Kenya
Edward Mutafungwa	Aalto University, Finland
Behailu Shiferaw Negash	University of Turku, Finland
Ethiopia Nigussie	University of Turku, Finland
Petri Sainio	University of Turku, Finland
Hussien Seid	Addis Ababa Science and Technology University, Ethiopia
Zelalem Sintayehu Shibeshi	University of Fort Hare, South Africa
Abiot Sinamo	Mekelle University, Ethiopia
Tesfa Tegegne	ICT4D Research Center, BiT, Bahir Dar University, Ethiopia
Woubishet Zewdu Taffese	Aalto University, Finland
Mesfin Belachew Tefera	Agricultural Transformation Agency, Ethiopia
Dereje Teferi	Addis Ababa University, Ethiopia
Nanda Kumar Thanigaivelan	University of Turku, Finland
Mthulisi Velempini	University of Limpopo, South Africa
Ruth Wario	University of the Free State, South Africa
Hlabishi Kobo	CSIR, South Africa
Kula Kekeba	Addis Ababa Science and Technology University, Ethiopia
Mebiratu Beyene	ICT4D Research Center, BiT, Bahir Dar University, Ethiopia

Web Chair

Mebiratu Beyene	ICT4D Research Center, BiT, Bahir Dar University, Ethiopia

Publicity Chairs

Abinew Ali	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia
Elias Wondemagen	Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia

Organized by

ICT4D Research center, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar

Hosted by

Bahir Dar University

Sponsoring Institutions

Contents

Wireless and Mobile Computing Track

Artificial Intelligence and Data Science Track

A Novel Method of Pre-processing Using Dental X-Ray Images by Adaptive Morpho Histo Wavelet Denoising (AMHW) Method

S. G. Saju Simon[1(✉)], X. Felix Joseph[2], Ararso Taye Waktola[2], and Derera Senay[2]

[1] Department of Biomedical Engineering, Noorul Islam University, Kumaracoil, India
sajusimon.nta@gmail.com

[2] Department of Electrical and Computer Engineering, Bule Hora University, Hagere Mariam, Ethiopia
felixjoseph75@gmail.com, arasotaye@yahoo.com, derjan06@gmail.com

Abstract. Dental X-ray imaging is one of the medical imaging techniques used by the dental practitioners. This method uses dental X-ray diagnostic file. This has been found to be very helpful to dental doctors to get more information on the oral disease in patients. X-ray radiation is used for wide range of medical imaging applications. Once the technique becomes a success, it produces good quality X-ray images. The X-ray image is very essential for the treatment planning and procedure for the patients. The X-ray image has lot of artifacts which can be removed through various preprocessing methods. In this paper a novel method for preprocessing is introduced through the fusion of adaptive histogram, morphological enhancement and wavelet de-noising (AWHM). The input X-ray image is initially checked with various other existing preprocessing methods such as adaptive histogram equalization, un sharp masking, Gaussian low pass and high pass methods, high pass adaptation, morphological enhancement, contrast enhancement and wavelet de-noising. AWHM is giving better result than all the other methods. The existing method and novel method is compared with various non-reference parameters and the reference parameter. The result of the novel method is better than all other existing methods. The Contrast Per Pixel technique is used for analyzing the pixel brightness in a more better way. Since the CPP value is higher in AMHW method it indicates AMHW method is the best among the pre-existing methods.

Keywords: Adaptive histogram equalization · Un sharp masking · Gaussian low pass and high pass methods · High pass adaptation · Morphological enhancement · Contrast enhancement and wavelet de-noising · AMHW method

1 Introduction

Dental X-ray imaging is a very useful diagnostic method in dentistry. Dental X-ray imaging is only possible on dental lesions. The dental diseases are characterized by variations in the trabecular pattern. These alterations involve depth of the trabecular

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 3–10, 2019.
https://doi.org/10.1007/978-3-030-26630-1_1

pattern, shape and the X-ray absorption amount in the lesion caused by dental tissues. The output from dental X-ray devices, provides high resolution X-ray images. It produces confused trabecular pattern structure and contrasts difference in the region of interest of dental lesions by human observers. Many researches have shown that minute dental lesion in the dental X-ray image cannot be detected by a dentist. The important applications in X-ray diagnosis in dentistry is periodontal bone loss. Periodontitis is one of the most common diseases present and also gingivitis is caused by plaque accumulation. Gingiva inflammation leads the destruction of soft and hard dental tissues. The limitations of X-ray diagnosis in dental diseases is already mentioned above. Image processing using computer is used to detect variations in contrasts; the dentist provides creativity and intelligence to make sense of the findings for the recognition process.

2 Proposed Work

In this paper a novel method for preprocessing is introduced. The existing preprocessing methods are adaptive histogram equalization filtering, unsharp masking filtering, Gaussian low pass filtering, Gaussian high pass filtering, high pass adaptation filtering, morphological enhancement filtering, contrast enhancement filtering and wavelet de-noising filtering.

Adaptive histogram equalization (AHE) is a medical image pre-processing techniques which is used to improve the pixel contrast in images. The AHE differs from normal histogram equalization which uses adaptive technique. The AHE redistributes the lightness value of the images and it is used to improve the normal contrast of the pixels and enhances the edges in each region. The adaptive step function performs Contrast Limited Adaptive Histogram Equalization (CLAHE). The enhancement method is better than regular histogram equalization for dental X-ray images.

Unsharp masking is an image sharpening preprocessing method which is found in digital image processing software such as Matlab. The name Unsharp is derived from the technique which uses a negative or blurred image to create a mask of the input image. Unsharp mask is combined with the input image and creates a new image which is less blurred than the input X-ray image. The function qspecial produces various predefined filters in the form of correlation kernels. This filter has the effect of masking edges and the fine details in the dental X-ray images.

```
A = imread('orginalimage.jpg');
b = qspecial('unsharp');
A2 = imfilter(A, b);
Imshow(A), ('Original Image')
Figure, imshow(A2), title('Filtered Image')
```

Gaussian filter is one of the preprocessing filters whose impulse function is Gaussian function. Such filters have the property of having no overshoot for step function input and minimizes the rise- fall time. They are based on frequency filter and time domain filter. Gaussian filter modifies the input X-ray images by mathematical convolution with Gaussian function and is called as weierstrass transforms. The Gaussian low pass and high pass helps minimizes the problem in the ideal low pass and high pass filter. The ringing effect which appears at the point because of some point transmission between one colour to another cannot be defined precisely (Fig. 1).

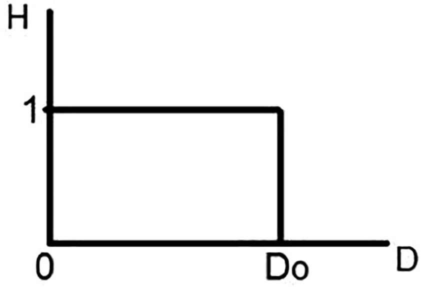


Fig. 1. Graphical representation of ideal low pass filter

The above graph represents an ideal low pass filter. The point Do does not tell whether the value would be 0 or 1. In order to reduce the effect, an ideal low and high pass filter followed by Gaussian low pass and high pass filter is introduced. The filtering concept in Gaussian low pass filter remains same but the transitions become different and smoother. The filtered output from the input dental X-ray image is represented graphically. From the smooth curve, the value of Do can be defined accurately y The Gaussian high pass filter and the ideal high pass filter are similar concepts, but the conversion is smoother in the former case (Fig. 2).

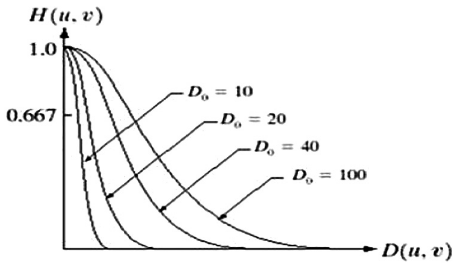


Fig. 2. Contrast enhancement denoising

Morphological enhancement is a non – linear operation in image processing. This operation is related to the shape and features of the dental X-ray image. These preprocessing operations are not based on numerical values but are based on their pixel values. This preprocessing is relative ordering of pixel values of dental X-ray images. This morphological enhancement can be used for gray scale images because the unknown light transfer function and minor interest on the proper pixel value.

The contrast enhancement is an image preprocessing techniques. It is a suitable method for X-ray dental image preprocessing. It helps to improve the details in X-ray images that are under study.

The wavelet – De-noising preprocessing method is used for removing artifacts from X-ray dental images. Artifacts can be white noise or random noise with no coherent noise produced by the algorithms or imaging machines.

3 Result and Discussion

3.1 Adaptive Morpho – Histo Wavelet Denoising (AMHW)

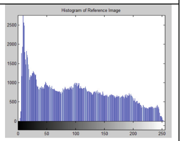
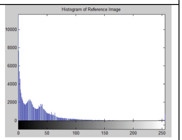
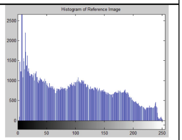
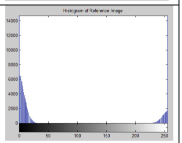
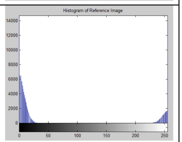
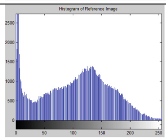
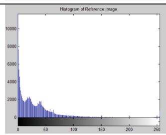
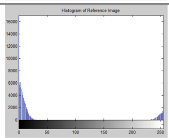
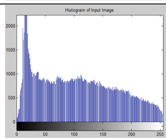
The AMHW is a novel method of preprocessing technique used for medical image processing from the fusion of adaptive morphological histogram wavelet de-noising. The algorithm of AMHW produces better result than any other method which was discussed earlier. The result is checked with various non – reference parameters such as Histogram graphical representation, Contrast per Pixel matrix, Blur Matrix and Anisotropic Quality Measure, Average Information.

The non – reference parameters provide more clarification from the end result. From the histogram graphical representation the novel method provides more information from the input medical dental X-ray image. The AMHW method provides more information than all other existing methods through the graphical representation as shown in the figures (Table 1).

Contrast Per Pixel or CPP which is used to quantify the quality of dental X-ray images. CPP is the average intensity difference between a pixel and the neighboring pixel. Contrast is the difference in color that makes them distinguishable. Contrast is determined by the difference in the brightness and color of the X-ray image. In case of blur matrix, it is the comparing of variations between the adjacent pixels before and after preprocessing with low pass filters. In the initial step, the computation of the intensity variations between adjacent pixels of the input image is to be done (Fig. 3).

Then apply a low pass filter and compute the variations between the adjacent pixels. Finally these intensity variations help to calculate the blur in the input dental X-ray image. If the change between dental X-ray image and the blurred image is more, it can

Table 1. Compression result of AMHW method and various other methods

Methods	Histogram Graphical Representation
Adaptive Histogram equalization	
Un Sharp Masking	
Gaussian Low Pass Method	
Gaussian High Pass Method	
High Pass Adaptation Method	
Morphology Enhanced Method	
Wavelet De-noising Method	
Contrast Enhanced Method	
ADAPTIVE MORPHO – HISTO WAVELET DENOISING (AMHW)	

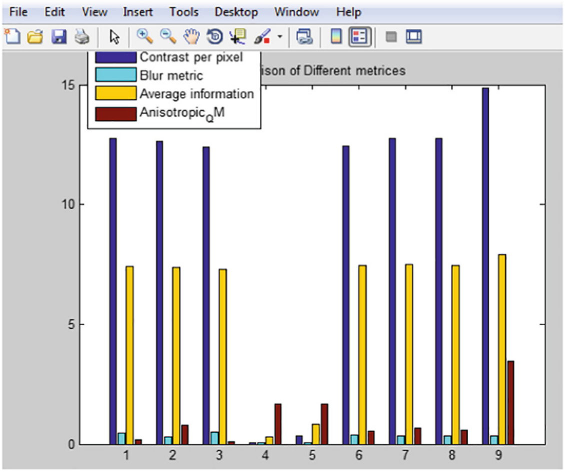


Fig. 3. Graphical representation of various matrix results

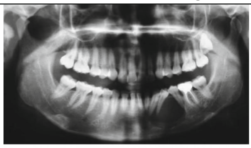
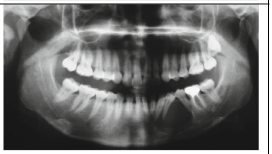
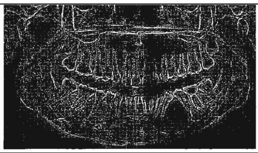
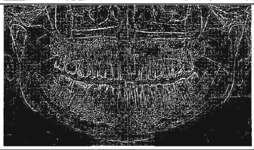
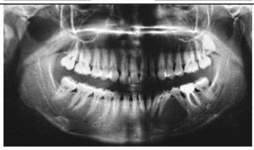
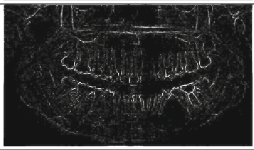
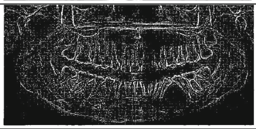
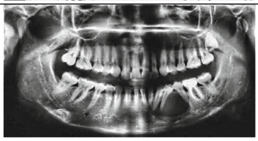
be concluded that the preprocessed image is sharp or if the variation between the blurred image and the input dental X-ray image is less, it means that the pre-processed image was already blurred. From the average information matrix, the amount of information from the X-ray image must be coded by using compression algorithm. The less entropy dental X-ray images having lot of black sky, very less contrast and large runs of pixels with the same. The perfectly flat X-ray images have entropy value zero. The entropy is calculated by the formula

$$\text{Entropy} = -\sum_{i} P_j \mathrm{Log}_2 P_j$$

From the above equation, the probability (P_i) is the difference between two neighbouring pixels and is equal to I, and $\log_2$ is the base 2 logarithms. The anisotropic ‘Q’ matrix is a no reference quality matrix. Here the input image is compared with its flipped, angle shifted version. The process is done for different angles. In each step the renyi entropy is taken between them (Table 2). The Anisotropic Quality measure is calculated using the formula

$$Q = (1/1 - \text{alpha})) * \log 2\,(\text{sum of renyi entropy})$$

Table 2. Final results of novel method (AMHW) and other existing methods

Methods	Pre Processed Image
Adaptive Histogram equalization	
Un Sharp Masking	
Gaussian Low Pass Method	
Gaussian High Pass Method	
High Pass Adaptation Method	
Morphology Enhanced Method	
Wavelet De-noising Method	
Contrast Enhanced Method	
Adaptive Morpho – Histo Wavelet Denoising (AMHW)	

4 Conclusion

Based on the above said preprocessing methods AMHW, Adaptive morpho histo wavelet de-noising makes a better quality result image after removing artifacts. The image from AMHW produces high pixel brightness than all other existing methods. Therefor the further processing of this dental image produces more information and it will be useful for analysis and diagnosis. Thus the novel AMHW method is a better method for preprocessing to get better result than any other existing method. This image which is obtained is very helpful for the further segmentation.

References

1. Yaroslavsky, L.P.: Local adaptive image restoration and enhancement with the use of DFT and DCT in a running window. In: Proceedings of SPIE, Bellingham, vol. 2825, pp. 2–13 (1996)
2. Mehdizadeh, M., Dolatyar, S.: Study of effect of adaptive histogram equalization on image quality in digital preapical image in pre apex area. Res. J. Biol. Sci. **4**(8), 922–924 (2009)
3. Regezi, J.A.: Periapical disease: spectrum and differentiating features. J. Calif. Dent. Assoc. **27**, 285–289 (1999)
4. Langland, O.E., Langlais, R.P. Preece, J.W.: Principles of Dental Imaging. Lippincott Williams & Wilkins (2002)
5. Noor, M., Khalid, N.E.A., Ali, M.H., Numpang, A.D.A.: Enhancement of soft tissue lateral neck radiograph with fish bone impaction using adaptive histogram equalization (AHE). In: The 2^{nd} International Conference on Computer Research and Development (2010)
6. Khalid, N.E.A., Manaf, M., Aziz, M.E., Ali, M.H.:CR images of metacarpal cortical edge detection-bone profile histogram approximation method. In: Intelligent and Advanced Systems, 2007, ICIAS 2007, pp. 702–708, 25–28 Nov 2007
7. Sakata, M., Ogawapp, K.: Noise reduction and contrast enhancement for small-dose x-ray images in wavelet domain. In: Presented at the Nuclear Science Symposium Conference record (NSS/MIC) (2009)
8. Aufrichtig, R., Xue, P.L.: Dose efficiency and low-contrast detectability of an amorphous silicon X-ray detector for digital radiography. Phys. Med. Biol. **45**, 2653–2669 (2000)
9. Tiwari, R.B.: Dental X-ray image enhancement based on human visual system and local Image statistic. In: Proceeding of the International Conference of Image Processing, Computer Vision and Pattern Recognition, vol. I&II, pp. 100–106 (2006)
10. Gonzales, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice Hall, Upper Saddle River (2002)
11. Thangavel, K., Manavalan, R., Aroquiaraj, I.L.: Removal of speckle noise from ultrasound medical image based on special filters: comparative study. ICGST-GVIP J. **9**(III), 25–32 (2009)
12. Pisano, E.D.: Contrast limits adaptive histogram equalization image processing to improve the detection of simulated speculations in dense mammograms. J. Digital Imag. **11**, 193–200 (1998)
13. Mahmoud, T.A., Marshall, S.: Medical image enhancement using threshold decomposition driven adaptive morphological filter. In: Presented at the 16th European Signal Processing Conference (EUSIPCO), Laussane, Swizerland (2008)
14. Foi, A.: Noise estimation and removal in MR imaging: the variance-stabilization approach. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1809–1814 (2011)
15. Grundland, M., Dodgson, N.A.: Decolorize: fast, contrast enhancing, color to gray scale on version. Pattern Recogn. **40**(11), 891–2896 (2007)

Knowledge Based System for Diagnosis and Treatment of Mango Diseases

Waltengus Birhanie[1(✉)] and Tesfa Tegegne[2]

[1] Assosa University, Assosa, Ethiopia
walteharamaya@gmail.com
[2] Bahirdar University, Bahirdar, Ethiopia

Abstract. Mango is one of the world's most important fruits of the tropical and subtropical countries and cultivated extensively as a commercial fruit crop. The mango fruit crop is also cultivated in all parts of Ethiopia. Mango trees in most parts of Ethiopia are developed from seedlings and are inferior in productivity and in fruit quality. Mango tree is attacked by different insects and diseases. With an end goal to address such issues and problems it is very essential to develop knowledge-based system that can provide advice for agriculture professionals and farmers to make possible diagnosis and treatment of mango diseases.

The aim of this study was developing knowledge based system for diagnosis and treatment of mango disease. Ambo Plant Protection Research Center and Assosa plant health clinic were selected area for data collection. Purposive sampling technique was utilized to choose domain experts for knowledge acquisition. In this way, in order to develop the proposed knowledge based system an implicit and explicit knowledge utilized and acquired through interview and document analysis respectively. The knowledge acquired through document analysis and interview is modeled using decision tree. The researchers used experimental research design to developed prototype system. The experiment was conducted using rule based technique. Finally, performance testing and user acceptance evaluation were performed. The performance according to the domain expert's evaluation is scored total of 81.8% and the user acceptance testing scored 86.29% performance. The overall performance of the system is 84.05%. The prototype system meets the objectives of the study. Updating the rules in the knowledge base of the system automatically, developing the system in local language with well designed user interface and applying other techniques are the future works of the study.

Keywords: Knowledge based system · Mango disease · Rule based system

1 Introduction

Agriculture is the corner stone of the development policy of the Government of Ethiopia. The country's economic development will depend, in large part on sustainable improvements in agriculture. Agriculture is the most critical sector in the Ethiopian economy since it straightforwardly bolsters around 85% of the population regarding business and occupation; it contributes around half of the nation's gross

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 11–23, 2019.
https://doi.org/10.1007/978-3-030-26630-1_2

domestic product (GDP) and it produces around 90% of the export earnings [1]. Agriculture is also the major source of food for the population and hence the prime contributing sector to food security [1, 2]. Ethiopia is characterized by having different agro-ecological zones and it accounts about a total area of 1.13 million km^2 [4]. A variety of fruit crops has been growing in different agro ecological Zones by small farmers, for subsistence and income generation. Mango is the most important fruit and it is the second in area coverage after banana in southern region and southwest of Ethiopia [5].

The area of mango production in the different regions is about 3789.47 ha in Oromia, 3375.89 ha in SNNPR, 652.56 ha in Benishangul Gumz, 246.85 ha in Amahara, 180.41 ha in Gambella, 44.5 ha in Dire Dawa, 33.52 ha in Somali, 118.20 in Tigray and 367.24 ha in Harari. The total area allotted for mango is about 8808.64 ha and the country annual production of mango from all mango grower regions is about 697,507 quintals [7]. Agriculture and plantation is an important and interesting research area everywhere in the world and Ethiopia is no exception. Nowadays available land area for a plantation is becoming scarce. This scarce resource is frequently wasted through our bad practices and improper management. Cultivation is a more economical but complex process. Diagnosis and treatment of mango fruit diseases for the maximum profit involves a sequence of tasks. These tasks and the whole process need a lot of expert knowledge and experience. But unfortunately, people having this type of knowledge are very limited. Their assistance is not available when the person who is going to cultivate needs it.

Knowledge based System is one of the important application-oriented branches of Artificial intelligence. The knowledge based systems approach attempts to model the domain knowledge of experts in their respective areas of specialization, for example, diagnosis, planning, forecasting etc. Expert System is based on the knowledge including not only models and data, but more emphasizing on experiences of domain experts [8].

A knowledge based expert System is an intelligent computer program that employs knowledge and inference procedures to address problems that are not easy enough to need human expertise for their solution. The knowledge necessary to perform at such level plus the inference procedures used can be thought of as a model of the expertise or the best practitioners in the field [9]. A knowledge based expert system is based on an extensive body of knowledge about a specific problem area. Knowledge is represented as data or rules that appear in syntax as physical patterns in electronic form in the knowledge base. The knowledge is represented as production rules in the form of condition-action pairs of IF—THEN format [10]. Knowledge engineers gather knowledge from domain experts and put it in such a form that system can use for inferring and reasoning using a knowledge representation technique. Marwaha [9] described as "the need of expert systems for technical information transfer in agriculture can be identified by recognizing the problems in using the traditional system for technical information transfer, and by proving that expert systems can help to overcome the problems addressed, and are feasible to be developed".

Farmers and agronomy expert require timely, accurate and location specific information in relation with different aspects of farming like the pests, diseases, weeds and fertilizer management, etc. for their crops from agricultural experts [12]. On the other hand, the complexity of a whole farming process is growing because it is constrained by many factors such as requirements, goals, regulations, etc. that farmer must satisfy or consider. To provide such information and to achieve an optimal crop plan, the automation is provided by computer-based systems termed as advisory systems. Knowledge based advisory system supports the farmers and experts in getting advices on many activities in farming process [11]. In this study, the researcher proposed knowledge based system to diagnosis and treat mango disease. It is an expert system that recommends farmers and agronomy experts about mango disease diagnosis and treatment.

2 Statement of the Problem

Fruit crops play an important role in the national food security of people around the world. They are generally delicious and highly nutritious, mainly of vitamins and minerals that can balance cereal-based diets. Fruits supply raw materials for local industries and could be sources of foreign currency. The mango fruit crop is cultivated in all parts of Ethiopia and while in southwest Ethiopia, mango is the first fruit crop grown [6]. The total area allotted for mango is about 8808.64 ha and the country annual production of mango from all mango grower regions is about 697,507 quintals [7]. But Mango tree is attacked by different insects and diseases such as, Anthracnose, Bacterial Black spot, Fruit fly, mango gall flies, Mango leaf coating, Mites, Mango seed weevil, Mealy bug, Powdery mildew, Scale, Spider mites, Mango tip borer, Stem-end rot, Termite, Thrips and White flies. Plant diseases are one of the most important reasons that lead to the destruction of plants and crops.

The major production constraints indicated by Bezu [13] were water shortage or erratic rainfall (79%) followed by disease and pest (75.7%) problems. Lack of knowledge and recommended production practices (nutrition, pruning, pest management etc.) and post-harvest losses were also noted as major problems of the mango growers. It is in agreement with CSA [3] report that stated mango production in Ethiopia fluctuates because of occurrence of diseases and lack of proper management [3]. Problems of pests and plant diseases are the main obstacles in increasing agricultural production. Damage occurs, both on the field during the cultivation process and warehouse storage. These conditions will significantly affect the income of farmers and the world's food supply. "Pests and plant diseases are still a complicated issue for owners of the plants, especially for those who do not have the basic knowledge of crop cultivation. The problem becomes more complex due to many types of pests and plant diseases. Identifying the plant diseases is not easy task; it needs experience and knowledge of plant and their diseases. Moreover, it requires accuracy in describing the symptoms of plant diseases [18]. To differentiate the cause of damage to crops, long enough experience are needed so there is no error in concluding the cause in order to take the right decision in an effort to control" [14]. In many countries today, farming has become technologically advanced and expert systems are widely used in the field of agriculture. In this way farmers can get experts opinion on their specific problems like selection of most suitable crop variety, diagnosis or identification of crop/livestock disorder, suggestion, and tactical decisions throughout production cycle from the expert system [15].

Although many applications have been developed for crop production; disease and pest management abroad, in Ethiopia there are very few works done to design knowledge base system for agricultural sector. However, to the best of the researchers' knowledge, no research has been done to apply knowledge based system for diagnosis and treatment of mango disease particularly in Ethiopia. Therefore, the researchers have been developed knowledge based system in mango fruit in order to diagnosis various disease and taking management decisions for the benefit of farmers and experts. Identifying, diagnosis and treating the mango diseases for the maximum yield production involves a sequence of tasks. These tasks and the whole process need a lot of expert knowledge and experience. But unfortunately, people having this type of knowledge are very limited. Their assistance is not available when the person who is going to cultivate needs it.

As a result, the aim of the study is to develop knowledge based system for diagnosis and treatment of mango disease. At the end this study will response the following research questions:

- What suitable domain knowledge exists in an explicit and tacit form for diagnosis and treatment of mango diseases?
- What are the appropriate trends or techniques taken by experts to diagnosis mango disease?
- How to model and represent the acquired knowledge for developing the knowledge-based systems?

3 Objective of the Study

3.1 General Objective

The general objective of this study is to develop a knowledge based system for mango diseases diagnosis and treatment using rule based reasoning approach.

3.2 Specific Objectives

- To acquire knowledge from domain experts and codified sources required for developing the system
- To model and represent the acquired knowledge using appropriate knowledge representation technique.
- To develop prototype knowledge based system for mango diseases diagnosis and treatment.
- To evaluate the overall performance of the prototype knowledge based system.

4 Scope of the Study

The scope of this study was limited to develop a prototype system for diagnosis and treatment of mango diseases and providing possible suggestions for decision making. The activity concerned in conducting this work includes literature review, problem identification, knowledge acquisition, modeling, representation and implementation or encoding. The prototype consists of knowledge base, inference engine, user interface, explanation facility and rule based reasoning mechanism. Even though the prototype includes all these components this system has limitation in automatic updating of the knowledge base by the user when the new factors are introduced.

5 Significance of the Study

The system developed enables to reduce the problem of the limited numbers of experts in giving preliminary diagnosis and treatment of Mango diseases and motivate further researches to be conducted in the area of agricultural expert system. The immediate beneficiaries of the study are primary agriculture workers and agriculture professionals or agronomist. Particularly, the prototype will have great significance to teach primary agriculture extension workers, general agronomy experts in order to have well understanding about mango diseases. As a result, those agriculture workers can use the system in diagnosing mango diseases on primary agriculture sectors where highly qualified mental professionals are unavailable. Additionally, the prototype can be used for agriculture professionals as a guide.

6 Methodology of the Study

6.1 Research Design

Dipanwita [16] used experimental research design to develop intelligent medical system for diagnosis of common disease by acquiring tacit and explicit knowledge from domain knowledge expert. The domain knowledge was acquired and then represented. The acquired and represented knowledge was inserted into the knowledge base. Based on the result of evaluation of inserted knowledge it changed again and again. In addition, during prototype development stages the sequence of the facts and rules were changes again and again until it fitted the best sequence. Moreover, the way to test and evaluate the performance of the prototype system by feeding the cases and records the result to compare it against with the decision made by domain experts in similar settings. Therefore, experimental research design and methodology is used in this study.

6.2 Sources of Data and Data Collection Methods

The researchers used primary and secondary data as source of information. In this study explicit and tacit knowledge is acquired from both codified (documented) sources and non-codified (non-documented) sources respectively. Non-codified sources of knowledge are acquired from agronomy experts (agronomists) who work in the Ambo Plant Protection Research Center and Assosa plant clinic research center by using interview and critique knowledge elicitation methods to filter the acquired knowledge. Similarly, codified sources of knowledge such as agricultural books, training manuals and journal agriculture articles are acquired by using document analysis technique.

6.3 Knowledge Modeling and Representation Method

Knowledge modeling is a cross disciplinary approach to capture and model knowledge. In this study the acquired knowledge was modeled using decision tree and represented using production rule which is one of the knowledge representation techniques. The prototype system uses backward chaining also called goal-driven chaining which begins with possible solutions or goals and tries to gather information that verifies the solution. The knowledge engineer collects Mango disease symptoms or cases and models it by using decision tree structure. For this research the knowledge representation method, rule based is chosen because it clearly demonstrates the domain knowledge. In a rule based system much of the knowledge is represented as a rule that is as conditional sentences relating statements of facts with one another. The following sample rules in the knowledge base of the prototype are expressed with natural language rules IF … THEN …

Rule 1:

If the symptom of mango diseases is Leaf spots
And there is flower, twig, and blossom blight
And Fruit rot
And "Tear staining" is that develops when spore-laden water droplets from infected twigs and panicles wash over and infects the fruit.
And the panicles (flower clusters) start to show as a small black or dark-brown spots
Then Anthracnose is diagnosis

Rule 2:

If the symptom of mango diseases is Leaf spots
And there is flower, twig, and blossom blight
And there is a profusion of rust-colored microscopic "spores" on the leaf surface
And Rust-red "spore" masses will also develop on infected stems
And Altitude is high
Then the Alga Spot (Red Rust, Green Scurf) is diagnosis

Rule 3:

If the symptom of mango diseases is Leaf spots
And there is flower, twig, and blossom blight
And Fruit rot
And "Tear staining" is that develops when spore-laden water droplets from infected twigs and panicles wash over and infects the fruit.
And the panicles (flower clusters) start to show as a small black or dark-brown spots

And the rot produces dark streaking of the water-conducting tissue
Then stem-end rot is diagnosis

Rule 4:

If the symptom of mango diseases is Leaf spots
And there is flower, twig, and blossom blight
And Bacterial leaf spot is noticed on the leaves as angular water soaked spots or lesions, surrounded by clear holes.
And longitudinal cracks also develop on the petioles.
And Cankerous lesions appear on petioles, twigs and young fruits.
Then the Bacterial Canker is diagnosis

These rules are added to the knowledge base using prolog programming language. These rules capture common evidence of problems associated with the symptoms for mango disease diagnosis. The detail is presented in the knowledge base.

6.4 KBS Implementation Tool

To develop the knowledge-based system, Prolog programming language was used. Specifically, SWI-Prolog editor has been chosen for this study. SWI-Prolog editor has debugging tool and it is flexible. It is portable to many platforms, including UNIX/Linux platforms and Windows. Additionally, it is non-commercial version of Prolog.

7 Implementation and Experimentations

This section presents the process of the knowledge based system development. The section also addresses coding and evaluation of the system. The represented knowledge is encoded and implemented using appropriate programming tool.

7.1 Architecture of the Proposed Prototype

Architecture defines how the system is constructed, describes what the critical components are and how they fit together. Architecture design is done in this section as well, which incorporates the knowledge base (facts and rules), explanation facility, inference mechanism, the knowledge base editor and the user interface. This proposed KBS has knowledge acquisition subsystem, knowledge base that contains facts and rules, backward inferencing, user interface and explanation facilities components as shown in the Fig. 1.

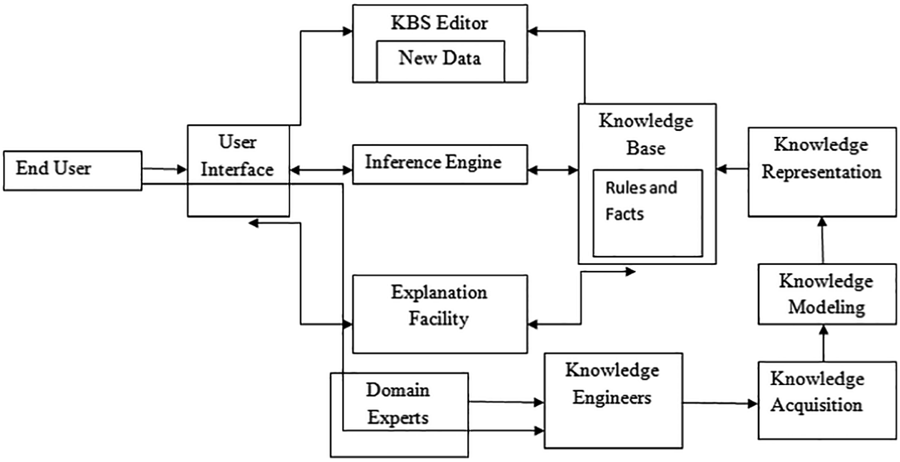


Fig. 1. Architecture of the developed prototype system

The knowledge acquisition subsystem that performs the eliciting and structuring of knowledge from the pathologist and relevant document is presented in the previous section. The core components of the developed KBS system are shown in Fig. 1. The main program that contains the inferencing, requests user to load the knowledge base and starts the identification process. The backward inferencing mechanism identifies disease through asking the characteristics features or symptom of each mango disease. The user interacts with the knowledge base system via user interface. Knowledge base is a set of rules or the encoded knowledge about diagnosis and treatment of mango disease. The validated knowledge is represented in the form of rules by rule-based representation technique and the rules are codified to the knowledge base of the prototype system using Prolog programming language. An inference mechanism consists of search and reasoning methods that enable the system to find solutions and if necessary, provide justification for its answers.

In addition, the system is capable of proving explanation for sign and symptoms of diseases of mango. But, the KBS does not learns and update its fact base when new facts were generated during reasoning.

The following Fig. 2 shows sample of how the prototype system accepts user response and provide the final advice as a solution for diagnosis and treatment. To interact with this system first the users must start the system by typing fruit followed by full stop (fruit).

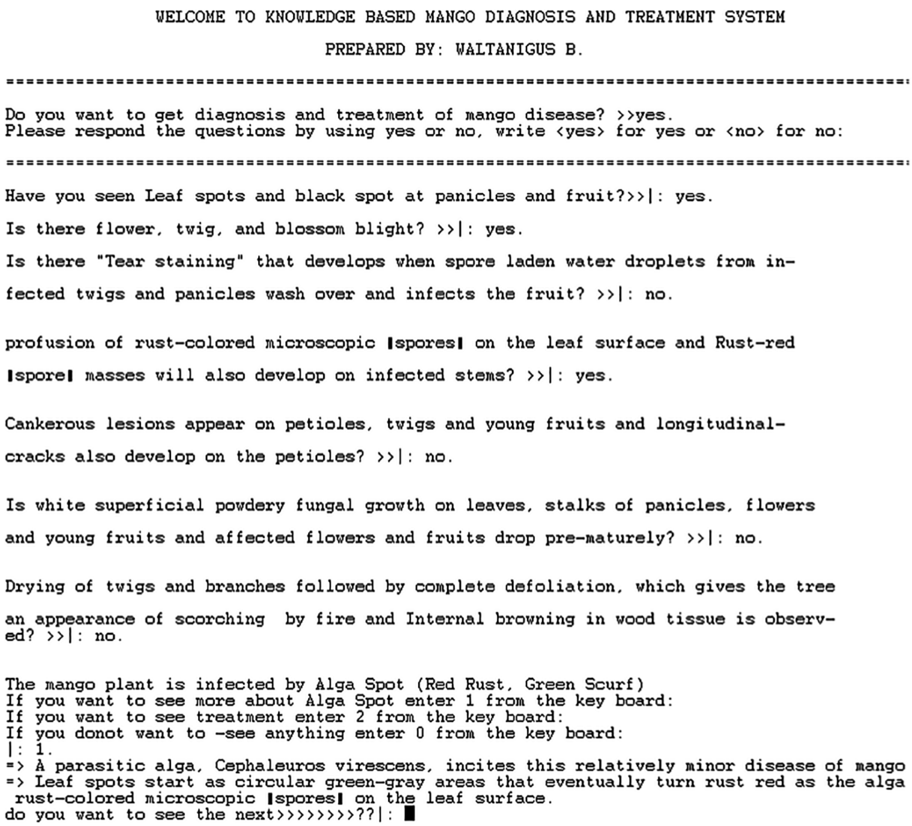


Fig. 2. How the prototype system accepts user response and provide diagnosis

One of the interesting features of knowledge based systems is their ability to explain themselves. Given that the system knows which rules were used during the inference process, it is possible for the system to provide those rules to the user as a means for explaining the results. In this study, explanation facilities such as; What, Why and Why-Not are incorporated [17]. A prototype system can provide facility of what signs and symptoms mean which is used during diagnosis and why it is selected as the signs and symptoms of mango diseases for diagnosis and treatments. Figure 4 show how a prototype system provides explanation facility for the signs and symptoms.

Researchers used backward chaining reasoning strategy. The explanation facility helps the Knowledge Based System to clarify and justify why such a digression might be needed [19].

```
              WELCOME TO KNOWLEDGE BASED SYSTEM FOR MANGO DISEASE DIAGNOSIS AND TREATMENT

                                 PREPARED BY: WALTANIGUS B.

==========================================================================================

Do you want to get diagnosis and treatment of mango disease? >>no.

WELL COME TO KNOWLEDGE BASED RECOMMENDER SYSTEM FOR MANGO DISEASE

press 1 to know History of Mango and Major Mango Diseases in Ethiopia.
----------------------------------------------------------------------------------
press 2 to know Major Symptoms of Each Disease of Mango Plant.
----------------------------------------------------------------------------------
press 3 to know Insects and Pests those Negatively Affect the Production of Mango.
----------------------------------------------------------------------------------
press 4 to know How Prevent Insects and Pests
----------------------------------------------------------------------------------
press 5 to know What Cultural management practices are used by Ethiopian Farmers.
----------------------------------------------------------------------------------
press 0 to exit from the keyboard
|: 5.
Grower farmers were opted to use variety of cultural practices in efforts to stop/mitigate
the effect of the pest in mango production.

The cultural practices used by Farmers are:

 -Pruning | to inhibit the transmission of the pest

 -Removing the infected trees- not to contaminate other normal tree or part of the tree

 -Smoking of mango tree for chess out the pest from the tree

 -Washing with soluble ash and soap

 -Thinning | for spacing among the planted trees through removing the tree,

 -Burning

 -Keeping the sanitary of soil under the tree

Please insert your choice number from the list>>>>>>>>>??|: █
```

Fig. 3. Sample dialogues which shows how prototype system provides information about cultural management practices

```
                WELCOME TO KNOWLEDGE BASED MANGO DIAGNOSIS AND TREATMENT SYSTEM

                              PREPARED BY: WALTANIGUS B.

========================================================================================== >

Do you want to get diagnosis and treatment of mango disease? >>yes.
Please respond the questions by using yes or no, write <yes> for yes or <no> for no:

========================================================================================== >

Please give (yes.) or (no.) answer to the following questions and
If the questions are unclear you can ask by entering (what.) and (why.)

Have you seen Leaf spots and black spot at panicles and fruit?(yes/no/what): |:what.

This is to mean have you seen a small round or roundish mark, holes which appears on the leaves

 of the plant. ?(yes/no/why): |: why.

Because leaf spot is the most important symptom of mango disease. So! have seen spots?(yes/no):
|: yes.
```

Fig. 4. Sample dialogues that shows how prototype system gives explanation facility

8 Testing and Evaluation of the Prototype System

To evaluate the performance of the system, two set of testing was conducted. In the first case, the system is tested for its performance on how diagnosis and providing treatment of mango diseases. In the second cases, the evaluation was conducted to evaluate the system's acceptance level by the users through directly interacting with the system. To test the performance of the system we have employed confusion matrix and the performance of the system was computed using recall, precision and F measure for measuring effectiveness. The accuracy of the prototype system is calculated as 81.8% based on the test cases which indicate that the prototype has a very good performance and the average evaluation result filled by the domain experts in the domain area is 86.29%, respectively. The overall performance of the prototype system is 84.05%. The users of the system had given a promising feedback. In general, the testing and evaluation result of the prototype system has achieved the objectives of the study. However, additional study is needed to bring complete implementation and use of knowledge-based system for diagnosis and treatment of mango diseases.

9 Conclusion and Recommendation

9.1 Conclusion

In Ethiopia most of the peoples are farmers, it's believed that more than 80% of the population depends on agriculture and it looks many challenges like plant diseases. Mango diseases are problematic plant diseases all over the word. It causes the destruction of mango plant in many countries especially in tropical area and developing countries where there is scarcity of agronomy experts (pathologist) and tools for diagnosis and treatments. Therefore it's logical to develop a knowledge based expert system to emulate the knowledge of agricultural experts. It gets attention on diagnosis and treatment of diseases. So, Rule based reasoning was used for the development of a prototype KBS and the developed prototype correctly assists agricultural professionals and farmers. However, in order to make the system applicable in the domain area, some adjustments like automatically updating the rules in the knowledge base of the system and incorporating a well designed user interface are needed.

9.2 Recommendation

The following recommendations were given based on the observed opportunities and uncover areas by this research.

- Integrating the application of data mining techniques with knowledge-based systems
- Developing the system in local language, like in Amharic, Afan Oromo and Tigrigna are recommended as future work.
- Last of all, this knowledge base system is not self learning, in the future learning component should be integrated

References

1. ICARDA: Ethiopia and ICARDA. Ties that Bind, no. 29. ICARDA, Aleppo, Syria, iv +20 pp. En (2010). ISBN 92-9127-245-0
2. Mellor, J., Dorosh, P.: Agriculture and the economic transformation of Ethiopia development strategy and governance division: Ethiopia strategy support program 2. In: International Food Policy Research Institute (2010)
3. CSA: Area and Production of Major Crops. Sample Enumeration Survey. Addis Ababa, Ethiopia (2009)
4. Berhe, K., Dessalegn, Y., Baredo, Y., Teka, W., Tegegne, A.: Smallholder-based fruit seedling supply system for sustainable fruit production in Ethiopia. Improving Productivity and Market Success (IPMS) of Ethiopia Farmers Project, Addis Ababa, Ethiopia (2007)
5. Central Intelligence Agency (CIA): Banana in Southern Region of Ethiopia, pp. 119–134 (2010)
6. Etisa, E., Ayele, L., Tadesse, D.: Review on the status of some tropical fruits in Ethiopia. In: Proceeding of the Inaugural and First Ethiopian Horticultural Society Conference, Addis Ababa, Ethiopia, pp. 39–44 (2006)
7. C.S.A (Central Statistics Authority), Addis Ababa, Ethiopia (2012/2013)
8. Dubey, S., Pandey, R., Gautam, S.: Literature review on fuzzy expert system in agriculture. Int. J. Soft Comput. Eng. (IJSCE) **2**, 289–290 (2013)
9. Marwaha, S.: AGRIdaksh—a tool for developing online expert system. Division of Computer Applications, Indian Agricultural Statistics Research Institute, p. 17 (2012)
10. Turban, E., Aronson, J.E.: Decision Support and Intelligent Systems. Prentice Hall International Inc., New Jersey (1990)

11. Ranjan, Y.: Farm advisory system for farmers of Northeastern States of India. Int. J. Innov. Eng. Technol. (IJIET) **5**, 169–172 (2015)
12. Islam, S.N.: Expert System on Wheat Crop Management. Development of Expert System through AGRIdaksh, New Delhi (2004)
13. Bezu, T.: Production Scenarios of Mango (Mangifera indica L.) in Harari Regional State, Eastern Ethiopia. Science, Technology and Arts Research Journal, Haramaya University, Dire Dawa, Ethiopia (2014)
14. Dewanto, S., Lukas, J.: Experet system for diagnosis pest and disease in fruit plants. In: EPJ Web of Conferences, Universitas Bina Nusantara, Computer Engineering Department, Jakarta, Indonesia (2014)
15. Mercy, P., Rani, N.: Expert systems in agriculture: a review. J. Comput. Sci. Appl. **3**(1), 59–71 (2011)
16. Dipanwita, B., Bairagi, S., Panse, N., Shinde, N.: Disease diagnosis system. J. Comput. Sci. Inform. **1**(2), 48–51 (2011)
17. Dagnew, G.: Designing a knowledge based system for blood transfusion (2012)
18. Ambisa, B.: Develop a knowledge based system for coffee diseases and pests' diagnosis and treatment, vol. 1, no. A thesis submitted in partial fulfillment of the requirements for the degree of Master of Information science (2012)
19. Gebremariam, S.: A Self-Learning Knowledge Based System for Diagnosis and Treatment of Diabetes (2013)

A Genetic Algorithm-Based Approach for Test Case Prioritization

Getachew Mekuria Habtemariam and Sudhir Kumar Mohapatra(✉)

Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
getachewmekuria19@gmail.com,
sudhirmohapatra@hotmail.com

Abstract. Software maintenance is the most important and expensive activity in the process of Software Development Life Cycle (SDLC). In the maintenance stage, after modification, the software goes through verification and validation. Regression testing is performed at the maintenance stage of SDLC to ensure that the old functionalities are working perfectly. Test case prioritization, which is about making a sequence of the test case, is one of the important parts of regression testing. Test case prioritization(TCP), which is a class of NP-hard problem, can have a better solution using soft computing approach as per no-free-lunch theorem [1]. The same no-free-lunch theorem states that a soft computing approach yields a case-specific result. In this article, we propose a novel genetic algorithm approach to solve the TCP problem. The proposed algorithm is experimentally compared with random technique. For the experiment three-benchmark program from the software artifact infrastructure repository is selected. From the experiment, it was found that the Average Percentage of Fault Detection (APFD) of the GA technique gives a better result than random technique. Also, an experiment was conducted to record the execution time with different numbers of generations. The results confirm that the result is not directly dependent on the number of generation.

Keywords: Software maintenance · Regression testing · Test case prioritization · Soft computing · Genetic algorithm

1 Introduction

The effective application of ICT depends not only on the efficiency of the computer but also robustness of the software that programs the computer. Accuracy and security of software are two important features, which must be ensured, for robustness of the software. Thus, every software has to undergo the process of testing before it is subjected to use. Needless to say, that if a software is inefficient or lacks security then the very purpose behind its development is not only defeated, the economy of the organization is also adversely impacted. The software with high functionality needs stringent testing methodology [2, 3]. The growing size and complexity of the software make regression testing very important in the maintenance stage of SDLC. Test suite prioritization, an important part of regression testing is used to find an ordering of test cases. Test case prioritization promises to detect the faults early in the re-testing process. Thus, finding an optimal order of execution of the selected regression test cases

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 24–37, 2019.
https://doi.org/10.1007/978-3-030-26630-1_3

will maximize the error detection rates at less time and cost. Early detection of error helps us in concluding the testing process in case of time constraints.

A substantial part of the project cost is incurred in testing, which is around 30–50% [4]. As exhaustive testing is practically impossible, early detection of error helps to circumvent the testing process without going for the method of retest all [4]. Looking into the round the clock use of the software, the tester has the responsibility to make the software fault-free [5, 6]. Rather than using retest all approach, it would be better to partially test the software with those test cases, which have the possibility of revealing the maximum number of faults [7]. While the partial selection of the test case will not be always effective, in a resource-constrained environment it is implicit. The strategy of test case prioritization encompasses the execution of the test cases with the highest priority, consistent with some fitness metric and thereby leading to their eventual ordering [8–10]. Rothermel et al. outlined the test case prioritization problem and propounded numerous solutions [9].

The test case prioritization problem, defined (by Rothermel et al.) is stated underneath:

The Test Case Prioritization (TCP) Problem:
Given: T, a test suite; PT, the set of permutations of T; f, a function from PT to the real numbers.
Problem: Find T' ∈ PT such that $(\forall T'' (T'' \in PT) (T'' \neq T') [f(T'') \geq (T'')])$
The PT represents all permutation of the test case set T and f is a function used to generate an ordering.

Solution to combinatorial problem with minimum cost is best possible using soft computing techniques [11]. Evolutionary algorithms, of which genetic algorithms (GAs) are a subclass, rely on search. GAs use Darwin's "Survival of fittest" theory. In the case of scheduling problems, of which regression test case prioritization is an example, the application of GAs has been shown to be effective [12]. Thus, prioritization techniques based on fault exposing potential of the test case is essentially heuristic [9]. The TCP, which is a class of knapsack-problem belongs to NP-hard [13]. In this research-article, we have tried to probe into the use of fault exposing potential of the test case for test case prioritization.

The organization of this paper is as follows. In Sect. 2 we have specified existing test case prioritization technique. In Sect. 3 an algorithm based on the genetic algorithm for test case reduction is proposed and discussed. The detail execution of the testing process is discussed in Sect. 4 followed by experiment and result discussion in Sect. 5. In the last section, the findings of the paper were summarized.

2 Related Work

Yadev et al. proposed a prioritization technique based fuzzy logic taking multiple factors into considaration [14]. The factors used are fault detection rate of program, execution time and requirement coverage. Khatibsyarbini et al. review the published article on TCP and the review document gives an insight of test case prioritization research [15]. Elbaum et al. and Rothermel et al. used different criteria for test case prioritization in [9, 16]. The authors' used code coverage, branch coverage and calculable ability to reveal faults. These algorithms are the greedy based algorithm. Srivastava and Thiagarajan, using the change created by the program as a basic criterion, proposed TCP [17]. They assumed that for quick detection of error the test case should be ordered in terms of their high potential to cover the affected elements of the program. A model-based prioritization was proposed by Panigrahi and Mall [18, 19]. They made use of UML diagram to propose two TCP techniques, viz., S-RTP and H-RTP. The proposed method considered the affected nodes in the ESDG (Extended System Dependency Graph) model. The number of affected nodes exercised by a test case was used as the base to derive the sequence of the test case. The authors propose a white-box test case prioritization, selection, and minimization considering the reuse context [20]. In order to prioritize regression test cases, Li et al. further proposed other greedy strategies including the 2-optimal strategy [21].

The coverage of the relevant slice of a test case was used as a criterion for TCP by Jeffrey and Gupta [22]. They presented a greedy algorithm where the relevant slice was implemented. Relevant slice is a measure of the statements that influence or could influence the output of a program when executed on a regression test case [22]. Mohapatra et al. proposed a genetic algorithm-based TCP [24]. The multi objective prioritization they proposed was based on code coverage and severity of the test case. Smith et al. used call tree for TCP [23] while Mohapatra et al. used chemical reaction optimization [25] and ant colony optimization [26] for the purpose. A new test case prioritization algorithm is proposed to get better the rate of fault detection and cost reduction using Simulated Annealing (SA) [27]. Noguchi et al. proposed a framework using Ant Colony Optimization [28]. The proposed framework is prioritized test cases for black box testing on a new product using the test execution history collected from a similar prior product.

Various researchers have perused the pros and cons of numerous approaches to test case prioritization. As exemplary evidence, the influence of the greedy algorithm on test case prioritization was scrutinized by Rothermel et al. and conducted a thorough empirical study on the available greedy methods for TCP and found that the greedy strategy is always not suitable [9]. In the study, the authors experimentally proved that these greedy algorithms sometimes consume more time without producing an optimal result. This creates practical snag for the sole use of the greedy algorithm for TCP. In the backdrop of the above, the delving of the plausibility of applying soft computing approach for TCP by scrutiny of its NP-hard nature is an interesting proposition [13].

3 Proposed Algorithm

The algorithm creates a sequence of minimum test cases, capable of exposing all the faults in concert. Initially, the algorithm creates a chromosome of length two (*Chromolen = 2*). On this population, all the genetic operators are applied. If at any generation (epoch) any chromosome is able to expose all the faults then the algorithm stops and returns the sequence. Otherwise, the GA process starts for chromosome length three. The chromosome-length shifts to the $(l + 1)$ level in each iteration until a sequence with 100% fault exposing potential is detected. Once the algorithm generates a sequence the rest of the test cases (which are not in the sequence) may be added in any order.

```
Input:
  1. T={Set of test case}
  2. Test case fault detection details
  as a traceability  matrix
  3. Pc =Crossover probability
  4. Pm =Mutation probability
  5. maxGen= Maximum GA generation

  Output:
  Sequence of minimized test case dis-
  covering maximum faults(ρmin)

  Abbreviations used:
    cp: Crossover point
    mp: Mutation point

    Prio_GA(T, TR, maxGen, Pc, Pm)
    Begin
    l←2
    Popsize←20
    loop
    InitialPopulation(Popsize,l)
    g ← 1
    loop
    CalculateFitness(P,TR)
       Crossover(Pc,P,l)
       Mutation(Pm,P,l)
         If  (Any order of test suit dis-
   cover 100% fault)
         ρmin ← FindMax(P)
    else
         g← g+1
    until g<=maxGen
    l← l+1
    until l<= Chromolen
    return ρmin
    end

    InitialPopulation(Popsize,Chromolen)
    Begin
    a←1
    repeat
    Pa ←{Ø};
    repeat
    Pa ← Pa U { Randomly from T which was
    not in the chromosome}
    until | Pa |<= Chromolen
```

```
i←i+1
until i<=Popsize
end
```

CalculateFitness(P,TR)

```
begin
i←1
loop
```

$P_i = \frac{|\cup_{i=1}^{n} Ti \subseteq F|}{|F|}$

```
i←i+1
until i<Popsize
end
```

Crossover(P_c,P,l)

```
begin
for Pc percentage of chromosome
generate cp ∴1<=sp<=l
at point cp exchange chromosome
end
```

Mutation(P_m,P,l)

```
begin
for Pm percentage of chromosome
generate mp ∴1<=sp<=l
replace a duplicate test case
with a test case which is not
present in that chromosome
end
```

FindMax(P)

```
begin
sort the population in
ascending order of their
fitness value.
end
```

3.1 Initial Population

The function *InitialPopulation (Popsize, Chromolen)* is responsible to generate the initial population. It will consider the population size and chromosome length as the parameters. Permutation encoding was used in our algorithm. The function created a single chromosome at a time (in one iteration of the loop), with the length same as *Chromolen* value. The method is employed with utmost care in order to avoid the duplication of the test cases in a chromosome. A typical case of chromosome <T6, T7, T9, T71, T23, T12, T21>, with *Chromolen* = *7*, is represented in Fig. 1.

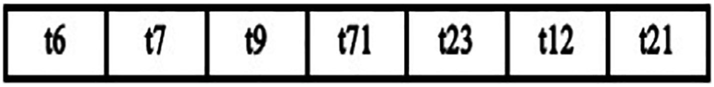


Fig. 1. Example of chromosome generated using permutation encoding

3.2 Selection

Rank selection method was used to select the chromosome to move to the next epoch. The function *FindMax(P)* was used for this, it sorts the individual chromosome based on their fitness value in ascending order. After sorting it select the best chromosome.

3.3 Crossover

Single point crossover is used in the algorithm. The crossover probability, $P_c = 0.6$ was used. *Crossover* (P_c, *P, l)* function performed the crossover activity. It could randomly select a point between one and the length of the chromosome. This served as the crossover point. Chromosomal-exchanges at the cross over points generated new off springs, as exemplified in Fig. 2.

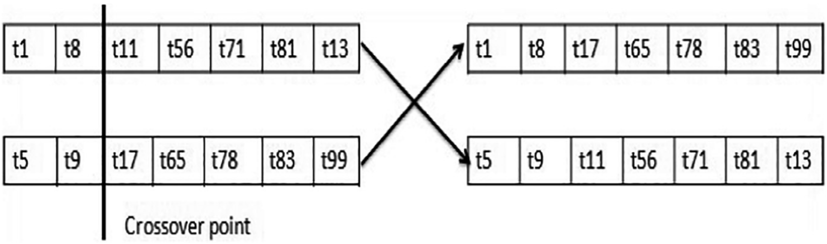


Fig. 2. Single point crossover applied to the parent to generate children

3.4 Mutation

The *Mutation (P_m, P, l)* method was used to perform the mutation operation of GA. The role of this operator in the algorithm is to remove duplicate test case from the chromosome. Sometimes we get a duplicate test case in a chromosome after crossover. *Mutation (P_m, P, l)* checks a chromosome and if there is any duplicate test case identified, replaces it with a new test case. The number of chromosomes for which mutation is applied will be decided by the parameter P_m *i.e.*, the mutation probability. In the present work, we take $P_m = 0.4$. The mutation process we followed is depicted in Fig. 3.

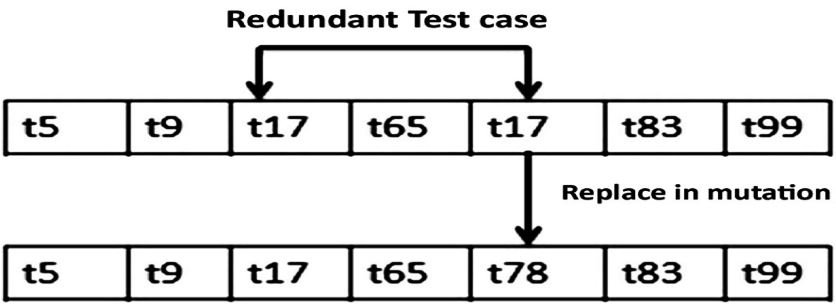


Fig. 3. Mutation operation applied to replace duplicate test case from chromosome

3.5 Fitness Function

This is the most important function, which decides the fitness of the chromosome to go to the next generation. The fitness function, used in the paper, relies on the Eq. (1)

$$\text{Fitness function} = \frac{\left|\mathrm{U}_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{Ti} \ \subseteq \mathrm{R}\right|}{|\mathrm{R}|} \tag{1}$$

Subjects to Ti ϵ T, where T is the current chromosome of *l* length

The fitness value of each chromosome varies in between 0 and 1. For example, if a chromosome of length 2 is {T3, T5}. Table 1 presents that test case T3 is able to expose faults $\{f1, f3\}$ and test case T5, $\{f3, f6\}$. The fitness value for the said example is

$$\mathit{Fitness\,value} = \frac{|\{f1, f3\} \cup \{f3, f6\}|}{6} = \frac{4}{6} = 0.67$$

The Eq. (1) was applied using the logical OR operator. The TR matrix was used for the fitness calculation of each chromosome. To understand, let us take a chromosome of length two i.e. T = {t2, t4} fitness value of T will be (based on TR matrix of Table 1)

$$
\begin{array}{r}
t2 = \{011110\} \\
t4 = \{001001\} \\
\hline
\text{OR}\ \{0\,1\,1\,1\,1\,1\}
\end{array}
$$

Chromosome T reveals five faults and the total number of faults present in the program is six, the fitness value of the T = $\frac{\{011110\}\ \mathrm{U}\ \{001001\}}{6} = \frac{\{011111\}}{6} = \frac{5}{6} = 0.83$

4 Proposed Model

The execution-details of the GA are shown in Fig. 4. Let F represents total faults, total test case represented by T and program to be tested is P. The execution-details of test case T for the previous version are collected from its test oracle. The test oracle is used to create the traceability matrix (TR). This TR matrix is a 0-1 matrix. The row of the matrix represents a test case and a column represents a fault. Table 1 presents an example of the relation between the test case and fault and how TR matrix is generated from Table 1.

Table 1. Traceability(TR) matrix generated from the table representing the test case and fault expose by them.

Test case	Fault expose by the test case					
	F1	F2	F3	F4	F5	F6
T1	1	1	1	0	0	0
T2	0	1	1	1	1	0
T3	1	0	0	0	0	1
T4	0	0	1	0	0	1
T5	0	1	1	0	0	0
T6	1	1	0	0	0	1
T7	0	0	0	1	0	1

$$
TR = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1
\end{pmatrix}
$$

In the above example, there are seven test cases and six faults. The entry in the cell TR_{ij} represents

$$TR_{ij} = \begin{cases} 1 & iftestcaseiexposefaultj \\ 0 & otherwise \end{cases}$$

The TR matrix is used by GA to generate prioritised test case T'. The whole methodology of the model is presented in Fig. 4.

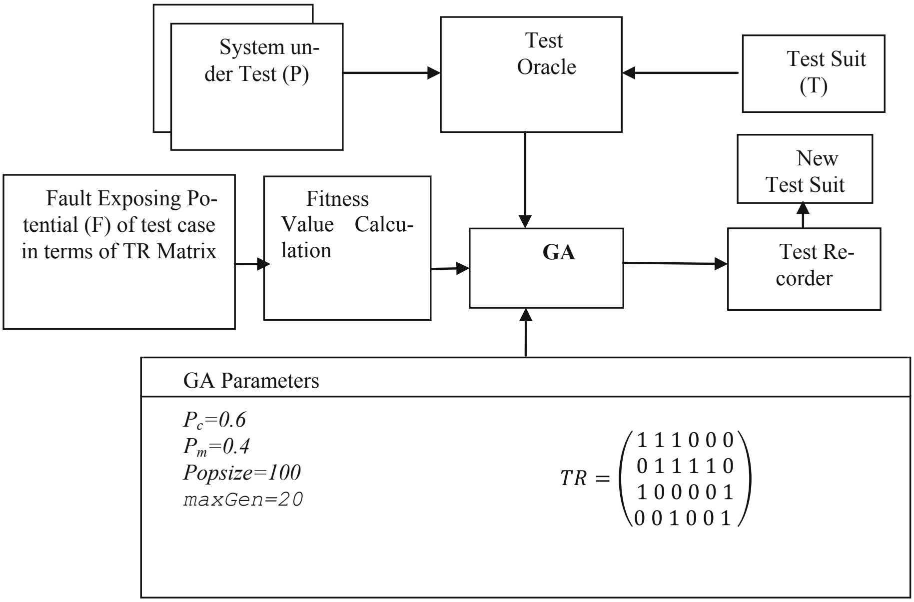


Fig. 4. The methodology applied to generate TCP order using GA

5 Experiment and Results

An empirical study was performed to find out the effectiveness of the technique. The major focus of the experiment was to measure the average percentage of fault detection (APFD) of the different subject program [22, 29]. The proposed model was compared to the much-used random technique. The experiment was carried out on three benchmark subject program from Software-artifact Infrastructure Repository (SIR) repository[1]. SIR is a repository of software-related artifacts meant to support rigorous controlled experimentation with program analysis and software testing techniques, and education in controlled experimentation. The details of the subject program are shown in Table 2.

[1] Software-artifact Infrastructure Repository, https://sir.csc.ncsu.edu/portal/index.php (Accessed 5th September 2018).

In order to carry out the experiment, we seeded fault on the benchmark program using Jumble[2]. The test cases were Junit[3] test cases, executed using Eclipse IDE[4]. The GA was implemented on MATLAB R2017a. It was executed on a PC with an Intel Pentium 3.2 GHz CPU and 1 GB memory running the Windows 10 operating system.

Table 2. Summary of programs used in experimentation

Program	Source file (LOC)	Test suite pool (T × R)	Mutation fault
Account	66	25 × 12	20
BST	130	57 × 22	21
Groovy	361	102 × 92	77

5.1 Research Questions

The execution of the experiments was dictated by the following two queries:

Q1: Does our GA increase early detection of fault more significantly than random technique?
Q2: Is GA based TCP efficient in terms of time and chromosome length?

5.2 Experiment to Evaluate APDF of the Different Subject Program Using Random and GA Based Technique

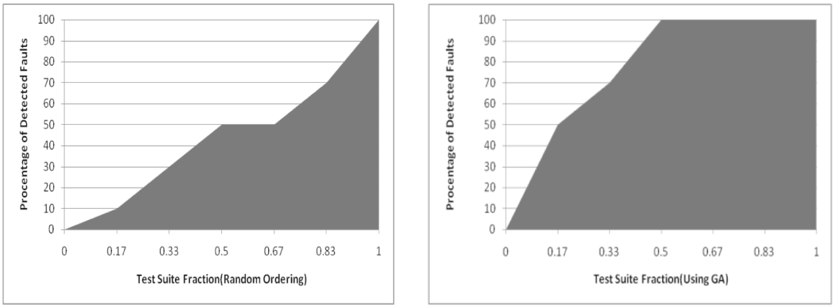
The experiments on the subject program were carried out to calculate APFD value for our GA based technique and random technique. The results are presented in Fig. 5.

Analysis of the APFD value of account program evinced that it could achieve 100% fault exposing potential after execution of 50% test case using GA prioritization. On the contrary, for groovy and BST, 100% achievement was noted after the execution of 67% of the test cases. In the case of random prioritization, 100% FEP was achieved only after execution of all the test cases (like retest all) for BST and account, whereas, in the case of groovy, it was achieved after 83% test case execution. The GA TCP detected all the faults much earlier than random prioritization.

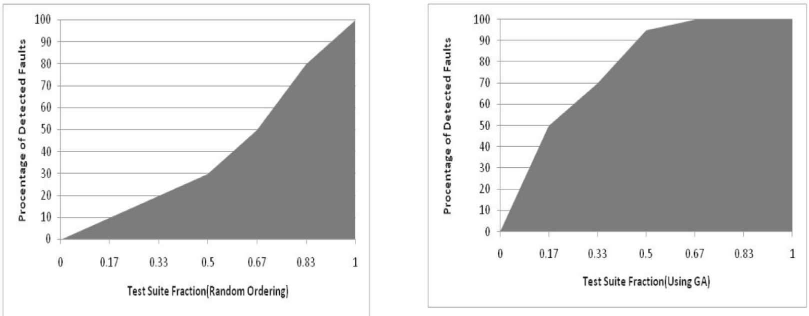
5.3 Experiment to Evaluate the Execution Time (Time Complexity) with Different Chromosome Length and Number of Generations

This experiment was conducted to assess the efficiency of the algorithm in terms of execution time. In order to perform this experiment, we considered variation in number of generations (*maxGen*). The three-subject program was tested for this parameter and their individual execution time was recorded (Fig. 6).

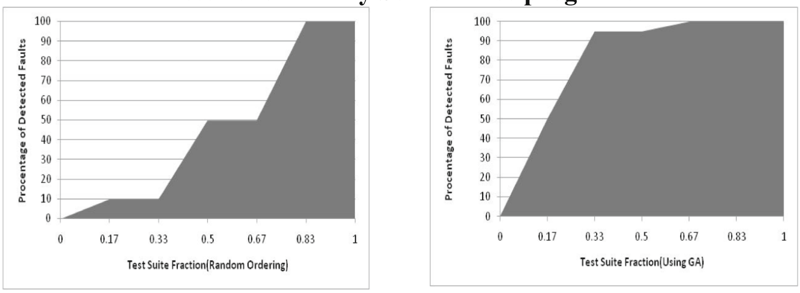
2 Jumble home page http://jumble.sourceforge.net (Accessed 7th September 2018).
3 JUnit's official website, http://www.junit.org/ (Accessed 7th September 2018).
4 The Eclipse Foundation website, http://www.eclipse.org/ (Accessed 7th September 2018).



Results for account program



Results for Binary Search Tree programs



Results for groovy programs

Fig. 5. Details of the APFD results of the subject program for GA and random TCP.

GA TCP generates quick result if a number of generations are less. The increase in the number of generations not always gives early result rather it increases execution time. The algorithm could efficiently generate the TCP order with less number of the epoch (generation).

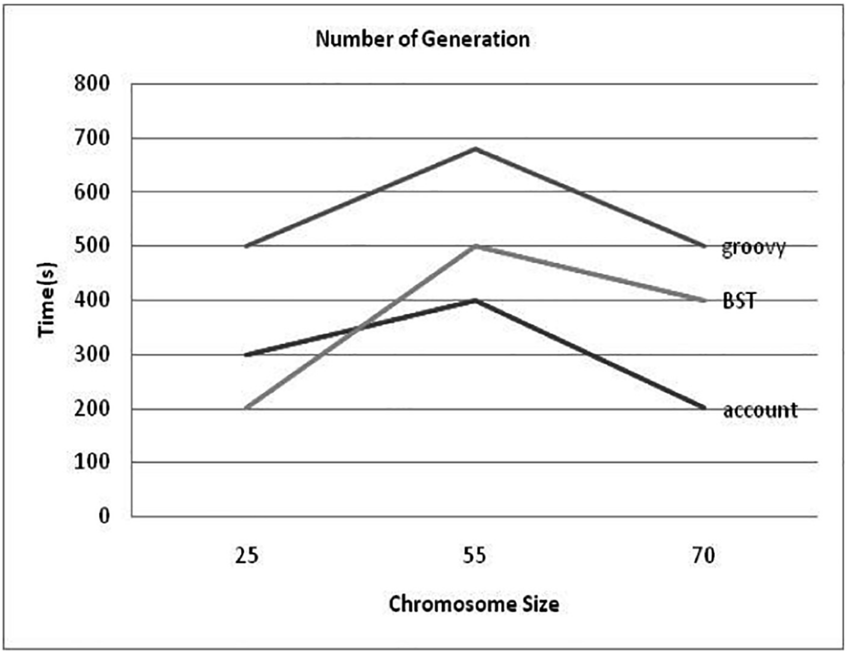


Fig. 6. Execution time results of the three-subject program with different chromosome size.

6 Conclusion

This paper presents a successful design of a GA algorithm to prioritize test cases. The efficiency of the algorithm was assessed with respect to random prioritization. APFD measure was used for the calculation of the efficiency. Analysis of the APFD measure for three-benchmark program gave early fault detection by GA with respect to random technique. GA TCP execution was faster as vouched by the exposition of 100% fault with less number of generations. Future research may be directed towards the improvement of the performance for some higher LOC benchmark programs. It would also be an interesting proposition to investigate whether prioritization technique modified using parallel GA could decrease the execution time with same or better APFD value.

References

1. Koppen, M., Wolpert, D.H., Macready, W.G.: Remarks on a recent paper on the "no free lunch" theorems. IEEE Trans. Evol. Comput. **5**(3), 295–296 (2001)
2. Ostrand, T.J., Weyuker, E.J., Bell, R.M.: Where the bugs are. In: ACM SIGSOFT Software Engineering Notes 2004 Jul 11, vol. 29, no. 4, pp. 86–96. ACM (2004)
3. Ahmed, A.: Software Testing as a Service. Auerbach Publications, Boca Raton (2009)

4. Ramler, R., Biffl, S., Grünbacher, P.: Value-based management of software testing. In: Biffl, S., Aurum, A., Boehm, B., Erdogmus, H., Grünbacher, P. (eds.) Value-Based Software Engineering, pp. 225–244. Springer, Heidelberg (2006). https://doi.org/10.1007/3-540-29263-2_11
5. Boehm, B., Huang, L.G.: Value-based software engineering: a case study. Computer **36**(3), 33–41 (2003)
6. Zhang, L., Hou, S.S., Guo, C., Xie, T., Mei, H.: Time-aware test-case prioritization using integer linear programming. In: Proceedings of the Eighteenth International Symposium on Software Testing and Analysis, 19 July 2009, pp. 213–224. ACM (2009)
7. Ashraf, E., Rauf, A., Mahmood, K.: Value based regression test case prioritization. In: Proceedings of the World Congress on Engineering and Computer Science, vol. 1, pp. 24–26 (2012)
8. Rothermel, G., Untch, R.H., Chu, C., Harrold, M.J.: Test case prioritization: An empirical study. In: Proceedings IEEE International Conference on Software Maintenance - (ICSM 1999). Software Maintenance for Business Change (Cat. No. 99CB36360), pp. 179–188. IEEE (1999)
9. Rothermel, G., Untch, R.H., Chu, C., Harrold, M.J.: Prioritizing test cases for regression testing. IEEE Trans. Softw. Eng. **27**(10), 929–948 (2001)
10. Wong, W.E., Horgan, J.R., London, S., Agrawal, H.: A study of effective regression testing in practice. In: Proceedings of the Eighth International Symposium On Software Reliability Engineering, pp. 264–274. IEEE, 2 November 1997
11. Beasley, J., Dowsland, K., Glover, F., Laguna, M., Peterson, C., Reeves, C.R.: Modern heuristic techniques for combinatorial problems, p. 320. Colin R. Reeves, New York (1993)
12. Husbands, P.: Genetic algorithms for scheduling. AISB Q. **89**, 38–45 (1994)
13. Gary, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. WH Freeman, New York (2002)
14. Yadav, D.K., Dutta, S.: Test case prioritization technique based on early fault detection using fuzzy logic. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1033–1036. IEEE, 16 March 2016
15. Khatibsyarbini, M., Isa, M.A., Jawawi, D.N., Tumeng, R.: Test case prioritization approaches in regression testing: a systematic literature review. Inf. Softw. Technol. **1** (93), 74–93 (2018)
16. Elbaum, S., Malishevsky, A.G., Rothermel, G.: Test case prioritization: a family of empirical studies. IEEE Trans. Softw. Eng. **28**(2), 159–182 (2002)
17. Srivastava, A., Thiagarajan, J.: Effectively prioritizing tests in development environment. In: ACM SIGSOFT Software Engineering Notes, vol. 27, no. 4, pp. 97–106. ACM, 22 July 2002
18. Panigrahi, C.R., Mall, R.: An approach to prioritize the regression test cases of object-oriented programs. CSI Trans. ICT **1**(2), 159–173 (2013)
19. Panigrahi, C.R., Mall, R.: A heuristic-based regression test case prioritization approach for object-oriented programs. Innov. Syst. Softw. Eng. **10**(3), 155–163 (2014)
20. Miranda, B., Bertolino, A.: Scope-aided test prioritization, selection and minimization for software reuse. J. Syst. Softw. **1**(131), 528–549 (2017)
21. Li, Z., Harman, M., Hierons, R.M.: Search algorithms for regression test case prioritization. IEEE Trans. Softw. Eng. **33**(4), 225–237 (2007)
22. Jeffrey, D., Gupta, N.: Experiments with test case prioritization using relevant slices. J. Syst. Softw. **81**(2), 196–221 (2008)
23. Smith, A.M., Geiger, J., Kapfhammer, G.M., Soffa, M.L.: Test suite reduction and prioritization with call trees. In: Proceedings of the Twenty-Second IEEE/ACM International Conference on Automated Software Engineering, pp. 539–540. ACM, 5 November 2007

24. Mohapatra, S.K., Prasad, S.: Evolutionary search algorithms for test case prioritization. In: 2013 International Conference on Machine Intelligence and Research Advancement, pp. 115–119. IEEE, 21 December 2013
25. Mohapatra, S.K., Prasad, S.: A chemical reaction optimization approach to prioritize the regression test cases of object-oriented programs. J. ICT Res. Appl. **11**(2), 113–130 (2017)
26. Mohapatra, S.K., Srinivas, P.: An ant colony algorithm to prioritize the regression test cases of object-oriented programs. Indian J. Sci. Technol. **9**(19), 1–8 (2016)
27. Maheswari, R.U., Mala, D.J.: Combined genetic and simulated annealing approach for test case prioritization. Indian J. Sci. Technol. **8**(35), 1–5 (2015)
28. Noguchi, T., Washizaki, H., Fukazawa, Y., Sato, A., Ota, K.: History-based test case prioritization for black box testing using ant colony optimization. In: 2015 IEEE 8th International Conference on Software Testing, Verification and Validation (ICST), pp. 1–2. IEEE, 13 April 2015
29. Elbaum, S., Malishevsky, A., Rothermel, G.: Incorporating varying test costs and fault severities into test case prioritization. In: Proceedings of the 23rd International Conference on Software Engineering, pp. 329–338. IEEE Computer Society, 1 July 2001

Stationary Wavelet Transform for Automatic Epileptic Seizure Detection

Gebremichael Shiferaw[1(✉)], Adane Mamuye[2], and Marco Piangerelli[3]

[1] Department of Information Systems, University of Gondar, Gondar, Ethiopia
gebrsh@gmail.com
[2] Department of Computer Science, University of Gondar, Gondar, Ethiopia
adothebigg@gmail.com
[3] Department of Computer science, University of Camerino, Camerino, Italy
marco.piangerelli@unicam.it

Abstract. Visual detection of epileptic seizure from EEG signal is being inefficient and time consuming. Computational EEG signal analysis techniques were then used in the diagnosis and management of epileptic seizures. In this study, we compared the performance of Discrete Wavelet Transform (DWT) and the Stationary Wavelet Transform (SWT) decomposition techniques with 22 wavelet functions (Coiflets (coif), Daubechies (DB) and Symlets (Sym) families) using support vector machine classifier. We used multichannel EEG dataset of the University of Bon Epilepsy Center. From this dataset, five statistical wavelet features: max, min, average, mean of absolute and standard deviation were extracted. In all of the wavelet functions except three, in the Coiflets family, the experimental result showed that SWT achieved better classification accuracy than DWT. SWT and DWT decomposition techniques registered 99.5% and 97.5% highest classification accuracies, respectively.

Keywords: EEG signal · Discrete wavelet transforms · Stationary wavelet transforms · Seizure detection · SVM Classifiers · Wavelet functions

1 Introduction

Epilepsy is a common chronic neurological disorder which affects around 50 million people of all ages all over the world. In low and middle-income countries, nearly three fourth of the epilepsy patients do not get the treatment they need (Pellegrino et al. 2014; Shi et al. 2017). The hallmark of epilepsy is recurrent seizures, termed epileptic seizures. A seizure occurs when parts of the brain producing abnormal electrical activity that spread throughout the brain and epilepsy is caused by sudden recurrent and transient disturbances of perception.

Monitoring the electrical activities from the brain through the electroencephalogram (EEG) has become an important data source for identification of epileptic seizure (Zandi et al. 2010). Electroencephalogram (EEG) records the

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 38–45, 2019.
https://doi.org/10.1007/978-3-030-26630-1_4

changes of the electrical activity in term of voltage fluctuations of the brain through multiple electrodes placed in the different location of the brain. EEG signal of epilepsy patients during a seizure shows patterns that are significantly different from the normal state of the brain with respect to space, time and frequency patterns. However, visual detection of epileptic seizure using EEG tool has a serious drawbacks. Visual detection is very time consuming and inefficient, especially in long-term recordings practitioners have to go through huge number of EEG data sheets. Additionally, disagreement among the practitioners on the same recording may possible due to the subjective nature of the EEG analysis. Moreover, the EEG patterns that characterize an epileptic seizure are similar to waves that are part of the background noise and to artefacts such as eye blinks and other eye movements, muscle activity, electrocardiogram and electrical interference. Due to these reasons, automated classification has been forwarded as valuable clinical tools for the scrutiny of EEG data in computationally efficient manner (Pellegrino et al. 2014; Tzallas et al. 2012; Orosco et al. 2013; Akareddy et al. 2013; Chen et al. 2015).

In developing the automated classifier, feature extraction and classification methods based on Wavelet and Fourier transformation were employed. However, previous studies have demonstrated that Discrete Wavelet Transform is among the most promising technique for extracting features from EEG signals (Gajic et al. 2014; Kabir et al. 2016; Tibdewal et al. 2016; Prince et al. 2016; Mahdi et al. 2011; Shete et al. 2014).

On the other hand, due to signal to noise ratio, the presence of noise may affect the recognition and localization of spikes. Thus, Stationary Wavelet Transform (SWT) has been used for EEG signal spike detection (Radmehr et al. 2013). The major advantage of SWT is the preservation of time information of the original signal sequence at each level. However, as the best of the researchers' knowledge there is no study that compares the accuracy of the SWT and the DWT classification techniques on the same EEG signal dataset. Thus, the focus of this study is to compare the DWT and SWT decomposition techniques, on the same signal dataset, using support vector machine (SVM) classifier.

2 Methodology

In this study, we follow four steps to decompose and classify the EEG signals.

2.1 Dataset

We used the EEG time series database from the Epilepsy Centre of the University of Bonn (Andrzejak 2001). This database consists of five EEG datasets (Sets A-E), identified as: O, Z, F, N and S; each set holds 100 segments of EEG signals of 23.6 s. The sampling frequency of each signal was 173.6 Hz, so each segment contains 4,097 samples. O and Z datasets were obtained from healthy subjects with eyes open and closed respectively; Datasets F and N were obtained during interictal states, period between seizures, in different zones of the brain and set

S was obtained from a subject during ictal state, seizure period (Tzallas et al. 2012). For this study, 200 EEG signal segments, 100 each from datasets A and dataset E, were selected, as shown in Figs. 1 and 2.

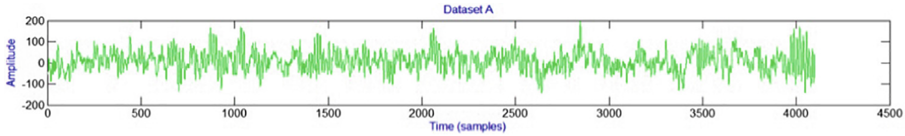


Fig. 1. Sample non-epileptic EEG signal block dataset A.

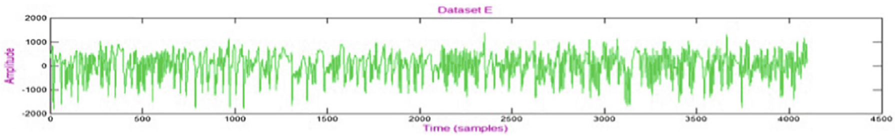


Fig. 2. Sample epileptic EEG signal block dataset E.

2.2 Signal Decomposition

A number of signal analysis techniques are available: including time-domain, frequency-domain, wavelet analysis. Linear analysis (time-domain and frequency-domain analysis) techniques might be insufficient approximation of the nonlinear signals (Guo et al. 2010). Since EEG signal has non-linear and non-stationary characteristic, wavelet analysis techniques (DWT and SWT) are more appropriate and reliable for EEG signal analysis. The two wavelet techniques differ in their way of decomposition. In decimation SWT modify the filter coefficients in each filtration level by padding zeros. In DWT after filtration in each stage, coefficients are destroyed by a factor. Due to decimation of coefficients, DWT fails to preserve translational invariance. In both cases, the maximum decomposition level is set to 7 based on the decomposition level computing equation given by (Chen et al. 2015). Three commonly used wavelet families Coiflets (coif), Daubechies (DB), and Symlets (sym) were used for decomposing EEG signals. The DWT decomposition using DB wavelet family is shown in Fig. 3. The DWT decomposition using DB wavelet family is shown in Fig. 3. As shown in Fig. 3, in DWT the filtered output has half the length of the input signal whereas in SWT the filtered output has the same length, as shown in Fig. 4.

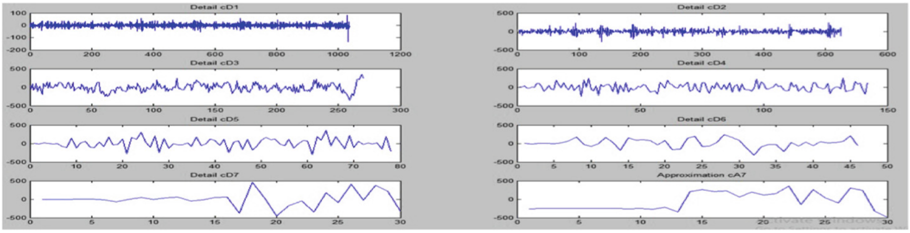


Fig. 3. Sample DWT DB8 detail coefficients of a sample EEG non-epileptic signal block.

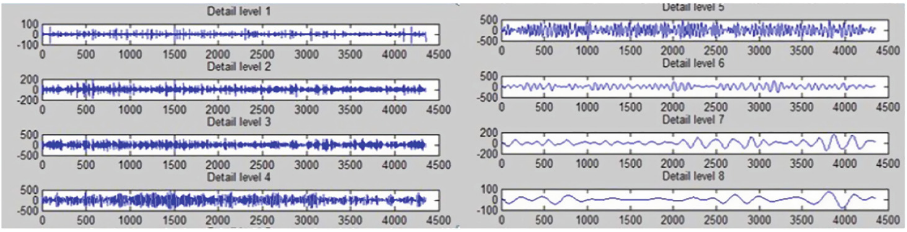


Fig. 4. Sample SWT DB8 detail coefficients of a sample EEG epileptic signal block.

2.3 Feature Extraction

In the signal processing, the low-frequency content (sub-bands) are the most representative part of the original signal. In the general human EEG rhythms (sub-bands) of the EEG signal, the clinical interest is more focused on: delta, theta, alpha, beta and gamma waves that are in the low frequency range (Gajic et al. 2014; Chen et al. 2017). Selecting low frequency sub-bands that best describe the behavior of EEG signals is important to improve seizure detection and classifier performance (Misiti et al. 1996). In this study, using DWT and SWT techniques the EEG raw signals were decomposed into the multi-level low-frequency resolutions.

A single decomposed EEG signal block is represented in the five sub-bands (D4–D7 and A7) frequency resolution spectrum. The 5 detailed low frequency sub-bands then used to extract features from each original raw signal. Each sub-band is divided into windows size called sub-block and each window size overlaps with the half of windows size. A single window size block is computed twice with its adjacent as shown on Fig. 5. Each of the five sub bands separated into 23 windows size and 46 sub-signal blocks. From each sub-block the statistical features, namely: maximum amplitude, minimum amplitude, average, mean and standard deviation are computed.

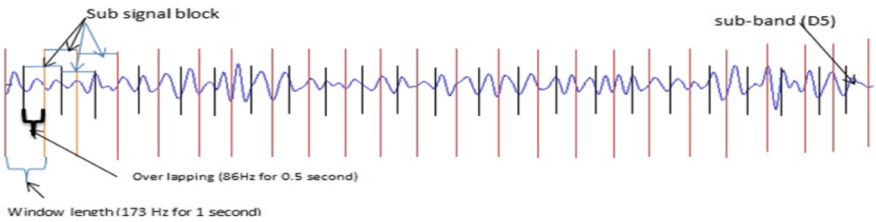


Fig. 5. A single sub-band divide in window size to form sub-signal block

2.4 Classifying Seizures

The ultimate goal of EEG signal detection is classification of signals based on predefined classes. SVM is a supervised machine learning approach that is widely used for data analysis and pattern recognition. Support Vector Machines (SVM) is the most common used supervised learning algorithms to separate EEG signal into epileptic and non-epileptic (Chen et al. 2015, 2017).

In this study, SVM classifier with RBF Kernel function was used. The training matrix contains the epileptic and non-epileptic data sets along with the class labels (class = 1 for epileptic and class = 0 for non-epileptic). The resulting SVM trained classifier, as shown in Fig. 6, contains the optimized parameters that are used to classify the test dataset.

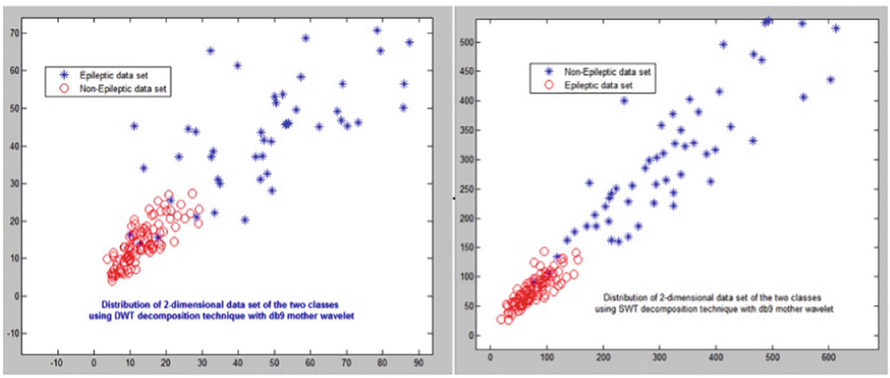


Fig. 6. Distribution of epileptic and non-epileptic signals using DWT and SWT techniques with DB9 wavelet function.

3 Results and Discussion

To compare the performance of the DWT and SWT decomposition techniques, training and testing datasets are randomly generated using 10-fold cross-validation. 200 EEG signals were randomly partitioned into 10 equal size subsamples. From the 10 subsamples, a single subsample (20 EEG signals) was taken as a testing data and the remaining k–1 subsamples (180 EEG datasets) are used as training data. The results from the folds then averaged (or otherwise combined) to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once. The test result of the DWT and SWT technique with DB9 mother wavelet function is shown on Fig. 7.

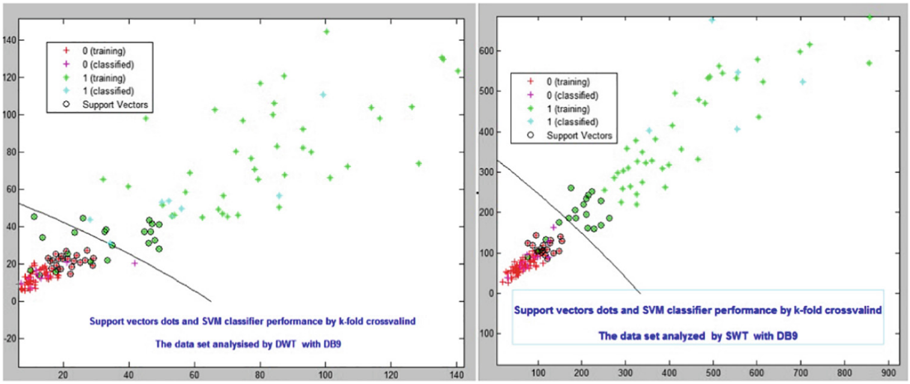


Fig. 7. Seizure prediction using DWT and SWT decomposition techniques

Based on common experimental parameters, the SVM classifier classified the EEG seizure dataset in the range 95% to 97.5% and 96% to 99.5% for DWT and SWT decomposition techniques, respectively. SWT performs better since instead of decimation it decomposes the EEG raw signal by filtering coefficients in each levels by padding them zeros. However, DWT does not implement the translation invariance, the decimation process in each level is eliminated. Due to decimation of coefficients, DWT fails to preserve translational invariance. Thus it suffers from lack of shift invariance. Lack of shift invariance assumes that any small translation in original signal can cause significantly large variations in the distribution of energy between wavelet coefficients. The average accuracy of DWT and SWT are 96.41% and 97.02%, respectively. SWT feature extraction technique scored the highest classification accuracy 99.5% with DB7 mother wavelet function Fig. 8.

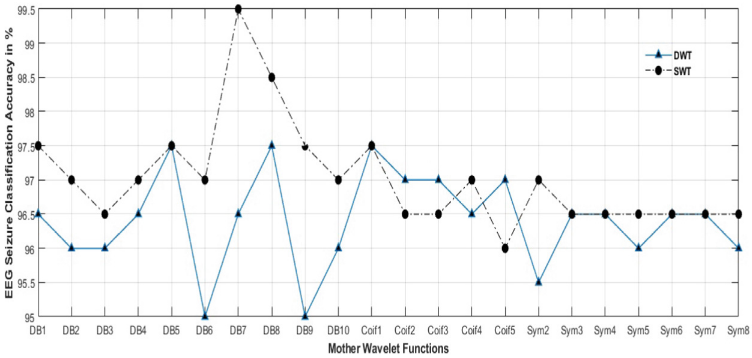


Fig. 8. DWT and SWT comparison

4 Conclusion

Visual detection of EEG signal is very time consuming and inefficient, especially in long-term recordings practitioners have to go through huge number of EEG data sheets. Due to these reasons, automated classification has been forwarded as valuable clinical tools. In this study we compared SWT and DWT decomposition techniques. The experimental result showed SWT mostly achieve high classification accuracy than DWT. The reason why this difference was existed: SWT instead of decimation it modifies the filter coefficients in each levels of stage by padding them zeros. But DWT on the other hand cannot performance translation invariance; hence, decimation process in each level is eliminated. Due to decimation of coefficients, DWT is not as good as SWT to preserve translational invariance.

References

Akareddy, S.M., Kulkarni, P.: EEG signal classification for epilepsy seizure detection using improved approximate entropy. Int. J. Public Health Sci. (IJPHS) **2**, 23–32 (2013)

Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E.: Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E **64**, 061907 (2001)

Chen, D., Wan, S., Bao, F.S.: EEG-based seizure detection using discrete wavelet transform through full-level decomposition. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1596–1602 (2015)

Chen, D., Wan, S., Xiang, J., Bao, F.S.: A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG. PLoS ONE **12**, e0173138 (2017)

Gajic, D., Djurovic, Z., Di Gennaro, S., Gustafsson, F.: Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. Biomed. Eng.: Appl. Basis Commun. **26**, 1450021 (2014)

Guo, L., Rivero, D., Pazos, A.: Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. J. Neurosci. Methods **193**, 156–163 (2010)

Juarez-Guerra, E., Alarcon-Aquino, V., Gomez-Gil, P.: Epilepsy seizure detection in EEG signals using wavelet transforms and neural networks. In: Elleithy, K., Sobh, T. (eds.) New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering. LNCS, vol. 312, pp. 261–269. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-319-06764-3_33

Kabir, E., Zhang, Y.: Epileptic seizure detection from EEG signals using logistic model trees. Brain Inform. **3**, 93–100 (2016)

Kalbhor, S.D., Harpale, V.K.: The review of detection and classification of epileptic seizures using wavelet transform. In: 2016 International Conference on Computing Communication Control and automation (ICCUBEA), pp. 1–5 (2016)

Mahdi, M.T.O.: A new fast epilepsy detection method using electroencephalogram signal processing. World Appl. Sci. J. **14**, 1119–1124 (2011)

Mallat, S.G.: A theory for multi-resolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989)

Orosco, L., Correa, A.G., Laciar, E.: a survey of performance and techniques for automatic epilepsy detection. J. Med. Biol. Eng. **33**, 526–537 (2013)

Pellegrino, G.: Analysis for automatic detection of epileptic seizure from EEG signals. Laurea Specialistica in Bioingegneria (2014). http://tesi.cab.unipd.it/46118/1/ThesisBio.pdf

Prince, P.G.K., Hemamalini, R.R., Kumar, S.: Epileptic seizure detection using EEG signals by means of stationary wavelet transforms. IJCTA, 291–329 (2016). https://healthdocbox.com/Epilepsy/69751917-Epileptic-seizure-detection-using-eeg-signals-by-means-of-stationary-wavelet-transforms.html

Radmehr, M., Anisheh, S.M.: EEG spike detection using stationary wavelet transform and time-varying autoregressive model. Int. J. Comput. Appl. **83**(13) (2013)

Shete, S., Shriram, R.: Comparison of sub-band decomposition and reconstruction of EEG signal by Daubechies9 and Symlet9 wavelet. In: 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT). IEEE (2014)

Shi, Y., Wang, S., Ying, J., Zhang, M., Liu, P., Zhang, H., et al.: Correlates of perceived stigma for people living with epilepsy: a meta-analysis. Epilepsy Behav. **70**, 198–203 (2017)

Shoeb, A.H.: Application of machine learning to epileptic seizure onset detection and treatment. Massachusetts Institute of Technology, 2 October 2009

Tibdewal, M.N., Dey, H.R., Mahadevappa, M., Ray, A.K., Malokar, M.: Wavelet transform based multiple features extraction for detection of epileptic/non-epileptic multichannel EEG. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1648–1653 (2016)

Tzallas, A.T., Tsipouras, M.G., Fotiadis, D.I.: Epileptic seizure detection in EEGs using time? Frequency analysis. IEEE Trans. Inf Technol. Biomed. **13**, 703–710 (2009)

Tzallas, A.T., et al.: Automated epileptic seizure detection methods: a review study. In: Epilepsy-Histological, Electroencephalographic and Psychological Aspects. InTech (2012)

Zandi, A.S., Javidan, M., Dumont, G.A., Tafreshi, R.: Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform. IEEE Trans. Biomed. Eng. **57**, 1639–1651 (2010)

Misiti, M., Misiti, Y., Oppenheim, G., Michel, J.P.: Wavelet toolbox: for use with MATLAB. citeulike.org (1996)

Classical Machine Learning Algorithms and Shallower Convolutional Neural Networks Towards Computationally Efficient and Accurate Classification of Malaria Parasites

Yaecob Girmay Gezahegn[1,3(✉)], Abel Kahsay Gebreslassie[1], Maarig Aregawi Hagos[1], Achim Ibenthal[2], and Eneyew Adugna Etsub[3]

[1] Mekelle Institute of Technology, Mekelle University, Mekelle, Ethiopia
yaecob.girmay@gmail.com, abelkahsay10@gmail.com, maarig2000@gmail.com

[2] HAWK University of Applied Sciences and Arts, Goettingen, Germany
achim.ibenthal@hawk.de

[3] Addis Ababa Instittute of Technology, Addis Ababa University, Addis Ababa, Ethiopia
eneyew_a@yahoo.com

Abstract. Malaria and its diagnosis methods need significant attention in Ethiopia. Studies show that for 46 laboratory professionals who were given 6 different positive and negative malaria slides, the detection error rates for plasmodium falciparum and vivax were 43.5% and 37%, respectively. Another similar study reports that the overall malaria diagnosis error rate is 40.4%. To circumvent these challenges, there needs to be a system that automatically and instantly analyzes and manipulates data with less bias so as to reduce misdiagnosis error, cost, workload of physicians and experts, and thereby improving the livelihood of the society. In this study, classical machine learning algorithms and shallow Convolutional Neural Networks (CNN) based classification for parasitized and uninfected malaria images are applied. While classical machine learning algorithms are requiring a feature design step for classification, CNNs learn to recognize objects end to end, implicitly without the need for feature engineering. Many machine learning algorithms such as K-Nearest Neighbors, Random Forest, Gradient Boosting, CNNs etc. are compared in this contribution. From our experimental observation, some of the classical models such as Gradient Boost, Random Forest and Stacked ensemble using color, texture and shape features achieved a classification accuracy of 96%, precision of 97%, recall of 97% and f1-score of 97% for the parasitized samples. A model ensemble of two custom CNNs and mini-VGGNet achieved a precision of 99%, recall of 96% and f1-score of 97% for the parasitized. These results of the CNN models are reproducible for each of them. All the approaches have comparable results, however, the classical machine learning models are resource efficient.

Keywords: Classical machine learning · CNN · Ensemble · Classification · Malaria · Automatic diagnosis

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 46–56, 2019.
https://doi.org/10.1007/978-3-030-26630-1_5

1 Introduction

Malaria is an acute infectious disease which is transmitted by Anopheles Mosquitos that carry plasmodium. It has five species which infect humans. Two of these species that are prevalent and have major impacts are plasmodium falciparum and plasmodium vivax. If a person is infected by falciparum, the disease progresses rapidly into cerebral malaria if not treated in time. Therefore, urgent diagnosis is a must [1].

According to a study by the World Health Organization (WHO), nearly half of the world's population is at risk due to malaria. The same study reports that globally there were 212 million cases of malaria and 429,000 deaths in 2015. More than 90% of the cases and deaths are in Africa [2, 3].

Convenient tropical climate for the growth of the parasites, inadequate health facilities, illiteracy, and poor socio-economic conditions made the fight against the disease hard [2]. Even though around 6.8 million lives have been saved from death since 2001, 70% of malaria illness and death happens in children below the age of 5 - still a major killer of children [4]. Hence, The Global Technical Strategy for Malaria has a goal to reduce malaria case incidence & mortality rate by 90% [5]. WHO also envisions to solve healthcare problems through innovative R&D with the theme, "Investing in knowledge for resilient health systems" [6].

A study in the Intensive Care Unit (ICU) of Jima University Specialized Hospital (Ethiopia) shows that the prescription error rate was found to be 52.5% [7]. Another study in the ICU of Tikur Anbessa Specialized Hospital (Addis Ababa, Ethiopia) shows that there is 40.7% of medication prescription error [8]. According to WHO, the main factors for diagnostic errors are low quality and scope of healthcare facilities, lack of competent and sufficient healthcare professionals, low teamwork and low learning and feedback among professionals when errors occur [8]. According to a study in the Tigray region of Ethiopia, out of 46 laboratory professionals who were given 6 different positive and negative malaria slides, their detection rates for plasmodium falciparum and vivax were 56.5% and 63%, respectively [9]. Another similar study reports that the overall malaria diagnosis error rate in Ethiopia is 40.4% and it varies in different parts of the country [10, 11]. Moreover, from our observation on day-to-day experiences of the society, and based on our visit to Tigray Health Research Institute (THRI), we have observed that the diagnosis error is high which is in line with the above reported cases. The main reasons that were identified in the health centers are, (i) few and poor medical facilities (laboratory set-up) with few skilled health professionals, (ii) lack of proper training, and (iii) all the laboratories use manual microscopy diagnosis. Microscopic diagnostics depends on the competencies and skills of laboratory technicians [12]. Manual diagnosis requires careful examination of red blood cells (RBCs) via microscope. A research in 2013 by Johns Hopkins University (USA) discloses that "diagnostic errors - not surgical mistakes or medication overdoses - accounted for the largest fraction of claims, the most severe patient harm, and the highest total of penalty payouts" [13].

To circumvent healthcare challenges, especially, in remote, resource constrained rural areas, automatic diagnosis method could solve the subjective and time consuming nature of manual diagnosis to more reliable and standardized diagnosis [14, 15]. In

[16], automated diagnosis using machine learning based medical image analysis has demonstrated more consistent and objective results compared to manual microscopy. A recent survey on algorithms used for malaria detection on images collected from thick blood smears is presented in [25]. In [26], multilayer perceptron neural networks is used for classification of malaria parasite. The authors of [27] used deep-learning based machine learning approach for detection of malaria from thick blood smear. However, this approach still did not address the problem with identification of the different species and their stages.

In this paper, different classical machine learning algorithms and convolutional neural networks have been employed to classify falciparum parasitized and non-parasitized thin blood smear images.

2 Feature Extraction, Classical Machine Learning Algorithms and Convolutional Neural Networks

Features such as color, texture and shape are widely used in order to efficiently characterize objects in images. Conventionally, there are different techniques which help us understand the semantics by appropriate representation of images. These methods are commonly known as feature extraction techniques, which however need to be designed by experts. As a result, performance of classifying, detecting or recognizing an object depends on how good the representation is and the representation in turn relies on how well the features are extracted. Following feature extraction, the next step is training using machine learning classifiers, in order to categorize the input images into their respective classes. Some of the classifiers include, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (logit), Decision Trees (DT), Naïve Bayes (NB) etc. These techniques have been widely used even with images that have complex features; for instance, SVMs can learn nonlinear classification boundaries. Even though conventional techniques have been widely used they have their own limitations [17].

To mitigate the limitations of the individual classifiers mentioned above, ensemble methods can be applied. Some of the ensemble techniques that are widely used are Random Forest (RF), AdaBoost (ABC) and Gradient Boost (GB). However, the approach is different from how we humans learn and perceive objects. So, deep-learning tries to bridge the gap, integrates extraction and classification as one module and learns intuitively as we humans do; learning via example and through experience [18]. The two important parameters, which make learning of objects as accurate as possible in CNNs, are large data sets for training purposes and a model with large learning capacity. CNNs employ prior knowledge learnt from neighboring pixels to compensate data which are missed [19]. Even though there are many techniques on natural images, the challenge of applying CNN techniques to the medical domain often lies in customizing existing architectures [20]. Figure 1 depicts a comparison of classical and CNN based classification of images.

By using classical machine learning algorithms, it is possible to detect and classify malaria infected blood smear images [1]. However, CNNs do end to end tasks without explicitly extracting features first, followed by classification [16, 17]. The main investigations reported subsequently can be categorized into (i) applying different classical techniques to extract features from infected and uninfected malaria datasets in combination with training via different classifiers, (ii) training CNN models on infected and uninfected malaria datasets and performance testing of classification results on a test set, and (iii) using ensemble methods for both classical machine learning algorithms and CNNs.

2.1 Ensemble Models

In ensemble methods, each model looks at slightly different features of the image to make its predictions, getting some fraction of the whole feature [21]. Decision of ensemble methods is based on gathering of information from multiple models, reducing generalization error, and improving overall performance while predicting. There are different types of ensemble learning. Here we have used AdaBoost, Gradient Boost, Stacking etc. for classical machine learning methods, and CNN based model ensembling for neural networks based machine learning techniques [22].

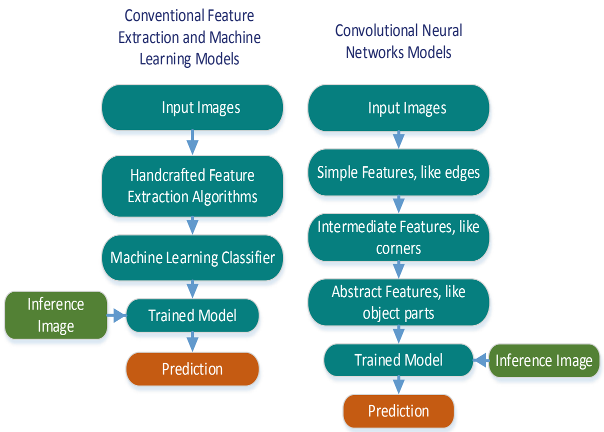


Fig. 1. Comparison of classical machine learning and CNN models.

Ensemble methods have two categories: sequential and parallel; in the first category we have AdaBoost that exploits the dependency between base learners, where the base learners are generated in sequence. In the second category, on the other hand, we have Random Forest, which exploits independence between base learners by generating parallel base learners.

Another Ensemble method is Bagging that stands for bootstrap aggregation which generates multiple versions of a predictor. One way to reduce the variance of an estimate is to take the average of multiple model estimates in case of regression and majority vote for classification. For the training of the base learners it uses bootstrap sampling [23].

Ensemble methods have numerous advantages. Most state-of-the-art academic benchmarks with a single model cannot be compared with these ensemble models. They are useful in weight transferring to other smaller models. However, increased training time and computational cost are disadvantages that should be noted [22].

3 Results, Discussion and Analysis

Different types of machine learning algorithms are applied as can be seen in the Tables 1, 2, 3, 4, 5, 6 and Figs. 2, 3, 4, 5 and 6. The two classes of malaria in the dataset (plasmodium falciparum infected and uninfected) used for training were taken from [24]. The classifiers are evaluated using accuracy, precession, recall and f1-score. However, using only accuracy as a performance measure does not always give correct result. Hence, to have a good knowledge about performance of a classifier used, the other mentioned metrics are also employed in tandem. Precision tells fraction of positive predictions that are actually positive whereas recall shows fraction of positive data predicted to be positive. If the recall is large, precision is small, few infected images are recognized, but includes no false positives (uninfected images). There is a tradeoff between precession and recall; high precision means low recall and vice versa. What is more, precision and recall are usually represented in a single number known as F-score (a harmonic mean of the two). In line with this, precise algorithms in this work, for example, found large positive (malaria images of falciparum) infected images. However, it includes few false positives as can be seen from the Tables and Figures. The confusion matrix shown in Fig. 4 clearly shows that 3% of the test data is false positive; the classifier confused 3% of the test data for non-parasitized (uninfected) with parasitized.

Tables 1 and 2 present results of the different algorithms for parasitized and uninfected cells respectively. Global feature extractors such as color channel statistics (mean and standard deviation) are used. Using just a color feature, Logistic Regression, Support Vector Machine, Random Forest, Gradient Boost, Linear Discriminant Analysis and Stacking Classifiers achieved an Accuracy, Precision, Recall and f1-score greater than 70%. Furthermore, Fig. 2 shows the performance of each classifier mentioned in Tables 1 and 2. As can be noted from the graphs, Gradient Boost, Random Forest and Stacking of classifiers got the highest classification accuracy. Different types of ensemble methods were also used in order to increase the performance of the classifiers as shown in Figs. 2, 3, 4 and Tables 1 and 4. Stacking ensemble is illustrated in Fig. 2. It consists of many of the classical machine learning algorithms mentioned in the table where kNN, RF, NB, GB, etc are base classifiers whose predictions are combined by logit. It can be depicted from Figs. 2, 3, 4, 5 and 6 that stacking achieves higher accuracy than individual classifiers.

Table 1. Classification result of different classical machine learning models using just the color feature for extraction of parasitized plasmodium falciparum.

	Type of classifier	Accuracy	Precision	Recall	f1-score
1	Naïve Bayes (NB)	0.63	0.63	0.61	0.62
2	Multinomial Naïve Bayes (MNV)	0.61	0.60	0.69	0.64
3	Bernoulli Naïve Bayes (BNB)	0.49	0.50	1.00	0.67
4	K-Nearest Neighbors (KNN)	0.66	0.67	0.65	0.66
5	Logistic Regression (Logit)	0.74	0.74	0.78	0.76
6	Decision Tree (DT)	0.67	0.67	0.68	0.67
7	Random Forest (RF)	0.71	0.73	0.70	0.72
8	Support Vector Machine (SVM)	0.74	0.75	0.77	0.76
9	Adaptive Boosting (ABC)	0.64	0.65	0.60	0.63
10	Gradient Boost (GB)	0.71	0.73	0.72	0.72
11	Linear Discriminant Analysis (LDA)	0.73	0.73	0.77	0.75
12	Stacking Classifiers	0.75	0.77	0.79	0.78

Table 2. Classification result of different classical machine learning models using just the color feature for extraction of uninfected RBC.

	Type of classifier	Accuracy	Precision	Recall	f1-score
1	Naïve Bayes (NB)	0.63	0.62	0.65	0.63
2	Multinomial Naïve Bayes (MNV)	0.61	0.63	0.54	0.58
3	Bernoulli Naïve Bayes (BNB)	0.49	0.00	0.00	0.00
4	K-Nearest Neighbors (KNN)	0.66	0.66	0.68	0.67
5	Logistic Regression (logit)	0.74	0.77	0.73	0.75
6	Decision Tree (DT)	0.67	0.67	0.67	0.67
7	Random Forest (RF)	0.71	0.71	0.74	0.73
8	Support Vector Machine (SVM)	0.74	0.77	0.74	0.76
9	AdaBoost (ABC)	0.64	0.63	0.68	0.65
10	Gradient Boost (GB)	0.71	0.72	0.73	0.73
11	Linear Discriminant Analysis (LDA)	0.73	0.76	0.71	0.73
12	Stacking Ensemble (SC)	0.75	0.78	0.76	0.77

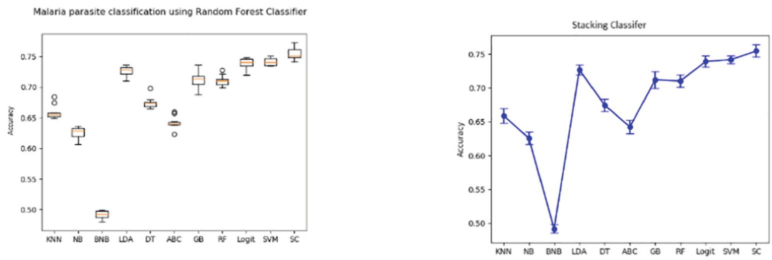


Fig. 2. Performance of different classifiers on the dataset (left), performance of stacked classifiers on the dataset (right) [using color features only].

Tables 3 and 4 present results of the different classical machine learning algorithms applied to classification of parasitized and uninfected cells along with ensemble methods, respectively, with more feature extraction techniques added. A combination of 'Hu-Moments' for shape extraction, color channels for color extraction and 'Haralick' for texture extraction were used. Furthermore, color channel statistics (mean, standard deviation and color histogram) were used to extract the color information of the images. As can be seen in the Tables and Figures, the additional features help improve the performance of the classifiers, even in the ensemble methods. The accuracy, precision, recall and f1-score jumped from around 70% to greater than 96% for the ensemble methods; an increment of greater than 20%. Figure 4, shows confusion matrix of random forest where the classifier is confused infected with non-infected malaria images for less than 3% of the test data while classifying.

Table 3. Classification result of different classical machine learning models using color, texture and shape features for extraction of parasitized plasmodium falciparum.

	Type of classifier	Accuracy	Precision	Recall	f1-score
1	Naïve Bayes (NB)	0.70	0.70	0.74	0.72
2	Bernoulli Naïve Bayes (BNB)	0.90	0.91	0.91	0.91
3	K-Nearest Neighbors (KNN)	0.58	0.58	0.57	0.58
4	Logit (Logistic Regression)	0.72	0.66	0.71	0.68
5	Decision Tree (DT)	0.95	0.95	0.97	0.96
6	Random Forest (RF)	0.96	0.97	0.97	0.97
7	Linear Discriminant Analysis (lda)	0.89	0.88	0.93	0.90
8	Support Vector Machine (svm)	0.75	0.75	0.80	0.77
9	Adaptive Boosting (AdaBoost, ABC)	0.92	0.91	0.97	0.94
10	Gradient Boost (GB)	0.96	0.97	0.97	0.97
11	Stacked Ensemble (SC)	0.96	0.97	0.97	0.97

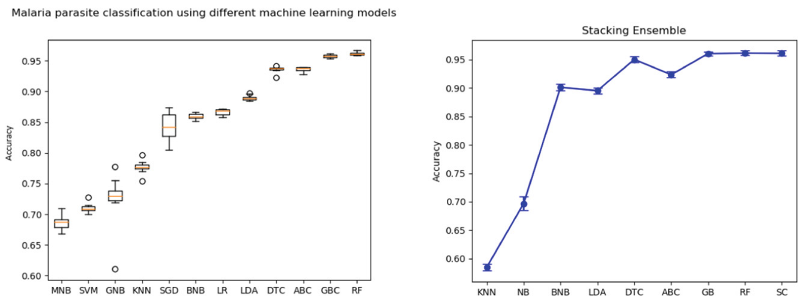


Fig. 3. Performance of different classifiers on the dataset (left), performance of stacked classifiers on the dataset (right) [using color, texture and shape features].

Table 4. Classification result of different classical machine learning models using color, texture and shape features for extraction of uninfected RBC

	Type of classifier	Accuracy	Precision	Recall	f1-score
1	Naïve Bayes (NB)	0.70	0.72	0.68	0.70
2	Bernoulli Naïve Bayes (BNB)	0.90	0.91	0.91	0.91
3	K-Nearest Neighbors (KNN)	0.58	0.58	0.59	0.59
4	Logistic Regression (logit)	0.72	0.69	0.63	0.66
5	Decision Tree (DTC)	0.95	0.97	0.95	0.96
6	Random Forest (RF)	0.96	0.97	0.97	0.97
7	Linear Discriminant Analysis (LDA)	0.89	0.92	0.87	0.90
8	Support Vector Machine (SVM)	0.75	0.79	0.73	0.76
9	Adaptive Boosting (AdaBoost, ABC)	0.92	0.97	0.90	0.93
10	Gradient Boost (GB)	0.96	0.97	0.97	0.97
11	Stacked Classifier (SC)	0.96	0.97	0.97	0.97

Tables 5 and 6, and Figs. 5 and 6, present customized convolutional neural networks, multilayer perceptron, mini-VGGNet and mini-GoogleNet. It is also depicted that ensemble methods were also used for the 3-layered, 4-layered CNN architectures and mini-VGGNet. We have ensembled three CNN architectures, Mini-VggNet, 3-layered and 4-layered customized CNN architectures. The training test split is 0.75/0.25 with total images of 27, 558.

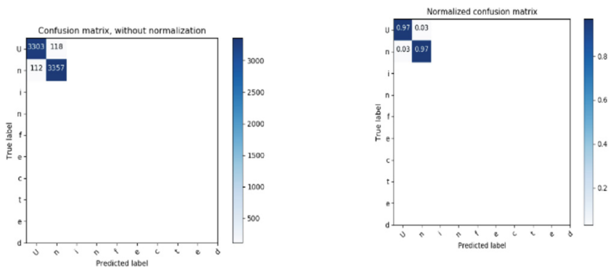


Fig. 4. Normalized and non-normalized confusion matrix for Random Forest classifier.

The overall performance is a little bit better than the classical feature extraction techniques. However, it is at the cost of computational time. Figures 6 and 7 depict the performance of our models during training and validation on the parasitized and uninfected malaria dataset. This is necessary because a model first needs to perform well in the training set in order to predict accurately on unseen images. Validation accuracy is necessary to make sure the models aren't overfitting. After completing training, performance for the two networks was evaluated on 6890 images, 3445 images from each class, that were held out for test time. The ensemble method

(Ensemble of 3 & 4 layer CNNs and mini-VGGNet) has highest performance with a precision, recall and f1-score of 0.99, 0.96 and 0.97, respectively, for the parasitized case, and 0.96, 0.99 and 0.97 for the uninfected cell images.

Table 5. Classification result of CNN models for parasitized plasmodium falciparum

	Type of classifier	Precision	Recall	f1-score
1	3-layered CNN	0.98	0.95	0.96
2	4-layered CNN	0.98	0.96	0.97
3	Multi-layer perceptron	0.84	0.49	0.62
4	Mini-GoogleNet (with 8-layers)	0.96	0.96	0.96
5	Mini-VGG (with 6-layers)	0.97	0.95	0.96
6	Ensemble of 3 and 4 layered CNNs	0.98	0.96	0.97
7	Ensemble of 3 and 4 layer CNNs and mini-VGGNet	0.99	0.96	0.97

Table 6. Classification result of CNN models for uninfected RBC

	Type of classifier	Precision	Recall	f1-score
1	3-layered CNN	0.95	0.98	0.96
2	4-layered CNN	0.96	0.98	0.97
3	Multi-layer perceptron	0.63	0.91	0.74
4	Mini-GoogleNet (with 8-layers)	0.96	0.96	0.96
5	Mini-VGG (with 6-layers)	0.95	0.97	0.96
6	Ensemble of 3 and 4 layered CNNs	0.96	0.98	0.97
7	Ensemble of 3 and 4 layer CNNs and mini-VGGNet	096	0.99	0.97

Fig. 5. Performance curves of 3-layers and 4-layers CNN models, from left to right.

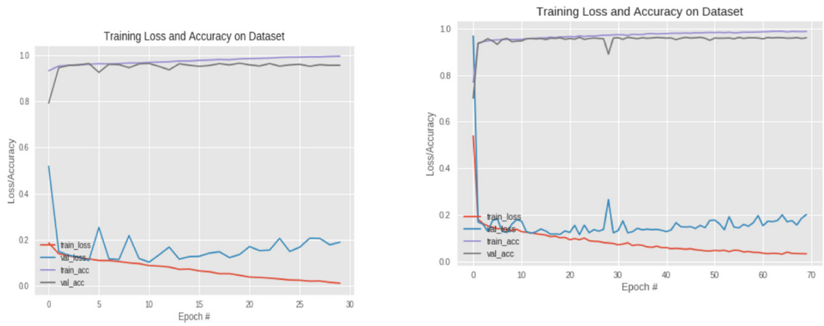


Fig. 6. Performance curves of Mini-GoogleNet and Mini-VGGNet, from left to right

4 Conclusion

The classification of parasitized and uninfected malaria was carried out using classical machine learning techniques and CNN architectures consisting of 3 and 4 layers, mini-VGGNet and mini GoogleNet. Ensemble methods have also been used for both, the classical machine learning and CNN models. The overall performance of the models, individually or using model ensemble, is good considering the number of layers used. What is more, the classical machine learning techniques have comparable result to the state-of-the-art CNN algorithms. The classical techniques don't require high performance computers or GPUs; they can easily be deployed into mobile devices. Furthermore, the shallow model with three CNN layers has comparable performance to the mini-VGG and mini-GoogleNet models. In order to get a better result, ensemble methods are employed in both cases, in the classical and in the CNN architectures. Some of the classical models such as Gradient Boost, Random Forest and Stacked ensemble using color, texture and shape features got a classification accuracy of 96%, precision of 97%, recall of 97% and f1-score of 97% for the parasitized. Similarly, model ensemble of two custom CNNs and mini-VGGNet got precision of 99%, recall of 96% and f1-score of 97% for the parasitized.

References

1. Poostchi, M., et al.: Image analysis and machine learning for detecting malaria. Transl. Res. **194**, 36–55 (2017)
2. Gallup, J., Sachs, J.: The economic burden of malaria. J. Trop. Med. **64**, 85–96 (2001)
3. Korenromp, E., et al.: World malaria report 2005, World Health Organization. Geneva, Technical report (2005)
4. http://www.who.int/mediacentre/factsheets/fs094/en/. Accessed 16 Sep 2017
5. Global Health Issues (2019). http://globalhealthstrategies.com/our-work/. Accessed 04 Jan 2019
6. World Health Organization: Investing in knowledge for resilient health systems: strategic plan 2016–2020 (2016)
7. Agalu, A., et al.: Medication prescribing errors in the intensive care unit of Jimma University specialized hospital, Southwest Ethiopia. J. Multidiscip. Healthc. **4**, 377–382 (2011)

8. Diagnostic Errors: Technical Series on Safer Primary Care. Geneva. World Heal Organization; Licence: CC BY-NC-SA 3.0 IGO (2016)
9. Megbaru, A., et al.: Performance of laboratory professionals working on malaria microscopy in Tigray, North Ethiopia. Hindawi J. Parasitol. Res. **2017**, 1–6 (2017). Article no. 9064917
10. Ayalew, F., Tilahun, B., Taye, B.: Performance evaluation of laboratory professionals on malaria microscopy in Hawassa Town, Southern Ethiopia. BMC Res. Notes **2014**, 1–8 (2014). Article no. 839
11. Yitbarek, T., Nega, D., Tasew, G., Taye, B., Desta, K.: Performance evaluation of malaria microscopists at defense health facilities in Addis Ababa and its surrounding areas. Ethiopia. PLoS ONE **11**(11), e0166170 (2016)
12. WHO. Malaria microscopy quality assurance manual - version 2. World Health Organization (2016)
13. David, E., Newman-Toker, M.D.: Diagnostic Errors More Common, Costly And Harmful Than Treatment Mistakes (2013). https://www.hopkinsmedicine.org/news/media/releases/diagnostic_errors_more_common_costly_and_harmful_than_treatment_mistakes. Accessed 07 Jan 2019
14. Reisenwitz, C.: How deep learning is changing healthcare part 1: diagnosis (2017). https://blog.capterra.com/how-deep-learning-is-changing-healthcare-part-1-diagnosis/. Accessed 04 Jan 2019
15. Bhatia, R.: Understanding The Deep Learning Advantage in Medical Imaging And How Tech Giants Are Rushing to Capitalize On It (2017). https://www.analyticsindiamag.com/understanding-deep-learning-advantage-medical-imaging-tech-giants-rushing-capitalize/. Accessed 05 Jan (2019)
16. Rajaraman, et al.: Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ **6**, e4568 (2018). https://doi.org/10.7717/peerj.4568
17. Razzak, M.I., Naz, S., Zaib, A.: Deep learning for medical image processing: overview, challenges and future. Computer Vision and Pattern Recognition, Cornell University, 22 April 2017
18. http://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/. Accessed online 07 Feb 2019
19. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 253–256 (2010)
20. Esteva, A., Kuprel, B., et al.: Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. Macmillan Publishers Limited, London (2017). Part of Springer Nature
21. Chollet, F.: Deep Learning with Python. Manning Publications Co., Shelter Island (2018)
22. https://blog.statsbot.co/ensemble-learning-d1dcd548e936. Accessed 07 Feb 2019
23. Breiman, L.: Bagging predictors. Department of Statistics University of California Berkeley, California 94720, Technical report no. 421, September 1994
24. https://ceb.nlm.nih.gov/repositories/malaria-datasets/. Accessed 10 Dec 2018
25. Jan, Z., Khan, A., Sajjad, M., et al.: A review on automated diagnosis of malaria parasite in microscopic blood smears images. Multimed. Tools Appl. **77**, 9801 (2018)
26. Días, G., González, F.A., Romero, E.: A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. J. Biomed. Inform. **42**(2), 296–307 (2009)
27. Quinn, J.A., et al: Deep convolutional neural networks for microscopy-based point of care diagnostics, arXiv: 1608.02989v1 [cs.CV], August 2016 and Sanchez, C.S.: Deep learning for identifying malaria parasites in images. Masters thesis, Artificial Intelligence School of Informatics, University of Edinburgh (2015)

Knowledge Based Decision Support System for Detecting and Diagnosis of Acute Abdomen Using Hybrid Approach

Alemu Workneh[1](✉), Dereje Teferi[2], and Alemu Kumilachew[3]

[1] Debre Tabor University, Debre Tabor, Ethiopia
a.workneh2005@gmail.com
[2] School of Information Science,
Addis Ababa University, Addis Ababa, Ethiopia
dereji.teferi@aau.edu.et
[3] Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia
alemupilatose@gmail.com

Abstract. Acute abdomen is one of the emergency diseases which is difficult for follow-up treatment. In developing countries, the mortality and morbidity rate are increasing due to misdiagnosis, delay, lack of knowledge and shortage of skilled manpower. These factors affect the quality of health care service in hospitals and reduce the quality of decisions made by physicians. This research attempts to investigate the applicability KBDSS using integration of rule based and case-based reasoning approach so as to improve the quality of decision made by domain experts, to provide effective and efficient services to the patients and to improve shortage of human expert in specific domain area. Domain knowledge is acquired using semi-structured interview technique. Domain experts are selected from Felege hiwot referral hospital in Bahir-Dar. In addition, secondary data is acquired from different sources following DSR methodology. The conceptual model of the knowledge-based system used a decision tree structure which is easy to understand and interpret the procedures involved in patient diagnoses. Based on the conceptual model, the prototype is developed with SWI prolog and java software tools. Beyond the domain expert, the result shows the system has accuracy of 99% for acute abdominal cause classification, and 66% for severity level identification. Whereas, the accuracy value of physicians has 84% for acute abdominal cause classification, and 50% for severity level identification with regard to the final result. Therefore, the KBDSS can diagnose patients with highest accuracy. We were taking the final observed value as pivoting point to test the KBDSS. In the evaluation of the system that classified the stored attribute according to the target problem is 71.33% accuracy. As the result shows, the three reasoning approaches: hybrid, case-based, and rule-based has an accuracy of 87.66%, 70%, and 60% respectively to retrieve target attributes for the target problem. This shows that hybrid reasoning approach is recommended to health care decision support system. Automation of KBDSS has a high contribution to establish truthful decision-making process in patient's acute abdomen treatment.

Keywords: Prolog · Java · Acute abdomen · KBDSS · Case-based · Rule-based · Hybrid

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 57–67, 2019.
https://doi.org/10.1007/978-3-030-26630-1_6

1 Introduction

Abdominal pain is the feeling of discomfort in abdominal cavity. The abdomen is the area surrounded by the lower margin of ribs, the upper part of diaphragm pelvic bone below and each side flanks. Pain arises from the tissues that surrounds the abdominal cavity which is muscles and skin. The term abdominal pain means sense of discomfort originating form organs in the abdominal cavity such us stomach, spleen, small intestine, large intestine, liver, gallbladder and pancreas (Jay et al. 2017). Abdominal pain is the current common problem which is chronic or acute (acute abdomen). Acute abdomen is the common emergency illness in developed and developing countries. Mortality and morbidity rate of the patients in developing countries due to acute abdomen is 18% and 20% respectively (Nyundo et al. 2013). This is because of misdiagnosis and delayed treatment. One of reasons for the prevalence of the problem is due to lack accurate computer aided supportive systems such as expert system, MYCIN, DSS and others are highly recommended to support the physician in making decisions. They support the physician early to know the disease and then decide the appropriate treatment immediately, since such disease are acute in nature and may lead to death or complications.

Artificial intelligence (AI) supports the human cognitive function. It is the paradigm shift in health care that increases the availability of patient data as well as it improves the diagnosis and treatment of the patient. For this purpose, it uses different algorisms and technique, such as knowledge-based reasoning techniques, NLP and ML (Fei et al. 2017).

A knowledge-based system (KBS) is a computer system which generates and utilizes knowledge from different sources. It supports experts in solving problems, especially complex ones, by utilizing artificial intelligence concepts and are mostly used in problem-solving procedures and to support human learning, decision making and actions. KBS is symbolic, heuristic, transparent and flexible. Knowledge-based systems are applicable in many areas such as health care, business (e-commerce), web information, agriculture (Leondes 2005). Knowledge based reasoning technique are the most common techniques for problem solving in artificial intelligence. Rule based reasoning, case-based reasoning, model-based reasoning and hybrid reasoning approaches are used for problem solving and decision making in different areas. The researchers used hybrid approach to minimize the drawback of individual reasoning approaches. The system was designed for health care to facilitates diagnosis and decision making in short time and with reduced cost, reduced mortality and morbidity rate using a knowledge base as recommended by (Zaraté and Shaofeng 2016).

Decision support system is a software which can replace the thinking process of a human and make logical decisions accordingly and could reduce the diagnostic errors as well as unwanted practice variation between physicians and improve the patient treatments in terms of time and cost (Prasath et al. 2013).

2 Problem Statement

Physicians are needed to provide the diagnostic knowledge to know the disease cause and provide appropriate measures such as extra diagnosis, further investigation and treatment of the disease. Most of the time, the traditional disease diagnosis takes extra time. Moreover, misdiagnosis happens due to different factors such as tiredness of the physician, similarity of cases and symptoms (Ray et al. 2016) (Ashley et al. 2013). This also forces the diagnosis to take more time. This length of time has a side-effect on patients especially for acute diseases. Most people think that the pain is like other chronic disease and take pain remedies, anti-acids, laxatives, or heating pads, which may complicate the problem. Therefore, acute abdomen needs special care, and require immediate diagnosis by a physician, otherwise it may lead to death. Studies conducted in Rwanda (Nyundo et al. 2013) revealed that the mortality and morbidity rate of patients in acute abdomen reach 18% and 20% respectively due to misdiagnosis and delayed treatment.

Furthermore, studies conducted in Nigeria (Adamu et al. 2010) on the waiting time of patients with abdominal pain in emergency department also asserted that 81.6% of the patients had surgical intervention more than 6 h later in developing countries. This waiting time is one hour in developed countries. The mean waiting time in Nigeria Ahmadu Bello University teaching hospital, were 22.3 $\pm$ 10.0 h with different factors. They conclude that the patient is waiting too long for emergency of acute abdomen, which resulted in increased mortality and morbidity rate of patients. In Ethiopia the mortality and morbidity rate were very high in case of acute abdomen. Regional health officials and health care providers discuss the way to reduce the acute abdomen mortality and morbidity rate and recommends good patient referral linkage and a quality surgical service in order to reduce its associated impacts and the case could be timely handled by training surgeons and physicians properly (Addisu et al. 2016). However, despite these efforts, the problem still persists. One of reasons for the prevalence of the problem is due to lack of accurate computer aided supportive systems. Expert systems such as MYCIN, DSS and others are highly recommended to support the physician in making better and faster decisions (Thomas et al. 2012). They support the physician to identify the disease early and then provide the necessary treatments immediately since such disease are acute in nature and may lead to death or complications unless treated in time.

The objective of this study is therefore to design and develop a Knowledge Based Decision Support System (KBDSS) for diagnosis and detection of acute abdomen.

To achieve this objective, the following activities are performed:

- Acquire tacit and explicit knowledge from domain experts for acute abdomen
- Analyze and model/design the KBDSS for AA using suitable technique, and
- Classify and identify the severity level of the acute abdominal pain

The output of this research:

- Benefit patients by providing easy and timely identification and diagnosis of acute abdomen which helps them to get quick treatment

- Helps physicians to apply better decision-making mechanism using information from the patients' history, ultrasound, X-ray, and white blood cells count.
- Reduces time and cost wasted due to misdiagnosis and unnecessary practice, and
- Reduces cost by using combinative knowledge, and increased.

3 Methodology of the Study

This research has developed a KBDSS for detecting and diagnosing of acute abdominal diseases. A design science research (DSR) methodology is followed in this research which is appropriate for design and development KBDSS. DSR is a research procedure for producing innovative construction intended to solve problems faced in the real world and to make a contribution to the theory of the discipline in which it is applied (Hamid 2014). The design science research methodology presented here incorporates principles, practices, and procedures required to carry out such research and meets three objectives: it is consistent with prior literature, it provides a nominal process model for doing DSR, and it provides a mental model for presenting and evaluating DSR (Ken and Tuure 2007).

3.1 Data Collection Method

In this research, both primary and secondary sources of knowledge are used to acquire the desired knowledge for the KBS. Primary knowledge is gathered from physicians and patients in Bahir Dar Felege-hiwot referral hospital using interview and observation. Secondary source of knowledge was collected by using document analysis. Acquiring knowledge from domain experts and professionals as well as organizing the collected knowledge in the knowledge base is one of the most important tasks of developing decision support system. One of the problems that professionals face is explaining their implicit knowledge. This makes knowledge acquisition the most difficult steps in the development of the model. Once the knowledge is gathered from the available sources, it is encoded in to the knowledge base.

3.2 Knowledge Representation

The knowledge acquired is represented using hybrid (integration of rule-based and case-based reasoning approaches) representation technique. Knowledge based system is a form of artificial intelligence that aims to capture the knowledge of human expert to support decision making. Knowledge base is the major component of the decision support system, since the quality of this unit determines how the user will deal with the system. The reason to use hybrid reasoning approach for the study was because it can alleviate the limitation of the single reasoning approaches and have better performance to represent knowledge for diagnosing acute abdomen.

3.3 Development Tools

Different tools are used for designing and developing KBDSS for acute abdomen.

Prolog: is a programming tool used in this study for designing and developing KBDSS. Prolog is a widely used programming language for artificial intelligence research. Prolog is used because it is easy to represent knowledge. As (Karde and Stefan 1986) stated, prolog programming answers the questions by logically following facts or rules.

Java: is the programming language used in this study to develop the interface. It was also used to provide a temporary storage for facts and cases in order to minimize the drawback of prolog programming language. Since patient history needs privacy and security, java programming language was recommended for developing the prototype of KBDSS (Tejinder and Jhunjhunu 2013).

MSQL: is used in the development of the prototype. Case-based reasoning approach needs a storage to store cases and other knowledge of domain experts. Therefore, MSQL database is used to develop case-library.

4 Model Development and Prototype Evaluation

Knowledge modeling is a cross disciplinary approach to capture and model knowledge of the domain experts into a knowledge-based system for purpose of preserving, improving, sharing, substituting, aggregating and diagnosing patients. Accordingly, in this research, it is used to simulate intelligence.

4.1 The Processes and Approaches of Decision-Making in KBDSS

The clinical decision support system passes different process stages in giving a decision to a problem. The processes include activities starting from collecting facts form domain experts about the given case to determination of final treatment. In healthcare system, different processes are applied before a final decision is reached for specific problem.

The aim of this research is developing a DSS for detecting and diagnosis of acute abdomen. Knowledge engineering research design was used to develop prototype system.

There are three common integration of rule-based and case-based reasoning approaches. These are rule dominant, case dominant and balanced approach.

4.1.1 Rule Dominant Integration

This is the way of integrating of rule-based and case-based reasoning in which rule-based reasoning is more applicable and case-based reasoning can be the supportive of rule-based reasoning.

4.1.2 Case Dominant Integration

Case-based reasoning is dominated to involve for solving a given problem and the rule-based reasoning can solve the problem when if case-based reasoning is not capable of solving a particular problem.

4.1.3 Balanced Approach

In this approach, the two-individual reasoning (rule-based and case-based) approaches have equal status for problem solving. The final result of individual reasoning's is combined to solve the target problem. The result that have been accessed from this type of reasoning approach is better than that can accessed by using the other two reasoning approach by using the same patient cases. According to (Jim and Ioannis 2009), balanced integration approach is best to be applied to health care. Since the two reasoning approaches have equally participated decision making. Due to this, balanced integration approach has been selected for the process of designing and developing an integrated approach for detecting and diagnosis of acute abdomen.

4.2 Architecture of the Prototype

Figure 1 shows the architecture of the proposed system. The inference engine is between the rule base and the working memory. Retrieving, reuse, revise and retain are included in case-based life cycle. Revising and retaining are done after integrating the two reasoning approaches.

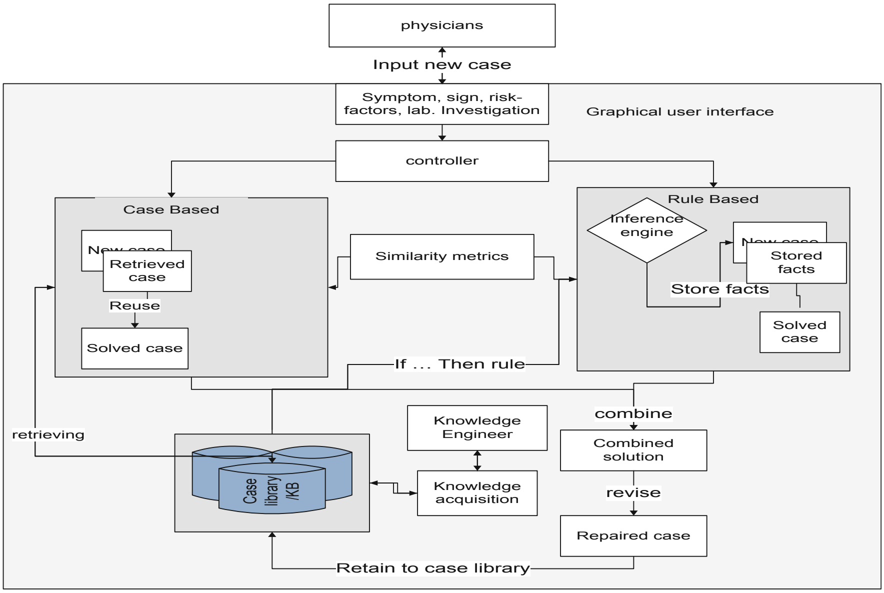


Fig. 1. Architecture of the proposed system

4.3 The Inference Engine

To achieve the goal of decision support system, inference engine helps to determine the order of testing the knowledge base rule, determine a Boolean (true, false uncertainty factor) application rules, and justifies user process of reasoning and generating the result in quantification form. The inference engine can calculate the probability of truth value and made decisions accordingly.

In prolog programming, one of the symptoms, sign, risk-factor or laboratory investigation is false, the inference engine draws, the investigation shows unknown result. It does not create approximate decision.

To solve this problem, temporary storage was created on java NetBeans 8.2 and retrieve facts which are stored in SWI-prolog 7.7.14 tools using prolog programming language to calculate the probability of the occurrence of truth value which helps to decide the severity level of the pain.

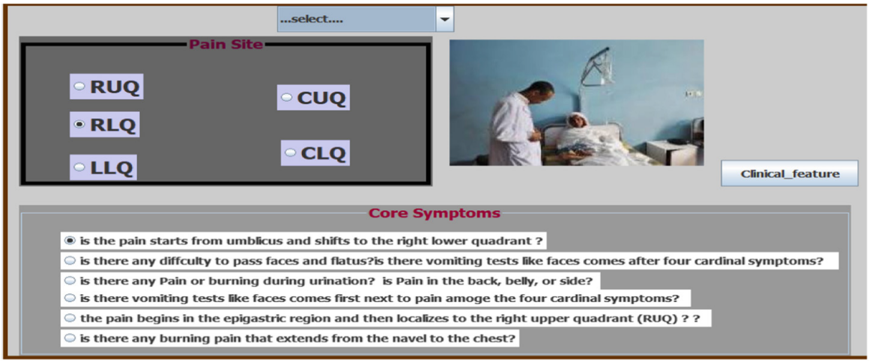


Fig. 2. Retrieving facts from fact base to temporary storage

Figure 2 shows, retrieving facts to store in java temporary storage. Which helps to improve effectiveness of the system. Since inference engine in prolog is difficult to infer from only few satisfied conditions. Therefore, in this system to infer a few satisfied conditions, first the system creates the temporary storage in java programming language. This can simplify the inference engine to draw the appropriate decision by calculating the truth value. After all, in this temporary storage, the system can verify the severity level of the pain using cosine similarity shown in Fig. 3.

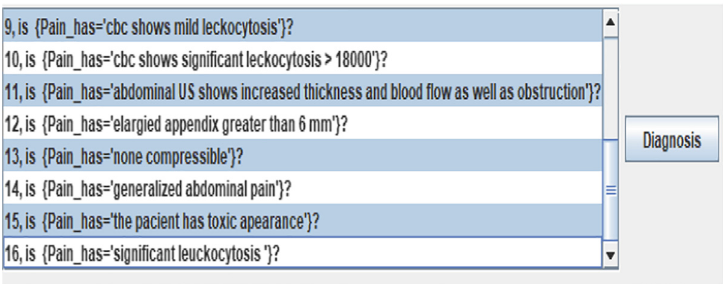


Fig. 3. Diagnosing patients using stored facts

Secondly the system uses case-based reasoning approach which compares the target problem with previously successfully diagnosed cases using cosine similarity metrics. If case-based response does not satisfy any of the condition, the inference engine gives a general recommendation as indicated below in Fig. 4.

Fig. 4. Recommendation wizard

4.4 Case Retrieval

Retrieving previous cases stored in the case library helps for the physicians to diagnose new similar cases. Before retrieving the stored case by physicians, there should be similarity comparison between the stored cases to be retrieved and the new case which helps to minimize the number of retrieved cases and minimize diagnostic analysis complexity by removing access of unimportant cases. There are two ways of retrieving the stored cases: by directly retrieving the stored case using the result of rule-based reasoning as an attribute and by using the site of pain for retrieval attribute.

Previously successful diagnosed cases for acute abdomen are retrieved in the order of accuracy of clinical investigation. Retrieval of the cases in case library by pain site is shown in Fig. 5.

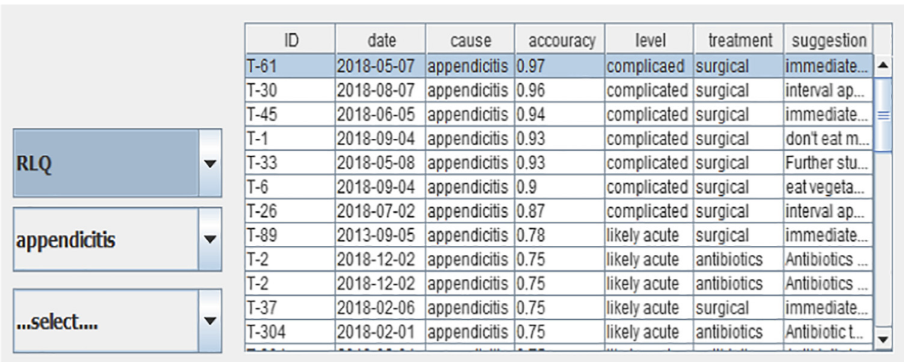


Fig. 5. Diagnosing patients using previous successfully diagnosed cases

4.5 Similarity Assessment

The similarity of each stored case to the target problem, typically represented as a real number in the range from 0 to 1, is computed using cosine similarity measure.

$$\text{Cos}\ \theta = \frac{A.B}{||A||||B||}$$

Where A is the previously diagnosed case or the facts and B is the new patient case.

$$\mathrm{A} = [\mathrm{a}_1, \mathrm{a}_2, \mathrm{a}_3, \ldots \mathrm{a}_\mathrm{n}]$$

$$\mathrm{B} = [\mathrm{b}_1, \mathrm{b}_2, \mathrm{b}_3, \ldots \mathrm{b}_\mathrm{n}]$$

$$\text{Cosine}(\mathrm{A},\ \mathrm{B}) = \frac{\sum_{i=1}^{n} aibi}{\sqrt{\sum_{i=1}^{n} ai}\sqrt{\sum_{i=1}^{n} bi}}$$

4.6 Cause Identification and Leveling Mechanism

In clinical decision-making, cause identification is an initial condition to take appropriate action for the particular cause of acute abdomen.

In this research, cause of acute abdomen was identified by using pain site (abdominal quadrants) and core symptoms. Once the quadrant is known, the pain can be identified using the cosine similarity value. To identify the severity level of the pain, threshold value was put as a pivoting point. Angle of difference between the stored knowledge and the target problem is computed to identify the cases that are very similar, similar and less similar. If the similarity value is greater than cosine of 30 then the pain is complicated, if the similarity value is between cosine of 45 and 30 the problem is likely acute whereas if the similarity is less than cosine of 45 the problem is less likely acute. The threshold value was set by discussion with domain experts.

4.7 Integrating Approach of Rule-Based and Case-Based Reasoning

Integration approach increases the confidence of physicians to diagnose acute abdomen. First the physician diagnoses the patients using rule-based and case-based reasoning and draws appropriate decision in parallel. Then the results of the two decisions can be integrated. For this integration, the mean value of the cosine similarity values of the two reasoning approaches is taken.

5 Conclusion and Recommendation

5.1 Conclusion

Acute abdomen is a type of sudden abdominal pain. In Ethiopia, 90% of surgical emergency is acute abdomen. It needed immediate, accurate and appropriate treatment.

Otherwise, it causes complication or death. KBDSS help physicians in decision making when diagnosing the patients within a short period of time. Knowledge sharing for domain experts in diagnosing patient is the main task of the system. The performance of the system shows promising acceptance by the system evaluators.

According to the system evaluators, the accuracy of the system 99% for acute abdominal cause classification and 66% for severity level identification whereas the accuracy value of physicians with related to the final result were 84% abdominal cause classification and 50% for severity level identification. Therefore, the KBDSS can diagnose patients with highest accuracy. We were taking the final observed value as pivoting point to test the KBDSS. The average result of each reasoning approach from the evaluator or domain experts, rule-based was 60% accuracy, case-based reasoning approach was 70% accuracy and hybrid reasoning approach were 87.66% accuracy. Therefore, hybrid approach was recommended for health care DSS with having the highest accuracy value compared with rule-based and case-based reasoning approaches. The average value of the system classifies attribute which stored in case library and fact base according to the target problem were 71.33% of accuracy. The evaluators and six patients were selected purposively. From this, the system can minimize delay during diagnosis, waste of resources for extra diagnosis and increase knowledge sharing between the domain experts. In addition to this automation of KBDSS has a high contribution for establishment of truthful decision making in acute abdomen patient's treatment.

5.2 Recommendation

Developing a decision support system for clinical diagnosis process is important to facilitate the process of diagnosis of a patient.

Acute abdomen is an emergency disease that cannot give the chance of follow up diagnosis and treatment. Therefore, it is recommended to use this developed KBDSS for diagnosing acute abdomen in health care system.

Integration of different reasoning approach reduces the limitation of individual reasoning. Therefore, it is recommended the researchers use the integration of more than one reasoning approaches and ML algorisms to study knowledge-based systems.

References

Adamu, A., Maigatari, M., Lawal, K., Iliyasu, M.: Waiting time for emergency abdominal surgery in Zaria, Nigeria. Afr. Health Sci. **10**(1), 46–53 (2010). https://www.researchgate.net/publication/46111242

Addisu, M., Tadess, A., Eyobe, T.: Pattern of acute abdomen in Dil Chora referral hospital, Eastern Ethiopia. Int. J. Collab. Res. Internal Med. Publ. Health **8**(11), 607–615 (2016)

Ashley, H., Bennet, B., Marie, C.: The evaluation of the acute abdomen. In: Moore, L., Todd, S. (eds.) Common Problems in Acute Care Surgery, pp. 19–31. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-6123-4_2

Fei, J., et al.: Artificial intelligence in healthcare: past, present and future. Stroke Vasc. **2**, 1–14 (2017)

Hamid, M.: Design science research as an approach to develop conceptual solutions for improving cost management in construction (2014)

Jay, W.M., Charles, P.D., Anand, B.S.: Abdominal Pain (causes, remedies, treatment). (Medicine Net.com) (2017). https://www.medicinenet.com/abdominal_pain_causes_remedies_treatment/article.htm. Accessed 24 Nov 2017

Jim, P., Ioannis, H.: Integration of rule-based and case based reasoning. Hybrid Intell. Syst. **6**(4), 189–209 (2009)

Karde, E., Stefan, J.: Prolog Programming in Artificial Intelligence. Addison-Wesley Publishing Company, Boston (1986)

Ken, P., Tuure, T.: A design science research methodology for information system researches. J. Manag. Inf. Syst. **24**, 45–77 (2007)

Leondes, C.T.: Intelligent Knowledge Based Systems. University of California, Los Angeles, Kluwer Academic Publishers, Los Angeles (2005)

Nyundo, M., Rugwizangoga, E., NtakiyirutaI, G., Kakande, I.: Outcome of emergency abdominal surgery at Kigali university teaching hospital: a review of 229 cases. East Cent. Afr. J. Surg. **18**(1), 31–39 (2013). Retrieved 27 Nov 2017

Prasath, V., Lakshmi, N., Nathiya, M., Bharathan, N., Neetha, P.: A survey on the applications of fuzzy logic in medical diagnosis. Int. J. Sci. Eng. Res. **4**(4), 1199–1203 (2013). http://www.ijser.org

Ray, S., Patel, M., Parmar, H.: Management of acute abdomen: study of 110 cases, **3**(2), 18–24 (2016). http://iaimjournal.com/. Accessed 24 Nov 2017

Tejinder, S., Jhunjhunu, R.: New software development methodology for student of Java programming language. Int. J. Comput. Commun. Eng. **2**(2), 194–196 (2013)

Thomas, M., Fabrizio, C., Shantale, C.: Decision support systems in medicine - anesthesia, critical care and intensive care medicine. McGill University, Intact (2012)

Zaraté, P., Shaofeng, L.: A new trend for knowledge-based decision support system design. Int. J. Inf. Decis. Sci. **8**(3), 305–324 (2016)

Modern CNNs for IoT Based Farms

Patrick Kinyua Gikunda[1,2(✉)] and Nicolas Jouandeau[2]

[1] Department of Information Technology,
Dedan Kimathi University of Technology, Nyeri, Kenya
patrick.gikunda@dkut.ac.ke

[2] Le Laboratoire d'Informatique Avancée de Saint-Denis (LIASD),
University of Paris 8, Saint-Denis, France
n@ai.univ-paris8.fr

Abstract. Recent introduction of ICT in agriculture has brought a number of changes in the way farming is done. This means use of Internet of Things(IoT), Cloud Computing(CC), Big Data (BD) and automation to gain better control over the process of farming. As the use of these technologies in farms has grown exponentially with massive data production, there is need to develop and use state-of-the-art tools in order to gain more insight from the data within reasonable time. In this paper, we present an initial understanding of Convolutional Neural Network (CNN), the recent architectures of state-of-the-art CNN and their underlying complexities. Then we propose a classification taxonomy tailored for agricultural application of CNN. Finally, we present a comprehensive review of research dedicated to applications of state-of-the-art CNNs in agricultural production systems. Our contribution is in two-fold. First, for end users of agricultural deep learning tools, our benchmarking finding can serve as a guide to selecting appropriate architecture to use. Second, for agricultural software developers of deep learning tools, our in-depth analysis explains the state-of-the-art CNN complexities and points out possible future directions to further optimize the running performance.

Keywords: Convolutional Neural Network · Farming · Internet of Things

1 Introduction

The global population is set to touch 9.6 billion mark by year 2050 [4]. The continuous population growth means increase in demand for food to feed the population [5]. Agriculture is the practice of cultivation of land and breeding of animals & plants to provide food and other products in order to sustain and enhance life [6]. Due to the extreme weather conditions, rising climate change and environmental impact resulting from intensive farming practices [8], farmers are now forced to change their farming practices. To cope with the new farming challenges, farmers are forced to practice smart farming [9], which offers solutions of farming management and environment management for better production.

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 68–79, 2019.
https://doi.org/10.1007/978-3-030-26630-1_7

Smart farming focuses on the use of information and communication technology (ICT) in the cyber-physical farm management cycle for efficient farming [10].

Current ICT technologies relevant for use in smart farming include IoT [11], remote sensing [12], CC [13] and BD [14]. Remote sensing is the science of gathering information about objects or areas from a distance without having physical contact with objects or areas being investigated. Data collected through remote sensing and distributed devices is managed by cloud computing technology, which offers the tools for pre-processing and modelling of huge amounts of data coming from various heterogeneous sources [15]. These four technologies could create applications to provide solutions to todays's agricultural challenges. The solutions include real time analytics required to carry out agile actions especially in case of suddenly changed operational or environmental condition (e.g. weather or disease alert). The continuous monitoring, measuring, storing and analysing of various physical aspects has led to a phenomena of big data [16]. To get insight for practical action from this large type of data requires tools and methods that can process multidimensional data from different sources while leveraging on the processing time.

One of the successful data processing tool applied in this kind of large dataset is the biologically inspired Convolutional Neural Networks (CNNs), which have achieved state-of-the-art results [17] in computer vision [18] and data mining [19]. As deep learning has been successfully applied in various domains, it has recently entered in the domain of agriculture [10]. CNN is a subset method of Deep Learning (DL) [20], defined as deep, feed-forward Artificial Neural Network (ANN) [21]. The CNN covolutions allow data representations in a hierarchical way [22]. The common characteristics of CNN models is that they follow the same general design principles of successive applying convolutional layers to the input, periodically downsampling the spatial dimensions while increasing the number of feature maps. These architectures serve as rich feature extractors which can be used for image classification, object detection, image segmentation and many more other advanced tasks. This study investigates the agricultural problems that employ the major state-of-the-art CNN architectures that have participated in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [23] with highest accuracy in a multi-class classification problem. ImageNet [26] classification challenge has played a critical role in advancing the CNN state-of-the-art [17]. The motivation for carrying out the study include: (a) CNNs has better precision compared to other popular image-processing techniques in the large majority of problems [27]. (b) CNN has entered in the agricultural domain with promising potential [28]. (c) all the CNN models that have achieved the top-5 error are successful when applied in other computer vision domain with remarkable results [27]. This review aims to provide insight on use of state-of-the-art CNN models in relation to smart farming and to identify smart farming research and development challenges related to computer vision. Therefore the analysis will primarily focus on the success on use of state-of-the-art CNN models in smart farms with a intention to provide this relevant information to the future researchers. From that perspective the research questions to be addressed in the study are:

(a) What is the role of CNN in smart farming?
(b) What type of the state-of-the-art CNN architecture should be used?
(c) What are the benefits of using state-of-the-art CNN in IoT based agricultural systems?

The rest of the paper is organized as follows: in Sect. 2, we present an overview of existing state-of-the-art CNNs architectures including their recent updates. Then we propose a taxonomy to provide a systematic classification of agricultural issues for CNN application. In Sect. 3, we present the existing state-of-the-art CNNs and their scope of application in agri-systems. We conclude the paper in Sect. 4.

2 Methodology

In order to address the research questions a bibliographic analysis in the domain under study was done between 2012 and 2018, it involved two steps: (a) collection of related works and, (b) detailed review and analysis of the works. The choice of the period is from the fact that CNN is rather a recent phenomenon. In the first step, a keyword-based search using all combinations of two groups of keywords of which the first group addresses CNN models (LeNet, AlexNet, NIN, ENet, GoogLeNet, ResNet, DenseNet, VGG, Inception) and the second group refers to farming (i.e. agriculture, farming, smart farming). The analysis was done while considering the following research questions: (a) smart farm problem they addressed, (b) dataset used, (c) accuracy based on author's performance metric, (d) state-of-the-art CNN model used. Its important to note that use of state-of-the-art deep learning has great potential, and there have been recent small comparative studies to analyse and compare the most efficient archtecture to use in agricultural systems. They include: Comparison between LeNet, AlexNet, and VGGNet on automatic identification of center pivot irrigation [29] and comparison between VGG-16, Inception-V4, Resnet and DenseNet for plant disease identification [30].

3 State-of-the-Art CNN: An Overview

CNNs typically perform best when they are large, meaning when they have more deeper and highly interconnected layers [31]. The primary drawback of these architectures is the computational cost, thus large CNNs are typically impractically slow especially for embedded IoT devices [32]. There are recent research efforts on how to reduce the computation cost of deep learning networks for everyday application while maintaining the prediction accuracy [33]. In order to understand the application of the state-of-the-art CNN architectures in agricultural systems, we reviewed the accuracy and computational requirements from relevant literature including recent updates of networks as shown in

Fig. 1. The classical state-of-the-art deep network architectures include; LeNet [34], AlexNet [35], NIN [36], ENet [37], ZFNet [38], GoogleLeNet [39] and VGG 16 [40]. Modern architectures include; Inception [41], ResNet [42], and DenseNet [43].

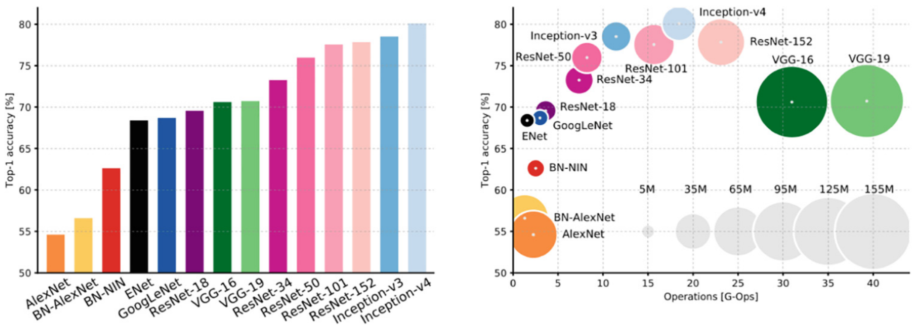


Fig. 1. Top-1 accuracy vs the computational cost. The size of the circles is proportional to number of parameters. Legend; the grey circles at the bottom right represents number of parameters in millions [32].

LeNet-5 is a 7-layer pioneer convolutional network by LeCun et al. [34] to classify digits, used to recognise hand-written numbers digitized in 32×32 pixel greyscale input images. High resolution images require more convolutional layers, so the model is constrained by the availability of the computing resources.

AlexNet is a 5-layer network similar to LeNet-5 but with more filters [35]. It outperformed Lenet-5 and won the LSVRC challenge by reducing the top-5 error from 26.2% to 15.3%. Use Rectified Linear Unit (Relu) [1] instead of Hyperbolic Tangent (Tanh) [2] to add non-linearity and accelerates the speed by 6 times. Droupout was employed to reduce over-fitting in the fully-connected layers. Overlap pooling was used to reduce the size of the network while reducing top-1 error by 0.4% and top-5 error by 0.3%.

Lin et al. [36] created a Network in Network (NIN) which inspired the inception architecture of googlenet. In their paper, they replaced the linear filters with nonlinear multi linear perceptrons that had better feature extraction and accuracy. They also replaced the fully connected layers with activation maps and global average pooling. This move helped reduce the parameters and network complexity.

In their article Paszke et al. [37] introduced an Efficient Neural Network (ENet) for running on low-power mobile devices while achieving state-of-the-art results. ENet architecture is largely based on ResNets. The structure has one master and several branches that separate from the master but concatenate back.

In their work Zeiler and Fergus [38] created ZFNet which won a ILSVRC 2013 [25] image classification. It was able to achieve a top-5 rate of 14.8% an improvement of the AlexNet. They were able to do this by tweaking the hyper-parameters of AlexNet while maintaining the same structure with additional deep learning elements. There is no record observed of use of ZFNet in agricultural systems despite the accuracy improvement. Each branch consists of three convolutional layers. The 'first' 1×1 projection reduces the dimensionality while the latter 1×1 projection expands the dimensionality. In between these convolutions, a regular (no annotation/asymmetric X), dilated (dilated X) or full convolution (no annotation) takes place. Batch normalization [60] and PReLU [61] are placed between all convolutions. As regularizer in the bottleneck, Spatial Dropout is used. MaxPooling on the master is added only when the bottleneck is downsampling which is true.

GoogleNet, a 2014 ILSVRC image classification winner, was inspired by LeNet but implemented a novel inception module. The Inception cell performs series of convolutions at different scales and subsequently aggregate the results. This module is based on several very small convolutions in order to drastically reduce the number of parameters. There has been tremendous efforts done to improve the performance of the architecture: (a) Inception v1 [39] which performs convolution on an input, with 3 different sizes of filters (1×1, 3×3, 5×5). Additionally, max pooling is also performed. The outputs are concatenated and sent to the next inception module. (b) Inception v2 and Inception v3 [41] factorize 5×5 convolution to two 3×3 convolution operations to improve computational speed. Although this may seem counterintuitive, a 5×5 convolution is 2.78 times more expensive than a 3×3 convolution. So stacking two 3×3 convolutions infact leads to a boost in performance. (c) In Inception v4 and Inception-ResNet [44] the initial set of operations were modified before introducing the Inception blocks.

Simonyan and Zisserman created VGGNet while doing investigation on the effect of convolutional network depth on its accuracy in the large-scale image recognition setting. The VGGNet took the second place after GoogLeNet in the competition. The model is made up of 16 convolutional layers which is similar to [35] but with many filters. There have been a number of update to the VGGNet archtecture starting with pioneer VGG-11(11 layers) which obtained 10.4% error rate [40]. VGG-13 (13 layers) obtains 9.9% error rate, which means the additional convolutional layers helps the classification accuracy. VGG-16(16 layers) obtained a 9.4% error rate, which means the additional $3 \times 1 \times 1$ conv layers help the classification accuracy. 1×1 convolution helps increase non-linearity of the decision function, without changing the dimensions of input and output, 1×1 convolution is able to do the projection mapping in the same high dimensionality. This approach is used in NIN [36] GoogLeNet [39] and ResNet [42]. After updating to VGG-16 it obtained 8.8% error rate which means the deep learning network was still improving by adding number of layers. VGG-19 (19 layers) was developed to further improve the performance but it obtained 9.0% showing no improvement even after adding more layers.

When deeper networks starts converging, a degradation problem is exposed: with the network depth increasing, accuracy gets saturated and then degrades rapidly. Deep Residual Neural Network(ResNet) created by He al. [42] introduced a norvel architecture with insert shortcut connections which turn the network into its counterpart residual version. This was a breakthrough which enabled the development of much deeper networks. The residual function is a refinement step in which the network learn how to adjust the input feature map for higher quality features. Following this intuition, the network residual block was refined and proposed a pre-activation variant of residual block [45], in which the gradients can flow through the shortcut connections to any other earlier layer unimpeded. Each ResNet block is either 2 layer deep (used in small networks like ResNet 18, 34) or 3 layer deep (ResNet 50, 101, 152). This technique is able to train a network with 152 layers while still having lower complexity than VGGNet. It achieves a top-5 error rate of 3.57% which beats human-level performance on this dataset. Although the original ResNet paper focused on creating a network architecture to enable deeper structures by alleviating the degradation problem, other researchers have since pointed out that increasing the network's width (channel depth) can be a more efficient way of expanding the overall capacity of the network.

In DenseNet which is a logical extension of ResNet, there is improved efficiency by concatenating each layer feature map to every successive layer within a dense block [43]. This allows later layers within the network to directly leverage the features from earlier layers, encouraging feature reuse within the network. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers, this helps alleviate the vanishing-gradient problem, feature reuse and reduce number of parameters.

3.1 Proposed Agricultural Issues Classification Taxonomy

Many agricultural CNN solutions have been developed depending on specific agriculture issues. For the study purpose, a classification taxonomy tailored to CNN application in the smart farming was developed as shown Fig. 2. In this section, we categorize use of state-of-the-art CNN based on the agricultural issue they solve:

(a) Plant management includes solutions geared towards crop welfare and production. This includes classification (species), detection (disease and pest) and prediction (yield production).
(b) Livestock management address solutions for livestock production (prediction and quality management) and animal welfare (animal identification, species detection and disease and pest control).
(c) Environment management addresses solutions for land and water management.

3.2 Use of State-of-the-Art CNN in Smart Farms

The Table 1, shows use of state-of-the-art CNN in agriculture and in particular the areas of plant and leaf disease detection, animal face identification, plant recognition, land cover classification, fruit counting and identification of weeds.

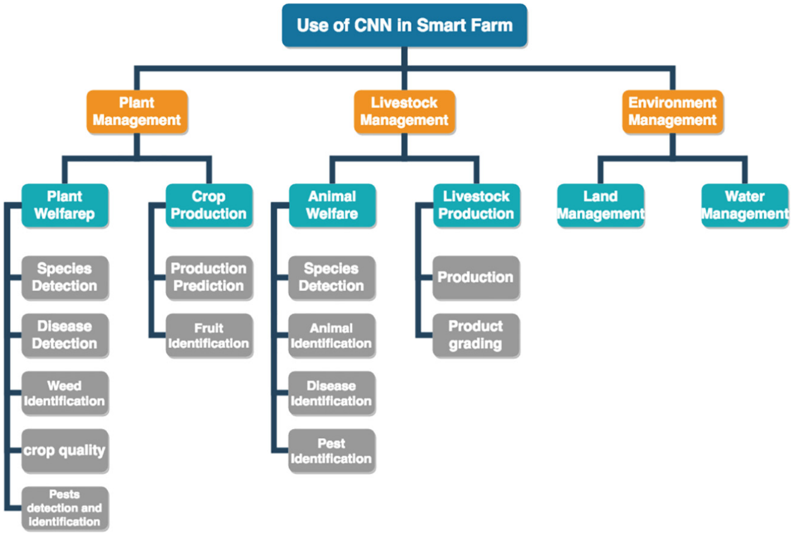


Fig. 2. Proposed classification taxonomy for CNN use in smart farm

It consist of 5 columns to show: the problem description, size of data used, accuracy according to the metrics used, the state-of-the-art CNN used and reference literature.

In their paper, Amara et al. [54] use the LeNet architecture to classify the banana leaves diseases. The model was able to effectively classify the leaves after several experiments. The approach was able to classify leaves images with different illumination, complex background, resolution, size, pose, and orientation. We also reviewed use of CaffeNet archtecture [59] in agricultural application, which is a 1-GPU version of AlexNet. The success of this model at LSVRC 2012 [24] encourage many computer vision community to explore more on the application of deep learning in computer vision. Mohanty et al. [7] combined both AlexNet and GoogLeNet to identify 14 crop species and 26 diseases(or absence thereof) from a dataset of 54,305 images. The approach records an impressive accuracy of 99.35% demonstrating the feasibility of the state-of-the-art CNN architectures. Other areas AlexNet has been used with high accuracy record include; identify plants using different plant views [49], identify plant species [48], identify obstacles in the farm [50] and leaf disease detection [3]. Because of its achievement to improve utilization of the computing resources GoogLeNet has been used in fruit count [53] and plant species classification [7]. VGGNet has been used in classifying weed [56], detect obstacles in the farm [51], fruit detection [47] and animal face recognition [55]. Like ResNet, DenseNet is a recent model that explains why it has not been employed significantly in farming, nevertheless it has been used in thistle identification in winter wheat and spring barley [52]. Since ResNet is a such a recent model, it have only been used by one author in fruit counting [53]. Many of the CNN developed for agricultural use depend on the problem or challenge they solve.

Table 1. Use of state-of-the-art CNN in Smart Farm

No.	Smartfarm problem description	Data used	Accuracy	CNN framework used	Article
1	Fruit detection	Images of three fruit varieties: apples (726), almonds (385) and mangoes (1154)	F1 (precision score) of 0.904 (apples) 0.908 (mango) 0.775 (almonds)	VGGNet	[46]
2	Detection of sweet pepper and rock melon fruits	122 images	0.838 (F1)	VGGNet	[47]
3	Recognize different plant species	Data set of 44 classes	99.60% (CA - correct prediction)	AlexNet	[48]
4	Recognize different plant	91 759 images	48.60% (LC-correct species classification)	AlexNet	[49]
5	Identify obstacles in row crops and grass mowing	437 images	99.9% in row crops and 90.8% in grass mowing (CA)	AlexNet	[50]
6	Identify crop species and diseases	54 306 images	0.9935 (F1)	AlexNet + GoogLeNet	[7]
7	Detect obstacles that are distant, heavily occluded and unknown	48 images	0.72 (F1)	AlexNet + VGG	[51]
8	Leaf disease detection	4483 images	96.30% (CA)	CaffeNet	[3]
9	Identify thistle in winter wheat and spring barley images	4500 images	97.00% (CA)	DenseNet	[52]
10	Predict number of tomatoes in images	24 000 images	91% (RFC-Ratio of total fruits counted) on real images, 93% (RFC) on synthetic images	GoogLeNet + ResNet	[53]
11	Classify banana leaf diseases	3700 images	96% (CA), 0.968 (F1)	LeNet	[54]
12	Identify pig face	1553 images	96.7% (CA)	VGGNet	[55]
13	Classify weed from crop species based on 22 different species in total	10413 images	86.20% (CA)	VGGNet	[56]
14	Detecting and categorizing the criticalness of Fusarium wilt of radish based on thresholding a range of color features	1500 images		GoogLeNet	[57]
15	Fruit counting	24 000 images	91% accuracy	Inception-ResNet	[53]
16	Automatic Plant disease diagnosis for early disease symptoms	8178 images	Overall improvement of the balanced accuracy from 0.78 to 0.87 from previous 2017 study	Deep ResNet	[58]

4 Conclusions and Recommendations

Despite remarkable achievement in use state-of-the-art CNN in agriculture in general, there exist grey areas in relation to smart farm that future researchers may look at. These areas may include; real-time image classification, interactive image classification and interactive object detection. State-of-the-art CNN is relatively a new technology that explain why the finding of the study about their use in smart farm is relatively small. However, its is important to note that models built from state-of-the-art architectures have a impressive record of better precision performance. In this paper, we aimed at establishing the potential of state-of-the-art CNN in IoT based smart farms. In particular we first discussed the architectures of state-of-the-art CNNs and their respective prediction accuracy at the ILSVRC challenge. Then a survey on application of the identified CNNs in Agriculture was performed; to examine the particular application in a smart farm, listed technical details of the architecture employed and overall prediction accuracy achieved according to the author precision metrics. From the study its evident of continuous accuracy improvement of the state-of-the-art CNN architectures as computer vision community put effort to perfect the methods. The findings indicate that state-of-the-art CNN has achieved better precision in all the cases applied in the agricultural domain, scoring higher accuracy in majority of the problem as compared to other image-processing techniques. Considering that the state-of-the-art CNN has achieved state-of-the-art results in prediction in general and high precision in the few farming cases observed, there is great potential that can be achieved in using the methods in smart farming. It has been observed that many authors apply more than one architecture in order to optimize the performance of the network without compromising the expected accuracy. This approach is very efficient in the observed cases, and we recommend similar hybrid approach when building robust IoT based networks which are computationally fair to the mobile devices. This study aims to motivate researchers to experiment and apply the state-of-the-art methods in smart farms problems related to computer vision and data analysis in general.

References

1. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. CoRR, abs/1505.00853 (2015)
2. Luo, P., Li, H.: Research on quantum neural network and its applications based on tanh activation function. Comput. Digit. Eng **16**, 33–39 (2016)
3. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. Comp. Int. Neurosci. (2016)
4. Godfray, H.C., et al.: Food security: the challenge of feeding 9 billion people. Science **327**(5967), 812–818 (2010)
5. Lutz, W.L., Sanderson, W.C., Scherbov, S.: The coming acceleration of global population ageing. Nature **451**, 716–719 (2008)
6. Bruinsma, J. (ed.): World Agriculture: Towards 2015/2030. An FAO Perspective. Earthscan, London (2003)
7. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. **7**, 1419 (2016)
8. Gebbers, R., Adamchuk, V.I.: Precision agriculture and food security. Science **327**(5967), 828–831 (2010)
9. Krintz, C., et al.: SmartFarm: improving agriculture sustainability using modern information technology (2016)
10. Kamilaris, A., Prenafeta Boldú, F.: A review of the use of convolutional neural networks in agriculture. J. Agric. Sci. **156**, 312–322 (2018). https://doi.org/10.1017/S0021859618000436

11. Weber, R.H., Weber, R.: Internet of Things: Legal Perspectives. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-11710-7
12. Ray, A.S.: Remote sensing in agriculture. Int. J. Environ. Agric. Biotechnol. (IJEAB) **1**(3) (2016)
13. Jinbo, C., Xiangliang, C., Han-Chi, F., et al.: Clust. Comput. https://doi.org/10.1007/s10586-018-2022-5
14. Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y.: Big data for remote sensing: challenges and opportunities. Proc. IEEE **104**, 2207–2219 (2016)
15. Waga, D., Rabah, K.: Environmental conditions' big data management and cloud computing analytics for sustainable agriculture. World J. Comput. Appl. Technol. **2**, 73–81 (2017)
16. Chen, M., Mao, S., Liu, Y.: Big data: a survey. MONET **19**, 171–209 (2014)
17. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**, 211–252 (2015)
18. Liang, M., Hu, X.: Recurrent convolutional neural network for object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3367–3375 (2015)
19. Poria, S., Cambria, E., Gelbukh, A.F.: Aspect extraction for opinion mining with a deep convolutional neural network. Knowl.-Based Syst. **108**, 42–49 (2016)
20. Goodfellow, I.J., Bengio, Y., Courville, A.C.: Deep learning. Nature **521**, 436–444 (2015)
21. Bhandare, A., Bhide, M., Gokhale, P., Chandavarkar, R. Applications of convolutional neural networks (2016)
22. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
23. http://image-net.org/challenges/LSVRC/2017. Accessed 02 Sept 2018
24. http://image-net.org/challenges/LSVRC/2012/. Accessed 02 Sept 2018
25. http://image-net.org/challenges/LSVRC/2013/. Accessed 02 Sept 2018
26. ImageNet. http://image-net.org. Accessed 21 Oct 2018
27. Zahangir, M.A., et al.: The history began from AlexNet: a comprehensive survey on deep learning approaches. https://arxiv.org/pdf/1803.01164
28. Kamilaris, A., Prenafeta-Boldu, F.X.: Deep learning in agriculture: a survey. Comput. Electron. Agric. **147**, 70–90 (2018)
29. Zhang, C., Yue, P., Liping, D., Zhaoyan, W.: Automatic identification of center pivot irrigation systems from landsat images using convolutional neural networks. Agriculture **8**, 147 (2018)
30. Chebet, E., Yujian, L., Njuki, S., Yingchun, L.: A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2018.03.032
31. Andri, R., Cavigelli, L., Rossi, D., Benini, L.: Hyperdrive: a systolically scalable binary-weight CNN inference engine for mW IoT end-nodes. In: 2018 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), pp. 509–515 (2018)
32. Canziani, A., Paszke, A., Culurciello, E.: An analysis of deep neural network models for practical applications. CoRR, abs/1605.07678 (2016)
33. HasanPour, S.H., Rouhani, M., Fayyaz, M., Sabokrou, M., Adeli, E.: Towards principled design of deep convolutional networks: introducing SimpNet. CoRR, abs/1802.06205 (2018)
34. LeCun, Y., Bottou, L., Bengio, Y.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
35. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**, 84–90 (2012)

36. Lin, M., Chen, Q., Yan, S.: Network in network. CoRR, abs/1312.4400 (2013)
37. Paszke, A., Chaurasia, A., Kim, S., Culurciello, E.: ENet: a deep neural network architecture for real-time semantic segmentation. CoRR, abs/1606.02147 (2016)
38. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
39. Szegedy, C., et al.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)
40. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)
41. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826 (2016)
42. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2015)
43. Huang, G., Liu, Z., Maaten, L.V., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017)
44. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4. In: Inception-ResNet and the Impact of Residual Connections on Learning. AAAI (2017)
45. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
46. Bargoti, S., Underwood, J.P.: Deep fruit detection in orchards. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 3626–3633 (2017)
47. Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C.: DeepFruits: a fruit detection system using deep neural networks. Sensors **16**, 1222 (2016)
48. Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P.: Deep-plant: plant identification with convolutional neural networks. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 452–456 (2015)
49. Reyes, A.K., Caicedo, J.C., Camargo, J.E.: Fine-tuning deep convolutional networks for plant recognition. In: CLEF (2015)
50. Steen, K.A., Christiansen, P., Karstoft, H., Jørgensen, R.N.: Using deep learning to challenge safety standard for highly autonomous machines in agriculture. J. Imaging **2**, 6 (2016)
51. Christiansen, P., Nielsen, L.N., Steen, K.A., Jørgensen, R.N., Karstoft, H.: DeepAnomaly: combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. Sensors **16**, 1904 (2016)
52. Sørensen, R.A., Rasmussen, J., Nielsen, J., Jørgensen, R.N.: Thistle detection using convolutional neural networks. In: EFITA WCCA 2017 Conference, Montpellier Supagro, Montpellier, France, 2–6 July 2017
53. Rahnemoonfar, M., Sheppard, C.: Deep count: fruit counting based on deep simulated learning. Sensors **17**, 905 (2017)
54. Amara, J., Bouaziz, B., Algergawy, A.: A deep learning-based approach for banana leaf diseases classification. In: BTW (2017)
55. Hansen, M.F., et al.: Towards on-farm pig face recognition using convolutional neural networks. Comput. Ind. **98**, 145–152 (2018)

56. Dyrmann, M., Karstoft, H., Midtiby, H.: Plant species classification using deep convolutional neural network. Biosyst. Eng. **151**, 72–80 (2016)
57. Hyun, J., Ibrahim. H., Irfan, M., Minh, L., Suhyeon, I.: UAV based wilt detection system via convolutional neural networks. Sustain. Comput. Inform. Syst. https://doi.org/10.1016/j.suscom.2018.05.010
58. Picona, A., Alvarez, A., Seitz, M., Ortiz, O., Echazarra, J., Johannes, A.: Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2018.04.002
59. Jia, Y., et al.: Convolutional architecture for fast feature embedding. In: ACM Multimedia (2014)
60. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML (2015)
61. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1026–1034 (2015)

A Collaborative Approach to Build a KBS for Crop Selection: Combining Experts Knowledge and Machine Learning Knowledge Discovery

Mulualem Bitew Anley[1(✉)] and Tibebe Beshah Tesema[2]

[1] Faculty of Informatics, University of Gondar, Gondar, Ethiopia
mulerkal@gmail.com, mulualem.bitew@uog.edu.et
[2] School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia
tibebe.beshah@gmail.com, tibebe.beshah@aau.edu.et

Abstract. This study proposed designing knowledge based solution through the collaboration of experts' knowledge with the machine learning knowledge base to recommending suitable agricultural crops for a farm land. To design the collaborative approach the knowledge was acquired from document analysis, domain experts' interview and hidden knowledge were extracted from Ethiopia national meteorology agency weather dataset and from central statistics agency crop production dataset by using machine learning algorithms. The study follows the design science research methodology, with CommonKADS and HYBRID models; and WEKA, SWI-Prolog 7.32 and Java NetBeans tools for the whole process of extracting knowledge, develop the knowledge base and for developing graphical user interface respectively.

Based on the objective measurement PART rule induction have the highest classifier algorithm which classified correctly 82.6087% among 9867 instances. The designed collaborative approach of experts' knowledge with the knowledge discovery for agricultural crop selections based on the domain expert, farmers and agriculture extension evaluation 95.23%, 82.2% and 88.5% overall performance respectively.

Keywords: Crop selection · Collaborative approach · Knowledge based system · Machine learning · Expertise knowledge

1 Introduction

Crop selection is a process of picking suitable crops for a farm unit based on the physical, economical, climatic and socio-political factors associated with the farm unit [1]. Those parameters are affecting the production yield of the crop. Farmers are facing a challenge during their cropping plan [1]. Taking a correct selection at suited time is directly affecting the whole product of the crop. Taking a selection is a very complex process as there are various factors affecting whole farming process. In order to enhance the decision, it needs to consider rainfall parameters, atmospheric condition, humidity, type of soil and many other factors [2].

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 80–92, 2019.
https://doi.org/10.1007/978-3-030-26630-1_8

Land evaluators and crop experts help farmers to choose the most productive crop that is best suited for a farm unit by assessing its land characters. But not all farmers get such advice and information, especially farmers living in developing country Ethiopia also no exception [3, 4]. On the other hand, agriculture is the main source of economy and the means of employment for nearly 85% of the population, meals supply of the home market and methods for outside cash acquiring. Beside of these, agricultural specialist assistance is not always available when the farmer needs it. In order to alleviate this problem a collaborative approach for combining experts' knowledge with results of knowledge discovery is a suitable solution that can overcome the scarcity of experts [3, 5]. Knowledge can be acquired from different sources such as making interview with the domain experts, document analysis, observation and automatic by using machine learning. Since tacit knowledge is personal and the knowledge expert may not tell all the knowledge she or he knows during the interview, there is hidden knowledge about the problem. Machine learning more general knowledge discovery techniques proposed for extracting hidden and previously unknown knowledge from datasets by different researchers [6].

2 Related Works

A study by Joshi [7] attempted to examine the fuzzy based intelligent system to predict most suitable crop. The fundamental goal of the work is to outline, improvement and organization of a fuzzy based agricultural decision support system which help farmers to make wise decisions regarding selection of a crop. They used a rule based system fuzzy decision support system (DSS) in light of MATLAB and they considered 15 parameters which are greatest among contemporary works and 22 different fuzzy rules.

Another study conducted by Ponnusamy [3] conducted on the design and development of a prototype knowledge based spatial decision support system. The investigation concentrates on assessing land assets and choosing suitable agricultural crops for a farm unit. The researcher considers the land evaluation process by including many experts from varied scientific domains. They at long last create a crop advisor expert system takes land information and recommend in consultation with the farmer a gathering of appropriate harvest decisions that are best as far as physical, monetary, social and political elements related with a farm unit.

A study by Abdulkerim [8] has conducted research on integrating data mining with knowledge based system on the network intrusion detection area and general objectives of his study was to construct a prototype knowledge base system which can update its knowledge base using the hidden knowledge extracted from the intrusion dataset by using data mining techniques. The researcher used knowledge discovery in database (KDD) model for the data mining task, rule based knowledge representation approach to represent knowledge, prototyping approach to develop the knowledge based system, WEKA 3.8.1 to mine hidden knowledge and compare the performance of classifiers, Java NetBeans IDE 7.3 with JDK 6 to develop an integrator application, PROLOG to represent rules in knowledge base, Precision, Recall, F-measure and True Positive rate to evaluate accuracy of the models and test cases to evaluate the performance of the KBS. The system was aimed at utilizing hidden knowledge extracted by employing an induction algorithm of data mining, specifically JRip from sampled KDDcup'99 intrusion dataset. They recommended applying integration of machine learning with a knowledge based system in other domain areas than intrusion detection specially the area that scarce of domain expert.

However, most of the research that has been conducted in the area of KBS their knowledge base was dependent on either domain expert knowledge only or on machine learning techniques which will limit the performance of the recommender. As pointed by [9, 10] each approach has its own limitation on the DSS. In the domain expert only knowledge acquisition experts do often provide a mixture of relevant, irrelevant and erroneous information. On the other side the machine learning approach has a limitation of the too dependent on the data that is provided from outside. According to [11] that have a problem of discovering knowledge might not be the needs of the real world and also has a problem of the rules surplus in depth and in quantity. Due to these problems the user of the rule is needed to decide many times during the use of the rule. Another problem the algorithm produces so many rules that are confused to choose the best rule to make a decision and the rules have too far from the facts so decision maker doesn't believe that rule to make a decision in real world decision making process. A study conducted by [12] stated that collaborating experts with a machine learning approaches was higher benefit to increase the accuracy of the developed knowledge base.

3 Methodology of the Study

3.1 General Research Approach

This study followed the design science research approach as suggested by [13] to design the collaborative approach based KBS that combine experts' knowledge with machine learning knowledge for agricultural extensions and farmers to crop selection decision support system with GUI. According to [13] the design science research process have six basic steps those are problem identification and motivation, definition of objectives for a solution, design and development, demonstration and evaluation lastly communication.

3.2 Data Source

For this study, we used crop production and land information data sourced from the central statics agency agricultural dataset from year 2009 up to year 2015 G.C and climatic dataset obtained from Ethiopian national meteorology agency covering from years 1952 to 2016. After, filtering by the crop type we have get 9867 among 64,327 instances. On the other hand, the NMA data are in Excel for each station by different sheets and monthly data are collected.

3.3 Knowledge Modeling and Representation

The researcher used by adopting the knowledge discovery process model to acquire knowledge from the central statistical agency (CSA) agricultural dataset and climate condition data from the national meteorology agency (NMA) for machine learning knowledge discovery. Machine learning (ML) methods build models based on previous observations which can then be used to predict new data. The model built as a result of a learning process that extracts useful information about the data generation process of the system using the previous observations. Machine learning methods take a set of data corresponding to the process and construct a model of that process in a variety of ways to predict that process. The resulting model can be applied to future data to attempt to predict the crop selection of the knowledge base. We used classification methods J48, PART and JRip decision trees and rule induction classification algorithms.

Expert knowledge base design approach for the development of the knowledge based system we used the CommonKADS knowledge based design model. CommonKADS gives handy documentation for framework configuration transform furthermore energize more awesome modularity and reusability of the model. Arranged with help knowledge engineers in selection knowledge representations and programming techniques with a specific end goal to create a decent outline of a KBS framework. Expects to do this in a way which is both nonspecific platform independent for as long as possible thus opening up possibilities for reusability and economical it encourages preservation of the structures within the expertise model. Nowadays CommonKADS are the primary methodology to support KBS engineering. It helps to perform a detailed analysis of knowledge tasks and processes [14].

3.4 Implementation Procedures and Tools

In order to mine the hidden knowledge from the preprocessed dataset and compare the performance of classifiers the researcher would use WEKA data mining tool. In order to process the data that was collected from the CSA we have used SPSS and Microsoft Excel software. To represent rules in the knowledge base and construct the prototype of crop selection knowledge based system, the researcher used SWI-Prolog. We have used JPL jar file for incorporating the SWI-prolog knowledge base into GUI and java NetBeans for developing the prototype of GUI system.

3.5 Validation and Evaluation Approaches

The performance of the model developed using machine learning techniques are evaluated based on accuracy of classifiers, Precision, Recall, F-measure and True Positive rate. The prototype of KBS is evaluated using system performance testing by preparing test cases. Moreover it is also tested by domain experts and users to ensure user acceptance and check the extent to which the system meets user requirements.

4 Experiment and Result Findings

4.1 Knowledge Modeling

The researchers have used expert knowledge and knowledge from machine learning knowledge elicitation. The expert knowledge are modeled by CommonKADS knowledge modeling. We represented the knowledge model with a unified modeling language (UML) class diagram. In the knowledge modeling stage UML was proposed as a standard notation for CommonKADS methodology for this research.

4.2 Knowledge Discovery

(I) Data Preprocessing. This research used weather and crop production sample survey datasets. CSA dataset was available in the SPSS software format in the CSA of Ethiopia database. Then, the researchers export those datasets into MS-excel for deciding on the usefulness of the data used as input for the stated machine learning methodology accordingly.

(II) Missing Value Handling Method. The researchers use mean imputation method for the meteorology agency dataset. This method is an effective method to fill missing values in the case of time series where the missed value is strongly related to its previous and next values. For NMA data total 11 attributes of the original dataset collected from since 1952 from different district and zonal stations. Among that dataset we use the data that are collected 2009–2015 G.C because of integrating with the CSA agricultural data and the remaining year data have high missing values from station to station. For the remaining NMA data attributes the researchers' uses to handles the missed attribute, the attribute mean to fill in the missing value. For this research CSA agricultural data have total 43 attributes of the original data set, 9 attributes were ignored because of their high missing values and 15 attributes that have no relation to the research goals have been removed. For the remaining attributes the researchers uses to handles the missed attribute values a global constant to fill in the missing value.

(III) Data Integration. The integration was done by being able to identify attributes which match in both data sets and integrating the data based on those variables. In this case the basis for my integrating was the district, season and year where these variables were the same. Other points to consider during data integration on the time unit chosen as a basis for recording time of the data.

(IV) Data Transformation and Data Conversion. The process of converting an attributes values from numeric to nominal and vice versa. Since, for the purpose of this research work, the researchers employed to attribute values convert depicted on Table 1.

(V) Attribute Selection. The attributes used for conducting this research were selected using domain experts investigation and information gain attribute evaluator and ranker search method which are relevant to the proposed study. Descriptions of the selected attributes that have been taken from the original dataset for experimentation are shown in Table 2.

Table 1. Converted attributes value

S. no.	Attributes	Previous values	Values after conversion
1	Field type	Numeric {1, 2, 3}	1 = Pure Stand, 2 = Mixed Crops 3 = Other Land Use
2	Percent of field in use	Numeric {0, 100}	0 = Land Use Only, 100 = Single Crop
3	Seed type	Numeric {1, 2}	1 = Improved Seed, 2 = Non Improved Seed,
4	Irrigation used	Numeric {1, 2}	1 = YES, 2 = NO
5	Irrigation source	Numeric {1, 2, 3, 4, 5}	1 = River, 2 = Lake, 3 = Pond, 4 = Harvested Water, 5 = Others
6	Fertilizer used	Numeric {1, 2}	1 = YES, 2 = NO
7	Fertilizer type	Numeric {1, 2, 3}	1 = Natural, 2 = Chemical, 3 = Both
8	Chemical fertilizer type	Numeric {1, 2, 3, 9}	1 = UREA, 2 = DAP, 3 = BOTH, 9 = NOT STATED
9	Crop damage	Numeric {1, 2}	1 = YES, 2 = NO
10	Prevention measure taken	Numeric {1, 2}	1 = YES, 2 = NO
11	Damage prevention	Numeric {1, 2, 3}	1 = Chemical, 2 = Non Chemical, 3 = Both
12	Extension program	Numeric {1, 2}	1 = YES, 2 = NO
13	Crop type	Numeric {1, 2, 7}	1 = BARLEY, 2 = MAIZE, 7 = TEFF

Table 2. Attributes used for experiment with description

S. no.	Attribute name	Representation type	Value	Description
1	Elevation	Nominal	High, medium, low	The elevation/Altitude of the district
2	Temperature	Nominal	Low, medium, high	Average temperature of each district
3	Fertilizer type	Nominal	Natural, chemical, Both	Type of fertilizer the farmers used
4	Rain fall	Nominal	High, medium, low	Annual rainfall in different stations
5	Field type	Nominal	Pure stand, mixed crop	A plot of land which is used to produce same or mixed crops
6	Seed type	Nominal	Improved, non-Improved	Varieties of seeds
7	Ownership	Nominal	Private, leased	Describes the owner of the farm land
8	Fertilizer used	Nominal	YES, NO	Fertilizer used for harvesting
9	Main season	Nominal	Belg, meher	Seasons for crop cultivation
10	Crop yield	Numeric	Low, medium, high	The total amount of crops produce
11	Irrigation used	Nominal	YES, NO	Farmers used irrigation for cropping or not
12	Percent of field in use	Numeric	Single crop, Land use only	The field in which used to weather cropping or for land use
13	Crop type	Nominal	Barley, Maize, Teff,	Describes most common cereal crops on north Gondar

4.3 Experimental Settings

After data preparation task completed and suitable for WEKA data miner software, we continue the experiment for building a classification model that can support a decision for farmers to choice crops based on their farmland. In order to do these the researchers conduct twelve different experimentations in four models from each three classification algorithms on integrated field information dataset and meteorology datasets to select the best model. In all experiment we have using 9867 instances were taken for training and model building. These experiments were done by using J48 decision tree classifier, PART rule based classifier and JRip rule based classifier algorithms. In the experimentation test set, cross-validation (10 folds) and percentage split (66%) training options employed and WEKA 3.8.0 data mining tool was used. The experiments conducted using confidence factor it ranges from 0.5-0.4 and minnumobj from 1–35 for every classifier algorithm since it permits accomplishing better precision compared with changing the default parameter values.

Experiment One Using J48 Decision Tree. The J48 algorithm is used for building the decision tree model. This experiment were conducted with two scenarios, J48 with pruned and J48 with unpruned tree. In addition the researchers used a test set, cross-validation and percentage split testing options. As a result pruned tree has been simple to interpret and understand the rules generated from the datasets and it makes the decision making easier. The pruned tree J48 model has 286 tree size and 172 numbers of leaves. The algorithm has correctly classified 8139 instances in other word 82.4871% among 9867 instances and 1728 instances in other word 17.5129% classified incorrectly and time taken to build model 0.74 s.

Experiment Two Using PART Rule Based Classifier. This experiment is done by using PART rule based classifier. This experiment was done by using testing set, cross-validation and percentage split test option belongs to the WEKA data mining software. Cross validation was better classified correctly for training the classification model. The model generates 108 rules. The algorithm has correctly classified 8151 (82.6087%) among 9867 instances and 1716 (17.3913%) classified incorrectly and time taken to build model 0.32 s.

Experiment Three Using JRip. This experiment was done by testing set, cross-validation and percentage split test option belongs to the WEKA data mining software. The algorithm generates a model with 32 rules and correctly classified instances are 7666 (77.6933%) and incorrectly classified instances are 2201 (22.3067%) from total number of instances of 9867. The algorithm takes 0.16 s to develop the model.

In algorithms evaluation, from above experiment, it is clearly observed that PART algorithm with the cross-validation test option accuracy and TP rate is better than other algorithms. As a result, it is reasonable to conclude that PART algorithm is better than J48 and JRip method for this study. Therefore, the model which is developed with the PART rule with cross-validation test option classification techniques is considered as the selected working model for the next use in the development of machine learning knowledge base of the KBS and some of the rules generated from the selected model are presented here below:

Rules

```
Rule 1:  If temperature = Low AND elevation = high AND
rainfall = Medium AND Irrigation_used = No AND Crop_Yield
= High: then crop type = BARLEY (1008.0/21.0)
Rule 2: If temperature = Low AND Crop_Yield = Low AND
Main_Season = Belg AND Field_Type = Pure: then crop type=
BARLEY (837.0/96.0)
Rule 3: if Fertilizer_Used = Yes AND Fertilizer_Type =
Both AND temperature = Medium AND Irrigation_used = No
AND Crop_Yield = High AND Field_Type = Pure: then crop
type= TEFF (11.0)
Rule 4: if Fertilizer_Used = Yes AND Fertilizer_Type =
Both AND rainfall = high AND Crop_Yield = Low: then crop
type= MAIZE (12.0/3.0)
```

4.4 The Proposed Solution Architecture

As depicted below on Fig. 1 the design of knowledge based system through the collaboration of expert's with the knowledge discovery to crop selection was begun from document analysis, interview the domain expert understanding the domain area and problem as a whole acquiring knowledge from the experts' and from dataset using machine learning approaches. This research proposed to use dual knowledge base the first knowledge base contained in the machine learning rules which was extracted from the dataset by using a PART rule induction algorithm. The other knowledge base was contained the experts knowledge in the rule representation format. The inference engine inferred from those dual knowledge base for the user input parameters.

The contribution of this study is design an approach that collaborate the experts knowledge with the knowledge extracted by machine learning. The knowledge based system of this research, domain experts, knowledge collaboratively works with machine learning rule. Other contribution of this study, we have developed a java based GUI in Amharic language. Here below is the discussion of the above architecture through the collaboration of expert's with the knowledge discovery for agricultural crop selection using GUI general collaborative approach.

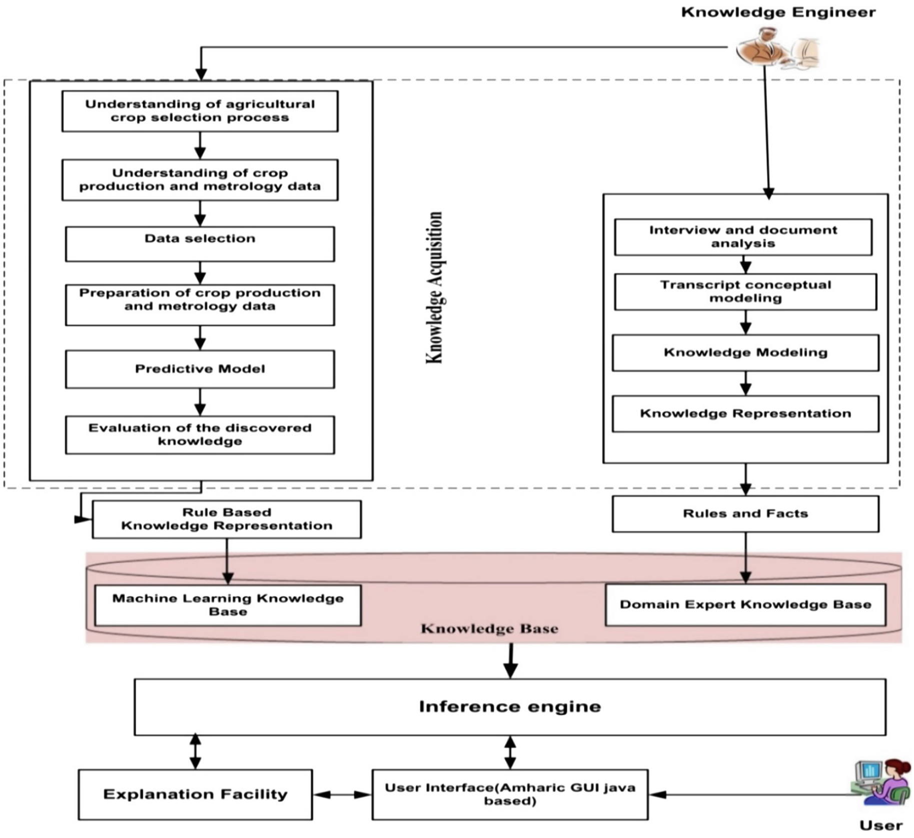


Fig. 1. Proposed collaborative approach architecture

(i) Inference Engine. After accessing the user interface the user feed the major land properties and climatic conditions parameter then the inference engine matches these with the existing rules in the knowledge bases and suggest most suited crop to that precise farming unit. It acts as an interpreter between the knowledge bases and the user interface here below the inference algorithms are discussed using flowchart depicted in Fig. 2.

(ii) User Interface. These studies is contributing to solve the gap that are rising on language barriers by designing and developing Amharic language GUI interface. Here below depicts on Fig. 3 the interface that the crop selection process pages.

As depicted in Fig. 3 shows the user means agricultural extension and farmers properly select the options, then based on the given input the system return barley (ገብስ) as a recommended crop for the given input parameters.

Proposed Algorithm for Collaborative Inference

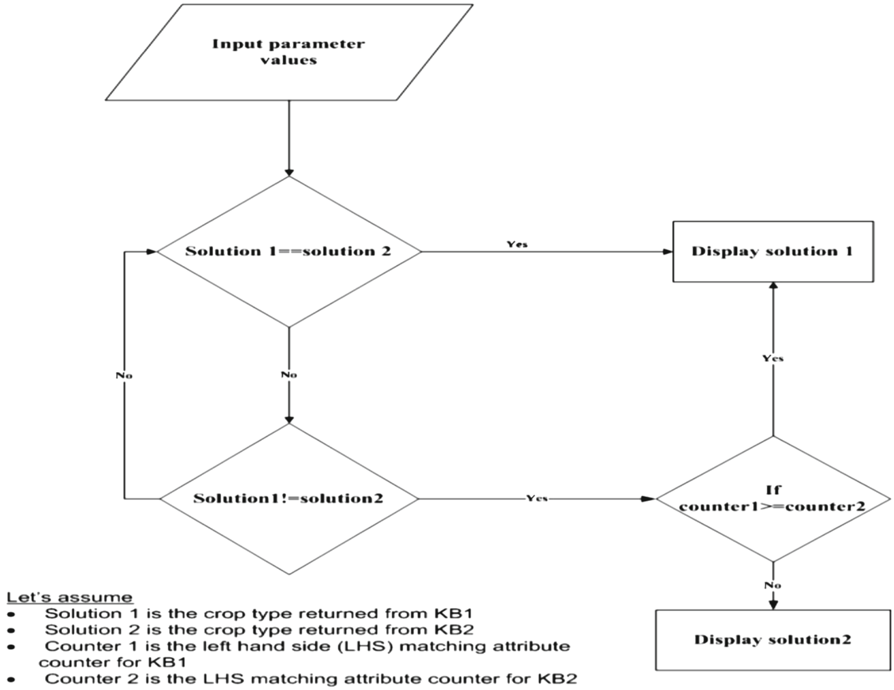


Fig. 2. Collaborative approach inference engine flow chart

Fig. 3. Crop selection recommendation interface.

5 Evaluation and Validation

5.1 System Performance Testing

To evaluate the designed collaborative approach and prototype knowledge based system we have prepared test cases for evaluating the performance of the proposed system. Purposively the researchers prepared a total of 21 test cases and distributed to 4 domain experts. For each of the three crops type's seven test cases are prepared. The domain expert expected to recommend crops based on the farm land area suitability and the estimated weather condition of the area for each test case. For comparing the performance of the experts and the KBS system confusion matrix is used. Based on the collected results from the domain experts evaluation of the total of 21 test cases the system classified 20 instances correctly to their corresponding classes and 1 instance classified incorrectly to other classes to the expected three crops. This means the proposed system's agricultural crop selection recommendation has 95.23% accuracy according to the evaluation performed by the agriculture domain experts.

5.2 User Acceptance Testing

To accomplish the objectives of evaluation the researchers select evaluators by convenience sampling method 10 farmers and 10 agricultural extension workers in the study area. After the discussion of the system, in order to assess the user acceptance of the proposed system, questionnaires were distributed. For this study the format of the questionnaire was adapted from [4] study conducted on cereal crop disease diagnosis prototype KBS. Accordingly, to the farmers' evaluation the system would have the 82.2% performance on the accuracy. The same evaluation questions were forwarded to the agricultural extension workers. According to the agricultural extension evaluation the system would have 88.5% performance for accuracy.

6 Conclusion and Recommendation

This study designed a collaborative approach to combine the domain experts' knowledge with the knowledge discovery result and implemented a prototype knowledge based system for the agricultural crop selection. The approach we used in the knowledge based system development was different from the previous research on the agricultural crop selection. In addition, this research includes the climatic factor parameters for recommending the optimal crop to the farm land. Information gain attributes ranking method and domain experts' evaluation were used and this study finds out that elevation, temperature, fertilizer type, rainfall, field type, seed type and soil type are the main factors to select crops to the farm land. We have done twelve experiments by using J48 decision tree classification algorithm as well as PART and JRip rule induction algorithms using test set, 10-fold cross validation and percentage split test options then after the experiment shows that the PART classification algorithm is the best classification algorithm to recommend the optimum crops for the farm land which classified 82.6087% correctly which means 8151 are classified correctly among 9867 instances. According to the selected farmers' evaluation the developed prototype succeeds 82.2% performance by the domain experts' evaluations have 95.23% accuracy and by the agricultural extensions evaluation the proposed solution succeed 88.5% performance.

We conclude that the collaborative approach based KBS we designed and developed prototype have the potential to increase the accuracy of knowledge based system and a way to collaboration of expertise with machine learning for decision support, at last for increasing the production of crops.

The machine learning knowledge base for this study was done by the knowledge engineer after the PART algorithm generated, further a researcher can also done to be automatically the machine learning rule read with the knowledge based system. In addition, automatic machine translation for extracting rules in different local language also recommended for further study.

Acknowledgment. We would like thank to all Gondar agricultural research center cereal crop research teams for their expert knowledge and guidance in developing the domain expert knowledge base, farmers and agricultural extensions for willing to evaluate the designed collaborative approach and developed prototype. We thank to the central statics agency and Ethiopian national meteorology agency staffs for providing the necessary data and their support regarding to the nature of the data to select relevant features used for conducting this research.

References

1. Rossiter, D.G.: ALES: a framework for land evaluation using a microcomputer. Soil Use Manag. **6**(1), 7–20 (1990)
2. Beed, D.: Software engineering decision support system for agriculture domain, vol. 431602, pp. 1–5
3. Ponnusamy, T.: Knowledge-based expert system for agricultural land use planning, February 2007
4. Tefera, E.: Developing knowledge based system for cereal crop diagnosis and treatment: the case of Kulumsa agriculture research center (2012)
5. Prasad, G., Babu, A.V.: A study on expert systems in agriculture. In: Extension of Technologies: From Labs to Farms, p. 297 (2008)
6. Heines, J.M.: Basic concepts in knowledge-based systems. Mach. Learn. **1**(1), 65–95 (1983)
7. Joshi, R., Fadewar, H., Bhalchandra, P.: Fuzzy based intelligent system to predict most suitable crop (2017)
8. Abdulkerim, M.: Towards integrating data mining with knowledge based system: the case of network intrusion detection. M.Sc. thesis, Addis Ababa University (2013)
9. Gaines, B.R., Shaw, M.L.G.: Integrated knowledge acquisition architectures. J. Intell. Inf. Syst. **1**(1), 9–34 (1992)
10. Webb, G.I., Wells, J.: Recent progress in machine-expert collaboration for knowledge acquisition. In: AI-CONFERENCE, pp. 291–298 (1995)
11. Huang, A., Zhang, L., Zhu, Z., Shi, Y.: Data mining integrated with domain knowledge. In: Shi, Y., Wang, S., Peng, Y., Li, J., Zeng, Y. (eds.) MCDM 2009. CCIS, vol. 35, pp. 184–187. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02298-2_28

12. Webb, G.I., Wells, J., Zheng, Z.: An experimental evaluation of integrating machine learning with knowledge acquisition. Mach. Learn. **35**(1), 5–23 (1999)
13. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007)
14. Schreiber, G., Wielinga, B., de Hoog, R., Akkermans, H., de Velde, W.: CommonKADS: a comprehensive methodology for KBS development. IEEE Expert **9**(6), 28–37 (1994)

Towards Fuzzy Analytical Hierarchy Process Model for Performance Evaluation of Healthcare Sector Services

George Wamamu Musumba(✉) and Ruth Diko Wario

Department of Computer Science and Informatics, University of Free State, Private Bag X13, Kestell 9866, South Africa
musumbaggw@gmail.com, wariord@ufs.ac.za

Abstract. All service providers strive to attain superior service quality (SQ), since the demand for provision of high quality customer focused services is constantly rising. In the healthcare sector, it becomes of utmost importance to understand consumer expectations and their needs to effectively remain competitive in the market. Increase in the per capita income of clients has made them increase their aspiration level to demand enhanced quality services. Currently, hospitals in Nairobi City County (NCC) in Kenya, do not offer services to patients based on their aspirations and in the process, they lose clients. There is need for hospital administrators to determine how patients perceive healthcare quality service, and employ techniques that match their expectations. This is probable by using patients to evaluate their services based on identified criteria. Human evaluations are normally imprecise thus uncertainty aspects should be integrated in the evaluation process. Techniques employed should be able to handle the subjective nature of human evaluators. Fuzzy analytic hierarchy process (FAHP) model has been designed to analyze the service quality in the healthcare sector based on SERVQUAL dimensions. The model is capable of handling the uncertainty and vagueness involved in the mapping of one's preference to an exact number or ratio. It has been used to measure the SQ of four hospitals from NCC by computing the priority of each of the dimensions of healthcare SQ attributes. The priority is then used for ranking the best hospital from the patient's perspective. Hospital managers can use the results of this assessment as a basis of employing strategies that would ensure they provide quality services to patients.

Keywords: Fuzzy analytical hierarchy process · SERVQUAL · Service quality · Performance · Healthcare

1 Introduction

All nations regard healthcare as a key component to their sustainable growth. The healthcare sector imparts directly to a nation's medical, social, political, ethical, business, and financial wellbeing. Creation of counties in Kenya when the new constitution was promulgated in 2010 [1], saw the healthcare services getting devolved and primarily became under the management of county governments. The national

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 93–118, 2019.
https://doi.org/10.1007/978-3-030-26630-1_9

government has recently restructured hospitals into five levels i.e. level one to five, and converting all former provincial hospitals into referral ones. The restructuring of the healthcare system, calls for better services for patients. There are numerous opportunities to improve services if business enterprise models are adopted in managing these healthcare facilities. Seemingly, it is evident that healthcare facilities are becoming autonomous while they strive to pursue targeted service outcomes, amidst budget constraints.

In most business enterprises, the more the clients, the more financially sound they are likely to be. This is also true for healthcare providers. Clients prefer to visit where they perceive to be getting quality services. Some works has defined healthcare quality as being able to use legitimate means to achieve desirable objectives implying a better state health-wise [2]. At the same time quality in healthcare can be thought of as achieving an improved healthcare service level for consumers [3]. It cannot be overemphasized that while striving to improve the quality of healthcare, use of the right technology, treatment in a timely manner, and depending on the demand, provision of adequate service, as well as ability of the practitioners to guarantee acceptable medical practice standards, are indeed necessary [4]. In a corporate point of view, better quality in healthcare can lead to the provider extending the client base and gaining a competitive edge in the market. This is admirable as it assures economic viability and profitability in the long run [5–7].

Healthcare quality is multifaceted. Some authors viewed it also as a multi-dimensional entity. The entities and actors that interactively affect it include, financier organizations and healthcare professionals who tend to the needs of the patients. Additionally, there is a category of healthcare consumers who also value the service level provided in terms of hospital staff relationship with patients, hospitality, privacy and patients' dignity considerations. This constitutes what they perceive as better clinical experience and outcomes.

Some service quality aspects as perceived by patients include how accessible, how responsive and how hospitable the healthcare facility is to the consumers. Researchers suggests that these aspects directly influence the perception of consumers on the facility's value and image. Additionally, patients' perception on quality also impacts indirectly on how they behave in regard to how they rank the facility. Depending on the rank, they may choose to visit the facility or seek medical services elsewhere. If patients are satisfied with the level of service, they are more likely to recommend the facility to others [8, 9]. It is for the same reasons that in today's healthcare context, it is necessary to have in place a systematic way of carrying out evaluation of patients' satisfaction [10].

In a number of research works in the healthcare domain, there exist many techniques that have been employed to determine satisfaction of stakeholders. These techniques are: Methods of Stated Importance (MSIs), Methods of Derived Importance (MDIs) and Multi-Criteria Decision-Making (MCDM). In MSIs [11–13], participants fill out a questionnaire that is well detailed which relates to the fundamental service quality aspects that are based on consumer expectations and perceptions. On the other hand, in MDIs [14, 15], participants fill a non-detailed simplified questionnaire. Participants are only required to provide their assessment on the perceptions on service quality aspects. Furthermore, they are also supposed to synthesize their perceived quality of service and then indicate their overall degree of satisfaction. Statistical derivation of the quality expectations are done in phase two, following the survey in the first phase. As stated in a benchmarking work by LuPo [16], *"derivation is on the basis of relations among the collected perceptions and the overall satisfaction degree"*.

The third category, MCDM base its' working on the principle that considers customers' attitudes toward a specific service. Consumers in this case, give their assessment based on how they perceive the importance of the service aspect when they compare it with another service aspect [17]. For instance, the importance of "responsiveness" compared to "empathy" when they are assessed on their contribution to service quality in hospitals. The importance relationship is normally between two aspects of the service. In this regard, MCDM techniques like (1) Analytical Hierarchy Process (AHP) [18], (2) VIsekriterijumska optimizacija I KOmpromisnoResenje (VIKOR) [19], (3) Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [20] among others, become strong candidates that can be employed in the design and development of models that can be used to assess service quality criteria. Recent application of the above methods in healthcare domain can be found in [16, 21, 22].

Use of the above methods in development of applications can result in imprecise or unreliable outcomes. The length of the questionnaire in SIM, may encourage unreliable results from participants who may want to save on time but still want to please the questionnaire administrators [23]. Authors of this study, are of the opinion that there could be a high tendency of users to rate highly the importance of all service aspects whenever they are requested to directly rate them, potentially leading to overestimation of quality expectations in addition to lower discriminatory power characterization. These methods use linguistic variables in determining preferences of service quality aspects. Research works have stated that using linguistic variables to assess the dimensions of service quality, is plagued with uncertainty, subjectivity and vagueness [24, 25].

To overcome the aforementioned weaknesses, a model should be designed to help healthcare team assess the healthcare service quality. This study proposes a technique based on the ServQual model [26]. ServQual model has been extensively used in literature in assessing service quality in other domains. The proposed model also incorporates the AHP technique to reliably estimate expectations aspects of service quality [27, 28]. AHP is regarded as the most established MCDM method used in various decision situations [29]. It is possible to integrate both quantitative and qualitative techniques in AHP. The ease with which AHP is able to acquire the required data enables it overcome some of the issues that are critical to the decision making process. Some of these critical issues as quoted from Lupo [16] are, "*existence of well documented respondents' tendency to select the central category of the evaluation scale to express their judgments [30], the influence in the evaluation process of the categories number of the evaluation scale, and the form and the type of related linguistic variables"* [31, 32].

Stakeholders while providing their judgments on the service quality aspects, just like it happens in expression of evaluation judgments in other domains, inherently exhibit uncertainty, subjectivity and vagueness [33–35]. This study considers integration of fuzzy set theory (FST) to MCDM techniques. FST was proposed by Zadeh [36]. The result of this integration has been applied in management science, in the decision-making [37] processes, service performance evaluations [38, 39] and information retrievals [40, 41]. The proposed model will provide support to decision makers and administrators in the healthcare sector at the time they need to make choices on the strategies they deem efficient to improve service quality.

2 Literature Review

There exists different interpretations and implementations of service quality as shown in literature. Selecting the best solution for a particular problem is usually the main objective of decision makers. Some decision makers (DMs) consider a single criterion for making decisions. However, a majority of DMs take into account a number of inconsistent and non-commensurable objectives with multiple criteria as it is insufficient to make decisions based on a single criterion. To make decisions based on multiple criteria, some papers propose application of multiple criteria decision making (MCDM) techniques. In essence, this happens in situations where decision criteria are contradictory in nature. In these cases, individuals are facilitated to make decisions based on their preferences [42]. In MCDM approach to decision making, a problem is hierarchically structured with several levels. In level one objective is defined, while level two consists of the set criteria used by individuals to select the best alternative for the decision situation. Sub criteria, sub sub-criteria can be in level three and four respectively. If there is no further division of sub sub-criteria then the list of alternatives to be considered takes the last level of the hierarchy.

The hierarchical structure effectively divides a complex problem into smaller portions and the problem is tackled using both top down and bottom up approaches. Decision makers analyze the criteria, sub criteria to the lowest level in the hierarchy using top down approach and the computation of the analysis is done using bottom up approach. After making the decisions, based on the hierarchical analysis, that have in essence considered issues related to the problem sub divisions, the next step involve the reconstruction of these sub divisions representing an inclusive view of the decision makers (DMs) [43].

Assessment of the service quality, which is the focus of this study, can be done using different qualitative and quantitative criteria that can be structured hierarchically. This effectively suggest the appropriateness of employing the MCDM techniques to address the evaluation of the aspects of quality of service. Literature has also reported several ways that have been used to measure the service quality. They include among others, those based on statistical analysis, analytical hierarchy process (AHP), fuzzy AHP (FAHP), VIsekriterijumska optimizacija i KOmpromisno Resenje (VIKOR), Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE), and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [44].

In many works where MCDM methods have been used in decision situations, the majority of the problems have multiple, conflicting, and incommensurate criteria and/or objectives. There are two categories proposed by Hwang and Yoon [45] to classify the MCDM methods. First, MCDM methods are classified as multi-attribute decision-making (MADM) where emphasis is on the human participation and judgments. The second category is the multi-objective decision-making (MODM) where human participation is seldom emphasized. There is vast amount of research on human judgments and decision making that have continually depicted the human brain as being incapable of holding or consider unlimited amount of information at the same time [46]. These assertions points to its unreliability in taking decisions to solve problems in complex situations. Furthermore, human judgments, in such situations are imprecise which complicates the matter even more. This call for techniques that can simplify this uncertainty aspects in complex decision making environments.

As quoted from Musumba and Wario [47], *"AHP is a MCDM technique that derive weights of importance by using pairwise comparisons (PCMs) of alternatives from a multi-level hierarchical structure. The structure depends on the problem but can have the objective of the problem at the top of the hierarchy and the healthcare facilities as alternatives whose service quality is to be evaluated at the bottom level and the set evaluation criteria in the middle level. Sometimes the comparisons are not consistent. In that case, AHP uses the eigenvalue and consistency checking methods to improve their consistency as described in* [48]*".*

AHP is also one of the most applied MADM technique. It has the ability to impartially and logically grade the decision making attributes by reducing the personal biases and allowing for comparison of disparate alternatives. In addition, its flexibility to be integrated with various techniques like Linear Programming (LP), Quality Function Deployment (QFD), Fuzzy Logic (FL) among others [49, 50] made it a good candidate for consideration is this work.

AHP has a number of weaknesses, including its inability to synthesize values that are not crisp in form. As quoted from the works by Musumba and Wario [47], it is stated, *"in order to overcome AHP weaknesses, Mikhailov* [51] *integrated fuzzy logic* [52] *into AHP forming fuzzy AHP (FAHP). Fuzzy logic* [52] *changes the crisp values of AHP into a continuum thus making it able to deal with uncertainty and vagueness characterizing input variables. Therefore, FAHP can be used to make rational decisions under such conditions. Fuzzy logic is considered the best method compared to deterministic approaches, algorithmic approaches, probabilistic approaches and machine learning* [53] *for problems that users are not certain of the value of parameters to use."*

2.1 MCDM Application in Service Quality

Each industry's success is hinged on the main players' ability to deliver, build, understand and maintain high standards of their service quality. In this regard, in [54], description of service predictions was based on service quality expectations. The key driving forces behind evaluating the service quality are measurement of service performance, diagnosis of service problems, management of the delivery of service and optimal service provision to all customers [55]. It is noted that there exists methods that can be used in the evaluation of the service quality. These methods include MCDM techniques which have been applied in evaluating service quality in airline and other transport industries, websites and Internet services, tourism and the hospitality industry, the healthcare industry, the manufacturing industry, the banking sector, the education sector and service organizations.

2.2 MCDM for Service Quality in Healthcare Industry

Quality service is a key driver in as far as consumer satisfaction is concerned. There is increasing competition due to influx of healthcare providers globally. With increased competition, healthcare consumers have many options of healthcare providers they can consider whenever they need to seek medical attention. With this reality, many healthcare providers know that their continued success going into the future depends on their ability to competitively provide good healthcare service quality [56]. Many patients are concerned with service quality while seeking medical attention. Non satisfaction with the service offered to patients, have higher bearing on them not to return but opting to seek the same in other healthcare facilities. The onus remains with healthcare providers to enhance their healthcare quality if they have to maintain and grow their clientele [57]. Indeed, healthcare providers have very limited options in paying attention to consumers' expectation of service quality if they have to successfully deal with existing globalization challenges and intensive competition. Some of the critical challenges faced by hospital management are, how to meet patients' needs and how to determine the expected service quality [58].

In Shieh et al. [59], hospitals have been classified as service industries because their activities include, direct contact with people, frequent communication with people, provision of customized and professional medical services. Research outputs have shown that development of the service quality of hospitals has existed for many years. However, it is vital to evaluate the consumer service quality expectation to determine with certainty if indeed hospitals' administrators understand the importance attached to quality of service they offer. Furthermore, previous works have shown that the service quality of a hospital can be measured according to different qualitative and quantitative criteria. Consequently, the MCDM approach becomes a suitable choice to evaluate these consumers' expectations of the quality of service healthcare providers' offer. According to findings of a systematic review by Abbas [44], several works have employed the use of MCDM techniques to evaluate the service quality in the healthcare sector.

In the state of the art survey conducted by Kubler et al. [60], it is reported that FAHP evaluation has been applied to the healthcare service sector, where managers need to demonstrate how efficient and customer focused their services are. Büyüközkan et al. [61] developed an evaluation framework for evaluating the performance of service quality among hospitals. The criteria used in the framework include, the level of facilities/equipment, physical appearance of the hospital, its hygiene, responsiveness to consumer issues, degree of reliability, assurance level, empathy and the degree of professionalism exhibited by staff. They concluded that the two most important factors were empathy and professionalism. The study was extended by Büyüközkan and Cifci [50], focusing on the service quality of e-services offered through the hospital's website. They added a sensitivity analysis step to the framework.

As stated earlier, in AHP, the priorities assigned to alternatives depends on the decision makers' highly imprecise judgments, which are subjective in nature. This may lead to instability of the final ranking if there are small variations on the principal weights. It is a common occurrence of people witnessing frequent emergency situations in healthcare facilities which should be tackled in a proper manner. As a result, Ju et al.

[62] has proposed an evaluation index system of the emergency response capability based on FAHP. Their system use 2-tuple linguistic variables which has simplified the understanding of the computed scores by decision makers. In [50, 62], it is pointed out that motivation and skills of human resources significantly affects the hospitals' service quality.

2.3 The SERVQUAL Tool and the Service Quality

According to literature in the business domain as discussed in the works by Shahin and Janatyan [63] and Zaim et al. [64], service quality is defined as the comparison of service expectations with actual performance. Also, service quality as defined by both Parasuraman et al. [65] and Bitner and Hubbert [66] is the impression the consumer creates of the degree of superiority or inferiority of a service provider and its services. Prakash et al. [67] analyzed Grönroos's [68] and Parasuraman et al.'s [26] views on the perceptions of service quality and stated that service quality perceptions are based on multiple dimensions. They neither had any idea about nature nor content of those dimensions.

SERVQUAL tool as presented in [26, 69–73] has been used for a long time to measure healthcare service quality. The tool has the following dimensions, reliability, responsiveness, assurance, empathy and tangibles. When patients receive the appropriate treatment, the healthcare provided is considered to be reliable. Similarly, as discussed in [70], when the healthcare personnel promptly and politely attends to a patient, responsiveness dimension is fulfilled. Regarding the assurance dimension, it is highly rated if patients exude confidence in the qualifications and competence of the healthcare practitioners. Personalized attention to patients from the facility, resulting in better experiences relates to empathy. Lastly, aspects of cleanliness, the physical appearance of the equipment used and the healthcare facility [74] relates to tangibles. In other words, tangible dimension refers to features that are visually visible.

In the Unites Arabs Emirates, Jabnoun and Chacker [75] investigated how patients perceived quality of healthcare service from both private and public hospitals. In addition to reliability, responsiveness, empathy and tangibles dimensions, they also found out that supporting skills was another dimension of the service offered. In Sohail's [73] work to assess the quality of services from Malaysian private hospitals, it is indicated that patients prefer if all aspects of service exceeded expectations in all dimensions. In a study on a hospital in South Africa, it is reported that Boshoff and Gray [76] used communication, tangibles, empathy of nursing staff, assurance, responsiveness of administrative staff, security and physician responsiveness in their evaluations.

A study by Ramsaran-Fowdar [77] on private hospitals in the Mauritian healthcare services, indicated reliability dimension as one of the most important service quality aspect. In Indian hospitals, the SERVQUAL instrument was used by Padma et al. [78] to determine the dimensions of service quality. Patients and their family members/friends perspectives were put into consideration which were eventually proposed to measure the quality dimensions. Attendants' perspectives were also suggested to be part of the aspects of service quality evaluation.

In summary, the preceding literature review has revealed that SERVQUAL instrument can be used to measure the service quality in the healthcare sector. The instrument measures gaps in the service quality based on five dimensions. It is not easy to obtain information on customer expectations. Despite this reality, the “gap model” has continually been used by many studies to address the challenge. With the model, the strengths and weaknesses in specific quality attributes can be identified [79, 80]. This study proposes the use of the SERVQUAL instrument scale [76] to assess patients’ perceptions of service quality for healthcare providers. Hospital environment is used. Fuzzy AHP is also used to deal with the subjective nature of human evaluators’ perception of service quality where linguistic variables are used.

3 Methodology

3.1 Fuzzy Analytical Hierarchical Process (FAHP)

This study has proposed to use SERVQUAL dimensions for assessing service quality. These dimensions become subjective when human assign to them values of importance. It is needed to define meaning of every word used as labels where boundaries of their values become undefined, fuzzy or uncertain. Furthermore, based on individual evaluators’ subjective perception on the level of importance of service quality aspects, the judgment of events by people may be significantly different [81]. Therefore, the study introduces fuzzy numbers for expressing linguistic variables appropriately. Fuzzy numbers are introduced on AHP hence fuzzy AHP (FAHP).

3.2 Fuzzy Set and Triangular Fuzzy Number

As pointed out earlier, while attempting to address the weaknesses in AHP, Mikhailov [51] introduced fuzzy logic [82] in AHP. It is noted from research works that the main focus of fuzzy set theory is on rationalization of uncertainty. Representation of vague data has been synonymous with fuzzy sets. Figure 1 shows a fuzzy number, which is characterized by a membership function. It differs from traditional set which defines an element as either belongs or does not belong to a set (i.e. 0 and 1). It implements grouping of data with boundaries that are not sharply defined (i.e. not crisp).

Methods or theories that implement ‘crisp’ definitions, for instance, the classical set theory, could be ‘fuzzified’. This is possible if the crisp set idea, is generalized to a fuzzy set with blurred boundaries. Extension of crisp theory and analysis methods to fuzzy techniques is beneficial. This can be demonstrated as with the ease it becomes in solving real-world problems. Inevitably, these problems entail some degree of imprecision in the variables and parameters measured and processed for the application. A triangular fuzzy number (TFN) is the special class of fuzzy numbers whose membership is defined by three real numbers, expressed as (a, b, c). The fuzzy triangular membership function gives the foundation for defining other types of membership functions such as general triangular function, right-angled triangular function and trapezoidal function. For example when $a = b$ for a right-angled triangular membership function such as $(1, 1, 3)$.

In this study, TFN has been used to decide the priority of one decision variable over other. A TFN has been given as $a \leq b \leq c$ where b, a, and c are the most likely, the lower bounds and upper bounds decision values, respectively. The TFN, is a linear piece-wise membership function, $\mu_n(x)$ of the form;

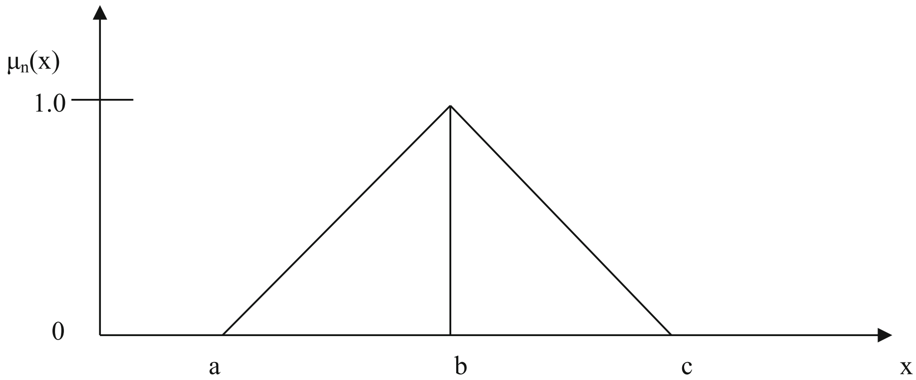


Fig. 1. Fuzzy triangular numbers membership

$$\mu_A = \begin{cases} (x-a)/(b-a), & a \leq x \leq b \\ (c-x)/(c-b), & b \leq x \leq c \\ 0, & \text{otherwise} \end{cases}$$

Where $\infty < a \leq b \leq x \leq c < \infty$

4 Fuzzy AHP Model

Involvement of experts in the healthcare played a bigger role in the design and development of hospital services performance model. The fuzzy AHP model, which is the result of this process adopted a group consensus method, informed by an initial focus group discussions with healthcare domain experts actively involved in the operations of healthcare facilities. The following steps resulted in the model:

(a) Definition of the service quality evaluation criteria and sub-criteria.
(b) Creation of the hierarchical structure based on AHP approach.
(c) Fuzzifying the evaluation criteria and sub-criteria.
(d) Using pair-wise comparison of fuzzified values to estimate the relative weights of elements.
(e) Performance assessment of the estimated elements relative weights.
(f) Defuzzification and performance measurement of the hospitals/healthcare facilities.
(g) Selection of the best healthcare facility based on the determined level of service quality.

4.1 Definition of the Service Quality Evaluation Criteria and Sub-Criteria

This study adopted the five SERVQUAL scale dimensions as the evaluation criteria as described in the following section.

- *Tangibles: Include the physical amenities, available equipment and the appearance of personnel.*
- *Reliability: How accurate and reliable are promised services performed.*
- *Responsiveness: Service performance promptness and willingness to assist customers.*
- *Assurance: This dimension considers employees' knowledge and courtesy and their ability to inspire trust in customers.*
- *Empathy: This dimension considers the ability of the employees to provide care and personalized attention to patients and other consumers.*

Experts in the healthcare sector were involved in sub-dividing these dimensions into sub-criteria. The outcome of the discussions are outlined in the following section. The sub-criteria were defined as questions from both healthcare consumers and providers perspectives. The sub-criteria are indicated after each dimension (See Tables 1, 2, 3, 4 and 5).

Table 1. Reliability

• What is the average time the doctor takes in assessing and treating the patient?
• How long does it take before patients return to hospital for treatment of similar ailment (Reoccurrence rate)?
• Are tests that are prescribed to patients relevant?
• Are tests conducted following in a defined set procedure?

Table 2. Responsiveness

• Do doctors respond to the patients' requests?
• Are there doctor's on-call at night?
• Do nurses respond to patients' requests?
• How quick and convenient is the process of medical care?
• Are emergency cases dealt with promptly by staff?
• Are administrative staff cooperative and helpful?
• How prompt is the ambulance service?

Table 3. Assurance

• Does the hospital physical organization provide safety and convenience to patients?
• Do doctors give confidence to patients?
• What is the extent of confidentiality between doctor and patient?
• Is the hospital reputable among peers?
• Do all staff give confidence to patients?

Table 4. Tangibles

• How clean is the hospital physical environment?
• What is the quality of equipment used for treating patients?
• How are staff organized in terms of appearance (including first-line)?
• How clean are wards in the hospital?
• How excellent is the food in the canteen?
• How excellent is the patients' food?
• How clean/hygienic is the canteen?
• Is water available all the time?
• Is electricity available all the time?
• Does the hospital provide an organized comfortable seating area?
• Is the estimated bill accurate?
• Is there a procedure for paying the final bill?
• Is the hospital well secured?
• Is there a counter for inquiry in the hospital?
• Does the hospital provide for movements of patients and persons with physical disabilities?
• Are beds provided to patients in good condition?
• How patient are doctors in hearing patient details?
• Are nurses patient with patients?
• Do pharmacy staff provide adequate instructions to patients?
• Are service staff (cleaners and ward boys) polite?
• Are patient conditions explained by doctors well to both patients and guardians?
• Do nurses give clear instructions to patients?
• How efficient are doctors when handling emergencies?
• How efficient are other staff when handling the emergency?
• How timely are the scheduled procedures handled?
• Are there exhibitions and displays for directions on living healthy?
• Does the hospital engage in the community training?

Table 5. Empathy

• Does the facility furnish patients with medical information related to their ailments?
• What is the average time patients take to register and / or get admitted?
• What is the average time patients take to be given medicine?
• What is the average time patients take to see a doctor?
• Is there a process in place for reserving a specific service e.g. doctor's consultation?
• How are the schedules for visitors and the duration of the visit?
• Are there facilities for families and visitors and in what state?
• Is the hospital able to respond to patients' requests appropriately?
• Does special requests by patients acceptable?
• When patients' relatives or families make requests, are they attended to?

4.2 Creation of the Hierarchical Structure Based on AHP Approach

The MCDM hierarchical structure created in based on AHP approach. As shown in Fig. 2, it consists of the SERVQUAL dimensions which are the evaluation criteria. Each criteria have sub-criteria that have been defined.

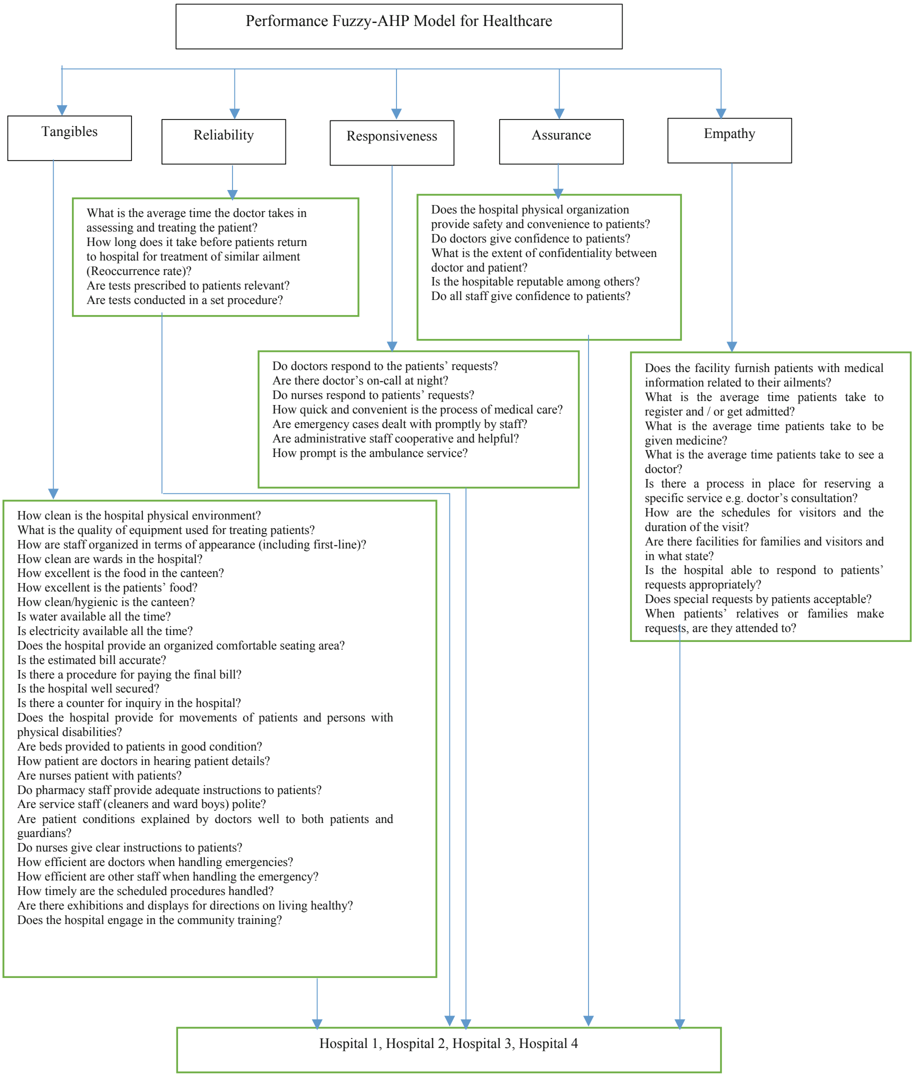


Fig. 2. Hierarchical structure of FAHP model

4.3 Fuzzifying the Evaluation Criteria and Sub-criteria

Using Table 6, the alphabets that have been used to represent linguistic scale, are assigned appropriate crisp values. The SERQUAL scale dimensions are represented using symbols (A, B, C, D, E) with matching nominal scales (equal importance, moderate importance over another, essential or strong importance, very strong or demonstrated importance and absolute importance) as suggested by Sinimole [83]. They are converted to crisp scale and then to fuzzy scale (see Table 6). It should be noted that linguistic symbols can still be converted directly to TFNs.

Table 6. Linguistic scale for importance (adapted from Sinimole [83] and modified)

Linguistic variables	Alphabets	Crisp no.	TFN	Reciprocal of TFN
Equal importance	A	1	(1, 1, 3)	(1/3, 1, 1)
Moderate importance over one another	B	3	(1, 3, 5)	(1/5, 1/3, 1)
Essential or strong importance	C	5	(3, 5, 7)	(1/7, 1/5, 1/3)
Very strong or demonstrated importance	D	7	(5, 7, 9)	(1/9, 1/7, 1/5)
Absolute importance	E	9	(7, 9, 9)	(1/9, 1/9, 1/7)

4.4 Using Pair-Wise Comparison of Fuzzified Values to Estimate the Relative Weights of Elements

The hierarchy was established with the main objective at the top, selection criteria (SERVQUAL dimensions) at second level and the considered healthcare facilities (hospitals) at the last level. In explaining how evaluation of the criteria, assuming the hierarchical structure has s alternatives with respect to a specific objective, which must be evaluated using t criteria, denoted $C_i(i = 1, 2, \ldots t)$. Let the weight of criterion C_i with respect to the objective be W_{Ci}.

Let the relative weight of alternative r $(1 \leq r \leq s)$ with respect to criterion C_i be W_{RCi}. The overall weights, denoted P_i $(1 \leq r \leq s)$ of s alternatives with respect to the objective are given by Eq. 1.

$$P_{WI} = AC \times (CO)^T \tag{1}$$

Where: P_{WI}: Overall weight of alternative r, AC: Relative weights of the criteria with respect to the objective, and CO: Relative weight of the alternatives with respect to criteria. Finally, the alternative with the highest weight value is selected. It is important to note that W_{Ci} $(1 \leq i \leq t)$ are the relative weights of criteria C_i while W_{RCi} are relative weights of alternatives, in this case, the hospitals. These relative weights are computed for elements at level 2 (for dimensions) of the hierarchical structure, then at levels 3 (for hospitals).

To compute the relative weights of the elements in the hierarchy, first, relative weights (weights at a specific level) of alternatives at each level of the hierarchy are calculated. The higher the weight, the better (or preferable) the alternative is. Assessment of relative weights is performed through pairwise comparison of the alternatives (alternatives in this case refers to elements for comparison at each level e.g. hospitals and dimensions), using the Saaty scale. The values of alternatives to be compared are assigned by evaluators (patients) using some chosen method. This results in so called, pairwise comparison matrices (PCM) of alternatives at a level in the hierarchy. For example, considering Fig. 1, relative weights of criteria C_i with respect to the objective are computed at level 2 because the criteria are at level 2 of the hierarchy.

Secondly, at level 3, relative weights of alternatives (hospitals) with respect to criteria C_i are computed. This is because evaluators compare and assign evaluation values to alternatives according to how they satisfy criteria C_i. Saaty [84] suggested that alternatives can be assigned a crisp (exact) value to show how important the alternative is vis-a-vis others according to how evaluators rate them. The exact/crisp values are fuzzified using the triangular fuzzy number approach. For example alternative 1 weighed against alternative 2 may have numerical value 9 (which means alternative 1 is absolutely important than 2), the numerical value of alternative 1 weighed against alternative 3 can be 5 (alternative 1 is strongly or essentially important than 3) while the numerical value of alternative 2 weighed against alternative 3 may be 7 (alternative 2 has very strong or demonstrated importance than 3). Saaty [84] proposes the eigenvalue method to compute pairwise comparison matrices. Each alternative weighed against itself is of equal importance (1). At any level of the hierarchy the comparisons form a square matrix.

Pair-wise comparisons (PCMs) for all criteria at a level in the hierarchy are done. As has been shown earlier, once crisp values are fuzzified, PCM of the same is then performed using the scale given in Table 6. For this study, pair-wise comparisons are done to assess the relative importance between two dimensions by assigning values to them as per the individual evaluators' perception. The PCM is defined as:

$$\bar{\mathrm{A}} = \begin{bmatrix} 1 & \bar{\mathrm{a}}_{12} & \ldots & \bar{\mathrm{a}}_{1n} \\ \bar{\mathrm{a}}_{21} & 1 & \ldots & \bar{\mathrm{a}}_{2n} \\ \ldots & \bar{\mathrm{a}}_{n} & m-1 & 1 \end{bmatrix} \tag{2}$$

where $\bar{\mathrm{a}} = (A_{jk}, B_{jk}, C_{jk})$ is the relative importance of each dimension in pair-wise comparison and A_{jk}, B_{jk}, C_{jk} are the lower bound value, most likely value and upper bound value of the TFN. There exist various techniques for estimating the fuzzy priorities W_i, where $W_i = (W_i^a, W_i^b, W_i^c)$ and $i = 1, 2, \ldots, n$, from the judgment matrix. Logarithmic least-squares technique has been employed in this study due to its effectiveness and reasonableness [81]. Triangular fuzzy weights are computed to represent the relative importance of the criteria according to the values assigned by individuals. The computation of the triangular fuzzy weights using the logarithmic least-squares method, is done using the following equation:

$$W_p = (W_P^a,\ W_P^b,\ W_P^c),\ \mathrm{k} = 1, 2, 3, \ldots \mathrm{N} \tag{3}$$

where,

$$W_P^s = \frac{\left(\prod_{k=1}^{K} a_{jk}^s\right)^{\frac{1}{n}}}{\sum_{i=1}^{n}\left(\prod_{k=1}^{K} a_{jk}^s\right)^{\frac{1}{n}}} \quad \mathrm{s} \in \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$$

Afterwards, computation of the normalized priorities to give the overall weight of each criterion is done.

4.5 Computation of Relative Weights to Rate Hospitals

Patients give feedback on their perception of each service quality by hospitals by rating each hospital based on each dimension. Triangular fuzzy numbers' weighting is used in measuring the patients' evaluations (feedback). The linguistic scales are converted to TFN (see Table 6) to assess the hospitals' performance in its quality of service. Let $PS_j^k = (a_{jk}, b_{jk}, c_{jk})$ denote the normalized triangular fuzzy number used to measure the performance of the j^th^ hospital with respect to the k^{th} criteria ($k = 1, 2, \ldots, N$; $j = 1, 2, \ldots, S$, where S = number of hospitals being compared).

4.6 Computation of Group Decisions

The evaluations are carried out by several individual evaluators. It is cumbersome to use individual evaluators hence the need for aggregation of the group evaluation. Geometric mean can satisfactorily deliver fuzzy group weightings. It has been applied in the fuzzy values (a_{jk}, b_{jk}, c_{jk}) from each individual evaluators. Literature show that geometric mean operations have been used for aggregating group decisions [85], especially where AHP is applied:

$$a_{jk} = \left(\prod\nolimits_{s=1}^{S} a_{jks}\right)^{\frac{1}{s}},\ b_{jk} = \left(\prod\nolimits_{s=1}^{S} b_{jks}\right)^{\frac{1}{s}} c_{jk} = \left(\prod\nolimits_{s=1}^{S} c_{jks}\right)^{\frac{1}{s}} \tag{4}$$

where $(a_{jks}, b_{jks}, c_{jks})$ is the fuzzy values obtained after evaluation hospitals' service aspects s ($s = 1, 2, \ldots$, S).

4.7 Computation of Overall Aggregated Score

The study use the computed fuzzy assessments PM_j^k with the relative importance W_p to calculate the overall aggregated weight of one hospital in relation to an overall aggregated weight of another. In essence, the weighted mean operator is used to calculate overall aggregated score (OAS) for each alternative as indicated in Eq. 5.

$$OAS_j = \left[\frac{1}{\sum W_j}\right] \times \left[(PM_j^1 \times W^1) + (PM_j^2 \times W^2) + \ldots + (PM_j^k \times W^p)\right] \tag{5}$$

where OAS_j denotes the fuzzy number indicating the overall aggregated score (or performance) of the j^{th} hospital. Zadeh's extension principle [82] is used to compute the membership function of these fuzzy numbers. However, it is noted that the form of OAS_j does not conform to TFNs, in as much as it is possible to approximate it as one, taking the form (U_j, V_j, W_j).

4.8 Using Center of Gravity to Defuzzify the Aggregated Fuzzy Set

Centre of gravity (COG) is used to defuzzify the OAS, which is a triangular fuzzy number. Defuzzification process results in crisp weight values. COG is preferred in this study because it is easy to use as opposed to other methods like maximum-membership principle, centre of area, weighted average and smallest of maximum. The overall hospital performance (OHP), is then calculated using the following equation:

$$OHP_j = \left[\frac{(C_j - A_j) + (B_j - A_j)}{3} + A_j\right] \tag{6}$$

where OHP_j is the overall hospital performance of the j^{th} hospital, C_j is the upper limit of the TFN, A_j is the lower limit of the TFN and B_j is the average of the TFN. With this outcome, the best healthcare facility (hospital) is selected.

5 Case Study

5.1 Hospitals' Data

This study considered a case of Nairobi city county, Kenya. Four large hospitals were selected to be evaluated based on the quality of their service. This sample was not representative of hospitals in the county but it was good for demonstration purposes. These hospitals included two public or government hospitals and two private hospitals. Patients who had made visits to these hospitals for a minimum of four times were invited to participate. Hospitals used in the study have been given anonymous names at the request of hospital administrators due to the sensitivity of the possible research findings. Use of site locations from the central business district has been employed to name the hospitals. The location does not necessarily indicate where the hospital are located. These hospitals include:

- Nairobi West Hospital (NW).
- Nairobi East Hospital (NE).
- Nairobi North Hospital (NN).
- Nairobi South Hospital (NS).

5.2 Design of the Evaluation Tool

To develop a reliable evaluation tool, to be used by stakeholders in the evaluation of service quality, the study employed the five dimensions (criteria) in the SERVQUAL instrument, namely tangibility, reliability, responsiveness, assurance and empathy. Researchers engaged experienced healthcare practitioners to examine the tool and used their recommendations to make necessary modifications to the tool. Domain expert opinions were used to obtain the weights associated with these criteria. Twenty (20) practitioners gave their opinions on how they rated the four hospitals in each dimension (criterion). In addition 20 patients gave their perceptions on the importance of each dimension when compared with each other in regard to how each contributed to the quality of service as witnessed in each of the four hospitals. All chronbach's alpha (a tool used to check reliability of the evaluation tool) values exceeded 0.8 and the consistency ratio (CR) was less than 0.1, hence ensuring reliability and confirming the appropriateness of the sub-criteria used for each dimension.

5.3 Evaluation of Criteria and Hospitals' Performance

The next step is to compute the PCMs of the values assigned to dimensions. This results in the fuzzy PCM (FPCM) in form of TFN *(a, b, c),* which is a pairwise comparison judgment matrix that indicates the preference of one dimension (D_i) over the other (D_j). See Eq. 7.

$$D_{ij} = \frac{D_i}{D_j} \quad for\ i,j = 1,2,3,\ldots n. \tag{7}$$

This procedure calculates the importance of each dimension from the stakeholders' opinions. This process created a fuzzy pairwise comparison matrix (FPCM) which was normalized to get a normalized fuzzy FPCM (NFPCM). From the NFPCM, normalized fuzzy weights for the five dimensions were calculated (see Tables 7, 8 and 9).

Table 7. Fuzzy pair-wise comparison matrix

	Tang.			Reli.			Resp.			Assu.			Empa.		
	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u
Tang.	1.000	1.001	3.000	0.150	0.230	0.451	0.140	0.202	0.350	0.300	0.781	1.000	0.300	0.811	0.921
Reli.	2.210	4.440	6.521	1.000	1.000	3.000	0.251	0.551	1.000	1.811	3.071	5.330	3.600	5.651	7.560
Resp.	1.960	5.131	7.180	0.531	1.531	2.100	1.000	1.000	3.000	0.871	1.411	2.470	4.440	6.561	7.860
Assu.	0.430	1.290	1.451	0.240	0.331	0.710	0.371	0.710	1.151	1.000	1.000	3.000	1.000	1.090	3.120
Empa.	0.431	1.230	1.310	0.131	0.180	0.281	0.130	0.160	0.240	0.321	0.920	1.000	1.000	1.000	3.000

Tang – Tangibles, Reli – Reliability, Resp – Responsiveness, Assu – Assurance, Empa – Empathy
A_l – Lower bound value, B_m – Middle value and C_u – Upper bound value

Table 8. Normalized fuzzy pair-wise comparison matrix

	Tang.			Reli.			Resp.			Assu.			Empa.		
	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u
Tang.	0.141	0.080	0.150	0.071	0.071	0.071	0.071	0.080	0.060	0.071	0.081	0.081	0.031	0.051	0.041
Reli.	0.311	0.340	0.330	0.491	0.310	0.460	0.131	0.210	0.170	0.421	0.211	0.421	0.351	0.370	0.341
Resp.	0.421	0.390	0.370	0.260	0.470	0.320	0.530	0.380	0.520	0.200	0.381	0.190	0.431	0.430	0.350
Assu.	0.061	0.100	0.071	0.120	0.100	0.110	0.200	0.271	0.200	0.230	0.270	0.230	0.100	0.070	0.140
Empa.	0.060	0.091	0.071	0.061	0.051	0.041	0.071	0.061	0.041	0.071	0.061	0.081	0.100	0.071	0.131

Table 9. Weights of each dimension

	A_l	B_m	C_u
Tang.	0.071	0.071	0.071
Reli.	0.310	0.280	0.330
Resp.	0.350	0.410	0.330
Assu.	0.131	0.141	0.141
Empa.	0.071	0.071	0.071

When pairwise comparison of the dimensions was performed, the results showed that ‘responsiveness’ was ranked with the highest overall weight followed by the ‘reliability’. By inference, it seems patients highly regarded doctors and other staff qualification, reliability and responsibility traits as critical factors that determined the level of service quality in the healthcare sector. If responsiveness and reliability dimensions are given proper attention, then patient trust goes high. Similarly paying more attention to tangibles, empathy and assurance, yields greater satisfaction by patients.

After determining the performance of the dimensions and by extension the performance of sub-criteria, assessment of the performance of the hospitals was carried out using the five SERVQUAL dimensions. In accomplishing this task, the SERVQUAL scale was the basis by which collection of patient’s feedback was done. As shown in Table 10, calculation of the weights for each dimension per hospital was done using TFN resulting in a performance matrix.

Table 10. Performance matrix

	NW			NE			NN			NS		
	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u
Tang.	5.441	7.000	8.580	4.091	5.830	7.561	3.630	5.251	6.890	1.581	3.120	4.611
Reli.	4.161	5.861	7.570	3.811	5.630	7.441	3.400	5.221	7.031	2.671	4.460	6.251
Resp.	4.421	6.041	7.640	4.191	5.861	7.591	3.450	5.111	6.761	1.771	3.380	4.961
Assu.	5.490	7.030	8.560	4.181	5.941	7.690	4.060	5.831	7.570	2.352	3.970	5.581
Empa.	4.740	6.410	8.081	3.560	5.300	7.060	3.380	4.961	6.540	1.491	2.910	4.331

The resultant weights were normalized with a normalized performance matrix as shown in Table 11 as the output.

Table 11. Normalized performance matrix

	NW			NE			NN			NS		
	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u	A_l	B_m	C_u
Tan	0.222	0.222	0.210	0.210	0.200	0.200	0.200	0.200	0.200	0.160	0.181	0.181
Rel	0.171	0.180	0.191	0.191	0.200	0.200	0.191	0.200	0.200	0.271	0.251	0.241
Res	0.181	0.191	0.191	0.210	0.210	0.200	0.191	0.191	0.191	0.181	0.191	0.191
Ass	0.232	0.222	0.210	0.210	0.210	0.210	0.230	0.222	0.222	0.240	0.222	0.222
Emp	0.200	0.200	0.200	0.181	0.191	0.191	0.191	0.191	0.191	0.151	0.160	0.171

Computation of the overall aggregated score (representing efficiency) of each hospital is carried out. This is done to remove any possible patients’ biases on their perceptions on any one of the dimensions and is achieved by normalizing the performance matrix values with each dimension’s relative weight. Equation 5 defined in Sect. 4.7 was used to compute the aggregated weights of each hospital.

Overall performance of each hospital is given in Table 12. To conclude the process of performance evaluation, as discussed in the previous sections, defuzzification of the fuzzy performance values is accomplished using the centre of gravity method. This results in performance score of each hospital.

Table 12. Triangular fuzzy overall performance matrix

	A_l	B_m	C_u
NS	0.21460	0.20890	0.21100
NN	0.19670	0.19890	0.20030
NE	0.20190	0.20190	0.20120
NW	0.18880	0.19250	0.19430

Table 13 shows the overall performance score after defuzzification process. From the Table, it is demonstrated that the patients used in this study ranked hospital NS, NE, NN and NW from the best to the worst respectively according to how they perceived the quality of their services.

Table 13. Scoring of the hospitals

Hospital	Weight	Position
NS	0.211541	1
NN	0.198620	3
NE	0.201670	2
NW	0.191861	4

6 Discussions

Healthcare sector is service industry and such it is difficult to quantitatively measure performance of quality of service. Specifically, these services have two sides which are, on one side, hospitals' staff giving care and the other side reception of the same by patients. In this study, fuzzy AHP model was developed and used to measure the service quality of the hospitals, eventually attempting to evaluate the hospitals. On one hand, researchers are of the opinion that this model can be used to effectively monitor the organizational performance of hospitals, enabling patients to identify a healthcare facility that would meet their expectations. On the other hand, they encourage decision makers to use it for analyzing their healthcare facilities' performance and make necessary improvements. In Table 13, the results show that for the hospitals whose perceived service quality performance was evaluated, patients ranked NS highest, followed by NE, NN and NW respectively. Interestingly, NS hospital which has been highly rated, also was perceived to be the best in reliability dimension. However, it was rated poorly both in tangibles and empathy dimensions.

Both hospitals NS and NN are private hospitals, while NE and NW are government hospitals. As shown in Table 13, hospital NE which was ranked second for its service quality is a leading public referral healthcare facility in the county. Its score however, is hardly better than NN and NW. In the third place is NN, a private hospital equipped with good medical infrastructures. From Table 11, NN was ranked second in both reliability and assurance dimensions while ranking third in tangibles, responsiveness and empathy dimensions. Its management should pay more attention to the third ranking dimensions (based on their weighting, starting with responsiveness, then tangibles and empathy in that order) to improve its level of service quality.

Service quality of hospital NW (a public hospital) is below par as it ranks fourth as per the evaluators. As shown in Table 11, its scores for both reliability and responsiveness dimensions are very low in comparisons to scores of other hospitals. In the same hospital, assurance, tangibles and empathy are ranked in that order as the highest to the lowest. Therefore, assurance and tangibles seems to be the main strengths for this hospital.

The study sampled both government and private hospitals, and it appears that government hospitals are performing poorly compared to the private hospitals on the quality of healthcare services they provide. The physicians in both sectors are well trained and experienced as they are trained from the same institutions and can therefore provide relatively equivalent quality of service. However, the government hospitals have limited facilities terms of infrastructure. As a result, the tangible dimension is greatly hampered. Therefore, these hospitals should put more attention to responsiveness, reliability, assurance and empathy.

This study has shown that by using a fuzzy linguistic framework, it has been possible to assess the overall performance of service quality of each dimension. With this, hospital administrators are able to select and implement the right strategies to improve the service quality, consequently enhancing customer experience. It has also been shown that FAHP model defined in this work, clearly differentiates the efficient and inefficient areas to inform necessary adjustments. Finally, it can be stated that using this model, it is possible to define the flow of processes that are necessary to evaluate the service quality of any given hospital.

7 Conclusion

This study proposed integration of fuzzy logic theory into AHP hence FAHP model for healthcare sector service quality assessment methodology. The methodology implements a fuzzy linguistic framework to compute the overall aggregated scores (OAS) for the selected four hospitals in Nairobi City County. The five SERVQUAL dimensions are used to determine the efficiency of these hospitals as perceived by healthcare consumers. Using this approach, hospital management can identify the dimensions in which they need to focus on for improvement. They can also be able consider a portfolio of quality strategies and select the optimal ones.

In this study, by using fuzzy linguistic framework, it has been made easy to assess and control the process of planning how to achieve desirable service quality. The framework has also made it possible to integrate and measure the service quality dimensions. Hospital managers can use the results of this assessment as a basis of employing strategies that would ensure quality of service in terms of giving care and general experience while patients are using their facilities. Customer satisfaction is easily achieved with higher service quality.

Each region has different social, cultural, economic and political settings which has effects on the operations of the hospital in service delivery. Hospital service delivery in itself is complex. Due to these differences, there are bound to be differences on how various factors play out in different hospital settings. Eventually trade-offs between decision criteria are inevitable. These were evident in this study, making FAHP the most appropriate tool that could be successfully applied to evaluate the service quality. This was because there were multiple dimensions to be used in decision-making by patients.

Use of FAHP to measure performance of hospitals had the following merits:

(a) The service at each hospital was multi-dimensional and while the dimensions were subjective in nature, FAHP could easily be used to model the measurement of performance of such systems.
(b) There exists fuzzy group decision making techniques that employ AHP approach. Use of multiple evaluators for the service quality performance measurement, makes the process, a group decision-making one.
(c) It is easy to identify inefficient aspects of service quality measurements by examining the specific hospital dimensions by use of this hierarchical analysis.

There is a paradigm shift in performance measurement in hospitals. Patients assess overall organizational performance of the hospital as opposed to assessing the individual staff performance. It is therefore necessary to have in place tools/techniques that can be used to evaluate the performance level of the organization's the systems and processes. To address this requirement, FAHP model for healthcare sector has been developed as a tool that can be applied to evaluate the level of service quality in hospitals in Nairobi City County. Hospital administrators can use the results of model to form the basis for employing strategies that ensures desirable service quality that satisfy patients. They can do this by improving the hospitals' operational performance.

The most sensitive task in the model entails, metric development whose weight matrix may be subjective. However, the subjective nature of the matrix weight has been reduced by the use of fuzzy linguistic variables. Researchers propose more exhaustive and substantial work with more data to improve performance of the model. It is also proposed that hospital facilities should be considered as one of the evaluation factor.

In conclusion, this study analyzed the service quality performance of the healthcare sector of the Nairobi City County using the FAHP model. The SERVQUAL instruments dimensions were used as evaluation criteria. The fuzzy nature of evaluators in expressing their judgments on service quality was handled using FST integrated in AHP. Service quality expectations were obtained using the AHP component of FAHP. Finally, the results show that healthcare service quality can be assessed and depending on the outcome, administrators can adjust their strategies to achieve desirable results.

References

1. The Constitution of Kenya (2010). https://www.wipo.int/edocs/lexdocs/laws/en/ke/ke019en.pdf. Accessed Aug 2018
2. Donabedian, A.: Quality assessment and assurance: unity of purpose, diversity of means. Inquiry **25**(1), 175–192 (1988)
3. Marshall, G.N., Ron, D.H., Rebecca, M.: Health status and satisfaction with healthcare: results from the medical outcomes study. J. Consult. Clin. Psychol. **64**(2), 380–390 (1996)
4. Li, L., Benton, W.C.: Hospital capacity management decisions: emphasis on cost control and quality enhancement. Eur. J. Oper. Res. **146**(3), 596–614 (2003)
5. Headley, D.E., Miller, S.J.: Measuring service quality and its relationship to future consumer behaviour. Mark. Health Serv. **13**(4), 32–42 (1993)
6. Sgroi, F., Di Trapani, A.M., Testa, R., Tudisca, S.: Strategy to increase the farm competitiveness. Am. J. Agric. Biol. Sci. **9**, 394–400 (2014)
7. Squatrito, R., Sgroi, F., Tudisca, S., Di Trapani, A.M., Testa, R.: Post feed-in scheme photovoltaic system feasibility evaluation in Italy: sicilian case studies. Energies **7**, 7147–7165 (2014)
8. Choi, K.S., Cho, W.H., Lee, S., Lee, H., Kim, C.: The relationships among quality, value, satisfaction and behavioral intention in health care provider choice: a South Korean study. J. Bus. Res. **57**(8), 913–921 (2004)
9. Lee, M.A., Yom, Y.H.: A comparative study of patients' and nurses' perceptions of the quality of nursing services, satisfaction and intent to revisit the hospital: a questionnaire survey. Int. J. Nurs. Stud. **44**, 545–555 (2007)
10. Hulton, L.A., Matthews, Z., Stones, R.W.: Applying a framework for assessing the quality of maternal health services in urban India. Soc. Sci. Med. **64**(10), 2083–2095 (2007)
11. Brown, T.J., Churchill, G.A.: Peter, J: Improving the measurement of service quality. J. Retail. **69**(1), 127–139 (1993)
12. Kaplan, R.S., Norton, D.P.: The balanced scorecard: measures that drive performance. Harv. Bus. Rev. 71–79 (1992)
13. Schvaneveldt, S.J., Enkawa, T., Miyakawa, M.: Consumer evaluation perspectives of service quality: evaluation factors and two-way model of quality. Total Qual. Manag. **2**, 149–161 (1991)
14. Kucukarslan, S.N., Nadkarni, A.: Evaluating medication-related services in a hospital setting using the disconfirmation of expectations model of satisfaction. Res. Soc. Adm. Pharm. **4**(1), 12–22 (2008)
15. Laschinger, H.S., Hall, L.M., Pedersen, C., Almost, J.: A psychometric analysis of the patient satisfaction ith nursing care quality questionnaire: an actionable approach to measuring patient satisfaction. J. Nurs. Care Qual. **20**(3), 220–230 (2005)
16. Lupo, T.: Benchmarking of airports service quality by a new fuzzy MCDM approach. Eur. Transp. Trasporti Europei **57** (2015)
17. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addion-Wesley, Reading (1975)
18. Saaty, T.L.: Decision making with the analytic hierarchy process. Int. J. Serv. Sci. **1**(1), 83–98 (2008)
19. Opricovic, S., Tzeng, G.H.: Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. Eur. J. Oper. Res. **156**(2), 445–455 (2004)
20. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications. Springer, New York (1981). https://doi.org/10.1007/978-3-642-48318-9

21. Li, Y., Hu, Y., Zhang, X., Deng, Y., Mahadevan, S.: An evidential DEMATEL method to identify critical success factors in emergency management. Appl. Soft Comput. J. **22**, 504–510 (2014)
22. Chang, T.H.: Fuzzy VIKOR method: a case study of the hospital service evaluation in Taiwan. Inf. Sci. **271**, 196–212 (2014)
23. Chou, C.C., Liu, L.J., Huang, S.F., Yih, J.M., Han, T.C.: An evaluation of airline service quality using the fuzzy weighted ServQual method. Appl. Soft Comput. **11**, 2117–2128 (2011)
24. De Battisti, F., Nicolini, G., Salini, S.: The Rasch model in customer satisfaction survey data. Qual. Technol. Quant. Manag. **7**(1), 15–34 (2010)
25. Hu, H.Y., Lee, Y.C., Yen, T.M.: Service quality gaps analysis based on fuzzy linguistic SERVQUAL with a case study in hospital out-patient services. TQM J. **22**(5), 499–515 (2010)
26. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A conceptual model of service quality and its implications for future research. J. Mark. **49**, 41–50 (1985)
27. Aktas, A., Cebi, S., Temiz, I.: A new evaluation model for service quality of healthcare systems based on AHP and information axiom. J. Intell. Fuzzy Syst. **28**, 1009–1021 (2015)
28. Baki, B., Peker, I.: An integrated evaluation model for service quality of hospitals: a case study from Turkey. J. Mult. Valued Logic Soft Comput. **24**(5–6), 453–474 (2015)
29. Lee, G., Lin, H.: Customer perceptions of e-service quality in online shopping. Int. J. Retail Distrib. Manag. **33**(2), 161–176 (2005)
30. Mardani, A., Jusoh, A., Zavadskas, E.K.: Fuzzy multiple criteria decision-making techniques and applications: two decades review from 1994 to 2014. Expert Syst. Appl. **42**(8), 4126–4148 (2015)
31. Calabrese, A.: Service productivity and service quality: a necessary trade-off? Prod. Int. J. Econ. **135**(2), 800–812 (2012)
32. Babakus, E., Boller, G.W.: An empirical assessment of the SERVQUAL scale. J. Bus. Res. **24**(3), 253–268 (1992)
33. Wasserman, G.S.: On how to prioritize design requirements during the QFD planning process. IIE Trans. **25**(3), 59–65 (1993)
34. Sgroi, F., Tudisca, S., Di Trapani, A.M., Testa, R.: Economic evaluation of aquaculture investments under conditions of risk and uncertainty in the Mediterranean Sea. Am. J. Appl. Sci. **11**, 1727–1734 (2014)
35. Sgroi, F., Tudisca, S., Di Trapani, A.M., Testa, R., Squatrito, R.: Efficacy and efficiency of Italian energy policy: the case of PV systems in greenhouse farms. Energies **7**(6), 3985–4001 (2014)
36. Zadeh, L.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
37. Certa, A., Enea, M., Galante, G.M., La Fata, C.M.: A multistep methodology for the evaluation of human resources using the evidence theory. Int. J. Intell. Syst. **28**(11), 1072–1088 (2013)
38. Montes, R., Sánchez, A.M., Villar, P., Herrera, F.: A web tool to support decision making in the housing market using hesitant fuzzy linguistic termsets. Appl. Soft Comput. **35**, 949–957 (2015)
39. Heradio, R., Cabrerizo, F.J., Fernandez-Amoros, D., Herrera, M., Herrera-Viedma, E.: A fuzzy linguistic model to evaluate the quality of library 2.0 functionalities. Int. J. Inf. Manag. **33**(4), 642–654 (2013)

40. Cabrerizo, F.J., Martínez, M.A., López-Gijón, J., Esteban, B., Herrera-Viedma, E.: Anew quality evaluation model generating recommendations to improve the digital services provided by the academic digital libraries. In: 2011 International Fuzzy Systems Association World Congress and the 2011 Asian Fuzzy Systems Society (IFSA 2011), Surabaya and Bali Island (Indonesia), pp. RW 202 1–RW 202 6 (2011)
41. López-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: Applying multi-objective evolutionary algorithms to the automatic learning of extended boolean queries in fuzzy ordinal linguistic information retrieval systems. Fuzzy Sets Syst. **160**, 2192–2205 (2009)
42. Bogetoft, P., Pruzan, P.: Planning with multiple criteria: investigation, communication and choice. J. Multi-Criteria Decis. Anal. **8**(2), 112 (1997)
43. Dodgson, J., Spackman, M., Pearman, A., Phillips, L.: Multi-criteria analysis: a manual. Department for Communities and Local Government, London (2009). http://eprints.lse.ac.uk/12761/
44. Abbas, M., Ahmad, J., Edmundas, K.Z., Zainab, K., Khalil, M.D.N.: Application of multiple-criteria decision-making techniques and approaches to evaluating of service quality: a systematic review of the literature. J. Bus. Econ. Manag. **16**(5), 1034–1068 (2015)
45. Hwang, C.L., Yoon, K.: Multiple Attribute Decision-Making Methods and Applications. Springer, Heidelberg (1981). https://doi.org/10.1007/978-3-642-48318-9
46. Simpson, L.: Do decision makers know what they prefer?: MAVT and ELECTRE II. J. Oper. Res. Soc. **47**(7), 919–929 (1996)
47. Musumba, G.W., Wario, R.D.: Towards group fuzzy analytical hierarchy process. In: Mekuria, F., Nigussie, E.E., Dargie, W., Edward, M., Tegegne, T. (eds.) ICT4DA 2017. LNICST, vol. 244, pp. 283–307. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95153-9_27
48. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
49. Saaty, T.L., Vargas, L.G.: Models, methods, concepts & applications of the analytic hierarchy process, vol. 175. Springer, New York (2001). https://doi.org/10.1007/978-1-4614-3597-6
50. Vaidya, O., Kumar, S.: Analytic hierarchy process: an overview of applications. Eur. J. Oper. Res. **169**(1), 1–29 (2006)
51. Mikhailov, L.: Deriving priorities from fuzzy pairwise comparison judgements. Fuzzy Sets Syst. **134**(3), 365–385 (2003)
52. van Laarhoven, P., van Pedrycz, W.: A fuzzy extension of Saaty's priority theory. Fuzzy Sets Syst. **11**(1), 199–227 (1983)
53. Ahmad, A., Basir, O., Hassanein, K.: Adaptive user interfaces for intelligent e-learning: issues and trends. In: Proceedings of the Fourth International Conference on Electronic Business (ICEB 2004), Beijing, pp. 925–934 (2004)
54. Tseng, M.L.: A causal and effect decision making model of service quality expectation using grey-fuzzy DEMATEL approach. Expert Syst. Appl. **36**(4), 7738–7748 (2009)
55. Li, L.X.: Relationships between determinants of hospital quality management and service quality performance – a path analytic model. Omega **25**(5), 535–545 (1997)
56. Min, H., Mitra, A., Oswald, S.: Competitive benchmarking of health care quality using the analytic hierarchy process: an example from Korean cancer clinics. Soc.-Econ. Plan. Sci. **31** (2), 147–159 (1997)
57. Chang, S.J., Hsiao, H.C., Huang, L.H., Chang, H.: Taiwan quality indicator project and hospital productivity growth. Omega **39**(1), 14–22 (2011)
58. Teng, C.I., Ing, C.K., Chang, H.Y., Chung, K.P.: Development of service quality scale for surgical hospitalization. J. Formos. Med. Assoc. **106**(6), 475–484 (2007)
59. Shieh, J.I., Wu, H.-H., Huang, K.K.: A DEMATEL method in identifying key success factors of hospital service quality. Knowl.-Based Syst. **23**(3), 277–282 (2010)

60. Kubler, S., Robert, J., Derigent, W., Voisin, A., Le Traon, Y.: A state-of the-art survey & testbed of Fuzzy AHP (FAHP) applications. Expert Syst. Appl. **65**, 398–422 (2016)
61. Büyüközkan, G., Cifci, G., Güleryüz, S.: Strategic analysis of healthcare service quality using fuzzy AHP methodology. Expert Syst. Appl. **38**(8), 9407–9424 (2011)
62. Ju, Y., Wang, A., Liu, X.: Evaluating emergency response capacity by Fuzzy AHP: and 2-tuple fuzzy linguistic approach. Expert Syst. Appl. **39**(8), 6972–6981 (2012)
63. Shahin, A., Janatyan, N.: Estimation of customer dissatisfaction based on service quality gaps by correlation and regression analysis in a travel agency. Int. J. Bus. Manag. **6**(3), 99–108 (2011)
64. Zaim, H., Bayyurt, N., Zaim, S.: Service quality and determinants of customer satisfaction in hospitals: Turkish experience. Int. Bus. Econ. Res. J. **9**(5), 51 (2010)
65. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: An empirical test of the extended gaps model of service quality. In: Marketing Science Institute Working Paper, pp. 90–122 (1990)
66. Bitner, M.J., Hubbert, A.R.: Encounter satisfaction versus overall satisfaction versus service quality: the consumer's voice. In: Rust, R.T., Oliver, R.L. (eds.) Service Quality: New Directions in Theory and Practice, pp. 72–94. Sage Publications, Thousand Oaks (1994)
67. Prakash, A., Mohanty, R.P., Kallurkar, S.P.: Service quality modelling for life insurance business using neural networks. Int. J. Prod. Qual. Manag. **7**(3), 263–286 (2011)
68. Grönroos, C.: Strategic Management and Marketing in the Service Sector. Swedish School of Economics and Business Administration, Helsingfors (1982)
69. Akter, S., Hani, U., Upal, M.: Service quality perception and satisfaction: a study in suburban public hospitals in Bangladesh. Int. J. Serv. Res. February (Special Issue), 125–145 (2008)
70. Jager, D.J., Plooy, D.T.: Are public hospitals responding to tangible and reliable service related needs of patients in the new South Africa? J. Manag. Policy Pract. **12**(2), 105–119 (2011)
71. Kuo, H.P.: Fuzzy comprehensive evaluation applied in the performance assessment of service quality of insurance companies in Taiwan. J. Am. Acad. Bus. **16**(1), 84–92 (2010)
72. Priporas, C.V., Laspa, C., Kamenidou, I.: Patient satisfaction measurement for in hospital services: a pilot study in Greece. J. Med. Mark. **8**(4), 325–340 (2008)
73. Sohail, M.: Service quality in hospitals, more favourable than you might think. Manag. Serv. Qual. **13**(2), 197–206 (2003)
74. Sajid, M.S., Baig, M.K.: Quality of health care, an absolute necessity for public satisfaction. Int. J. Health Care Qual. Assur. **20**(6), 545–548 (2007)
75. Jabnoun, N., Chacker, M.: Comparing the quality of private and public hospitals. Manag. Serv. Qual. **13**(4), 290–299 (2003)
76. Boshoff, C., Gray, B.: The relationships between service quality, customer satisfaction and buying intentions in the private hospital industry. South Afr. J. Bus. Manag. **35**(4), 27–38 (2004)
77. Ramsaran-Fowdar, R.R.: The relative importance of service dimensions in a healthcare setting. Int. J. Health Care Qual. Assur. **21**(1), 104–124 (2008)
78. Padma, P., Rajendran, C., Sai, L.P.: A conceptual framework of service quality in healthcare perspectives of Indian patients and their attendants. Benchmark. Int. J. **16**(2), 157–191 (2009)
79. Aagja, J.P., Garg, R.: Measuring perceived service quality for public hospitals (PubHosQual) in the Indian context. Int. J. Pharm. Healthc. Mark. **4**(1), 60–83 (2010)
80. Ladhari, R.: Alternative measure of service quality: a review. J. Manag. Serv. Qual. **18**(1), 65–86 (2008)
81. Chen, M.F., Tzeng, G.H., Tang, T.: Fuzzy MCDM approach for evaluation of expatriate assignments. Int. J. Inf. Technol. Decis. Mak. **4**(2), 277–296 (2005)

82. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning: parts 1, 2 and 3. Inf. Sci. **8**, 199–249 (1975)
83. Sinimole, K.R.: Performance evaluation of the hospital services–a fuzzy analytic hierarchy process model. Int. J. Prod. Qual. Manag. **101**, 112–130 (2012)
84. Saaty, T.L.: Multicriteria Decision Making: The Analytic Hierarchy Process. RWS Publications, Pittsburgh (1990)
85. Davis, M.A.P.: A multi-criteria decision model application for managing group decisions. J. Oper. Res. Soc. **45**(1), 47–58 (1994)

Design of Hybrid Neuro-Fuzzy Controller for Magnetic Levitation Train Systems

Yakob Kiros Teklehaimanot[1(✉)], Dereje Shiferaw Negash[2,3], and Endrias Alemayehu Workiye[3]

[1] Mekelle University, Mekelle, Ethiopia
yakobkiros81@gmail.com
[2] Addis Abeba University, Addis Abeba, Ethiopia
dnderejesh@gmail.com
[3] Adama Science and Technology, Adama, Ethiopia
endrias1873@gmail.com

Abstract. Maglev is a system in which the train runs levitated from the guideway by using electromagnetic forces between superconducting magnets (ferromagnetic materials) on board of the train and coils on the ground. The magnetic levitation train system are based on two types, electrodynamics suspension (EDS) and electromagnetic suspension (EMS). EDS is based on repulsive forces acting on a magnet and is inherently stable system and even has well robustness in many cases with open loop control. In this paper, we have assumed the EMS based train system. The electromagnetic suspension system is based on attractive forces acting on a magnet and is complex, unstable and the model is strongly nonlinear. In addition, due to the external disturbances like wind, the unbalanced magnetic forces between the guideway and the train, and parameter perturbation, the system model has greater uncertainty. This paper presents a hybrid neuro-fuzzy controllers for the magnetic levitation train system. The controllers are designed to bring the magnetic levitation system in a stable region by keeping the train suspended in the air in the required position in the presence of uncertainties. PID controller is used to generate the data which requires to train the hybrid controllers. The performance and robustness of the controllers have been compared by simulating the system with disturbances. After implementing and validating, the Matlab simulation results show that the performance of the system (overshoot, settling time, rise time and peak response) have improved and the controller have good robustness and adaptability.

Keywords: Magnetic levitation · EDS · EMS · Fuzzy controller · Hybrid neuro-fuzzy controller · Taylor's series linearization · Robustness · Adaptability

1 Introduction

Maglev train is a system which uses magnetic fields to levitate so there is no physical contact between the train and the track (guideways). The motion of

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 119–133, 2019.
https://doi.org/10.1007/978-3-030-26630-1_10

the maglev train is based on electromagnetism and magnetic fields. Electromagnetism is magnetism produced by an electric current [1,2]. So the basic idea behind electromagnets are that you can create a magnetic field by running an electric current through a conductor (wire). So the maglev train can levitate above it track or guideway, and propel forward using this magnetic field. There are two types of levitation s namely, EMS (Electromagnetic suspension) and EDS (Electrodynamics suspension). The electromagnetic suspension EMS uses attractive force system to levitate where as EDS uses repulsive forces to levitate [3,4].

The main advantage of EDS Maglev systems is that they are naturally stable minor narrowing in distance between the track and the magnets, which creates the strong forces to repel the magnets back to their original position. But, EMS maglev train system is inherently unstable. At high speed, it becomes difficult to maintain the correct distance between train and guideway. If this distance cannot be kept, the train will fail to levitate and come grinding to a halt. To protect this, EMS needs a complex feedback control system to ensure the train is always stable [3,5].

Additionally in EMS system, the air gap between the guide ways and train magnets is very small (8–10) mm, whereas the air gap in the EDS system may be as large as (100–150) mm. The small air gap of the EMS system implies much more stringent controls to maintain this small gap. So, the EMS magnetic levitation based train system is not stable because of this and different disturbances like wind and other factors. In this paper, a hybrid neuro-fuzzy controller is designed to control the levitation distance which is around 8–10 mm. The controller is designed in order to keep stable the suspension gab between the electromagnet and electromagnet rail (levitation gab) [4].

2 Mathematical Modeling

Mathematical modeling of the maglev train system is developed by considering different assumption. Since the train is levitated on the air, there is drag force which affects the stability of the train like wind.later on the simulation, the response is simulated by adding different disturbance to the system modeling and shown the results by comparing with and without disturbances.

Consider the transrapid maglev train system, which is simplified as single mass system on a rigid guide-way as shown in Fig. 1:

Where, mg is train weight (including the electromagnets), S_M is vertical gab between the guideway and the train, S_0 is vertical gab between the guideway and the equilibrium position of the train, $L(S_M)$ is coil inductance, R is coil resistance and U_M is input voltage.

2.1 Non-linear Modeling

The electromagnetic force produced by current is given by the Kirchhoff's voltage law [7,8]

$$u(t) = V_R + V_L = i_M R + \frac{d}{dt}(L(s_M)i_M) \tag{1}$$

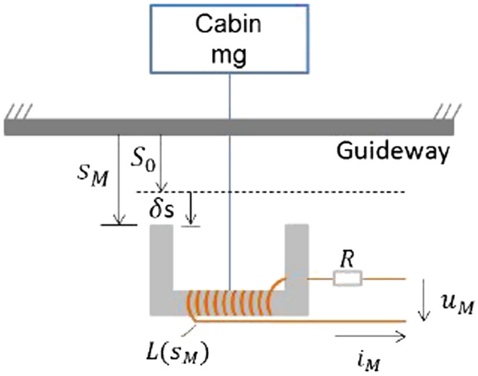


Fig. 1. Simplified Model of train levitation system [6]

The Net force is given by Newton's 3^{rd} law of motion while neglecting drag force of the air on the train [1]

$$f_{net} = f_g - f_m \tag{2}$$

Where,

$$f_m = \frac{\mu_0 N^2 A}{2} \left(\frac{i_M}{s_M}\right)^2 = C \left(\frac{i_M}{s_M}\right)^2 \tag{3}$$

and the net force is,

$$f_{net} = f_g - f_m = mg - C \left(\frac{i_M}{s_M}\right)^2 \tag{4}$$

The non-linear system model will be

$$\begin{aligned} m\ddot{s_M} &= mg - C \left(\frac{i_M}{s_M}\right)^2 \\ \dot{i_M} &= \frac{i_M}{s_M}\dot{s_M} - \frac{Ri_M s_M}{C} + \frac{Us_M}{C} \\ \dot{s_M} &= v \end{aligned} \tag{5}$$

State Space form of the non-linear model

Let $x_1 =$ Current (i_M), $x_2 =$ Velocity (v) and $x_3 =$ Position (s_M)

$$\begin{aligned} \dot{x_1} &= \frac{x_1 x_2}{x_3} + \frac{x_3}{C}(U - Rx_1) \\ \dot{x_2} &= -\frac{Cx_1^2}{mx_3^2} + g \\ \dot{x_3} &= x_2 \end{aligned} \tag{6}$$

2.2 Linear Modeling

To carryout controller design and analysis of maglev train system, the obtained non-linear model has to be linearized. Such linearization is done in the equilibrium point which can be calculated from,

$f_g = f(s_M, i_M)$ at S_0, I_0

At equilibrium, the force due to gravity f_g and the magnetic force are equal and opposite to each other so that the train levitates. We can linearize the model using Taylor's series expansion at $f(s_M, i_M)$ around the equilibrium point (S_0, I_0), where $s_M = S_0 + \delta s$ and $i_M = I_0 + \delta i$ [7,9,10].

Putting the linearized model into state space representation,

$$\begin{aligned} \begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \dot{x_3} \end{bmatrix} &= \begin{bmatrix} -\frac{RS_0}{C} & \frac{I_0}{S_0} & 0 \\ -\frac{CI_0}{mS_0^2} & 0 & \frac{CI_0^2}{mS_0^3} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \frac{S_0}{C} \\ 0 \\ 0 \end{bmatrix} [U], \\ y &= \begin{bmatrix} 0 \ 0 \ 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned} \tag{7}$$

The values of the System parameters which are used in simulation is shown in Table 1.

Table 1. System parameter values [6]

$m(kg)$	N	$A(mm^2)$	$S_0(mm)$	$L_0(h)$	$R(\Omega)$	$U(V)$	$I_0(A)$
10000	1000	80	10	0.1	1	140	140

Initial conditions, $S_0 = 10\,\text{mm}$ and $I_0 = 140\,\text{A}$. Substituting all the values, state space model of the system will be

$$\begin{aligned} \dot{x} &= \begin{bmatrix} -10 & 14 & 0 \\ -0.00014 & 0 & 0.00196 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 10 \\ 0 \\ 0 \end{bmatrix} U, \\ y &= \begin{bmatrix} 0 \ 0 \ 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned} \tag{8}$$

3 Controller Design

The controller simulation is implemented via four types of Adaptive Neuro-Fuzzy Inference System (ANFIS) control strategies. The control problem is to design a controller which computes the control signal (u) to magnetic levitation train system so as to keep the train stable at the required levitation position.

3.1 Hybrid Neuro-Fuzzy System

Fuzzy systems are popular in industrial application, but the development of a fuzzy system with good performance is not an easy task. The problem of finding membership functions and appropriate rules is frequently a tiring process of attempt and error. This lead to the idea of applying learning algorithms to the fuzzy systems. The neural networks, that have efficient learning algorithms, had been presented to automate or to support the development of tuning fuzzy systems. Neural networks and fuzzy systems can be combined to join its advantages and to cure its individual luck [11,12].

While neural networks are low-level computational structures that perform well when dealing with raw data, fuzzy logic deals with reasoning on a higher level, using linguistic information acquired from domain experts. However, fuzzy systems lack the ability to learn and cannot adjust themselves to a new environment [11]. On the other hand, although neural networks can learn, they are opaque to the user. Integrated neuro-fuzzy systems can combine the parallel computation and learning abilities of neural networks with the human-like knowledge representation and explanation abilities of fuzzy systems. As a result, neural networks become more transparent, while fuzzy systems become capable of learning [13–15].

3.2 ANFIS Controller

Adaptive Neuro-Fuzzy Inference System (ANFIS) is one of the most successful schemes which combine the benefits of these two powerful paradigms into a single capsule. An ANFIS works by applying neural learning rules to identify and tune the parameters and structure of a Fuzzy Inference System (FIS) [16]. According to the neuro-fuzzy approach, a neural network is proposed to implement the fuzzy system, so that structure and parameter identification of the fuzzy rule base are accomplished by defining, adapting and optimizing the topology and the parameters of the corresponding neuro-fuzzy network, based only on the available data. The network can be regarded both as an adaptive fuzzy inference system with the capability of learning fuzzy rules from data, and as a connectionist architecture provided with linguistic meaning [11,17].

3.3 Architecture of ANFIS

ANFIS is an adaptive network which is a multilayer feed-forward network composed of nodes connected by directed links, in which each node performs a particular function on its incoming signals to generate a single node output. Each link in an adaptive network specifies the direction of signal flow from one node to another; no weights is associated with the link. More specifically, the configuration of an adaptive network performs a static node function on its incoming signals to generate a single node output and each node function is a parameterized function with modifiable parameters; by changing these parameters, the node functions as well as the overall behavior of the adaptive network, are changed [18,19]. Figure 2 shows entire system architecture. With input/output data for given set of parameters, the ANFIS method models a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using either a back-propagation algorithm alone, or with a least squares type of method. The main objective of the ANFIS is to determine the optimum values of the equivalent fuzzy inference system parameters by applying a learning algorithm. The parameter optimization is done in such a way during the training session that the error between the target and the actual output is minimized [19].

The parameters to be optimized in ANFIS are the premise parameters and consequent parameters. In order to reduce the error measure, any of several optimization routines can be applied after constituting membership functions. The parameter set of an adaptive network allows fuzzy systems to learn from the data they have been trained. This thesis assumes that adaptive system under consideration has two inputs V_1 and V_2 and one output f. Let us scrutinize a first order Takagi, Sugeno and Kang (TSK) fuzzy inference system containing two rules [14].

$$\text{Rule 1:If}(v \text{ is } V_1) \text{ and } (d \text{ is } D_1) \text{ then } f_1 = p_1 v + q_1 d + r_1$$
$$\text{Rule 2:If}(v \text{ is } V_2) \text{ and } (d \text{ is } D_2) \text{ then } f_2 = p_2 v + q_2 d + r_2$$

Where p_1, p_2, q_1, q_2, r_1 and r_2 are linear parameters and V_1, V_2, D_1 and D_2 are non linear parameters, in which V_1 and D_1 are the membership functions of ANFIS (antecedent). p_1, q_1 and r_1 are the consequent parameters. To reflect adaptive capabilities, we use both circle and square. A circle indicates fixed node whereas square indicates adaptive node i.e. the parameter can be changed during adapting or training [14,18].

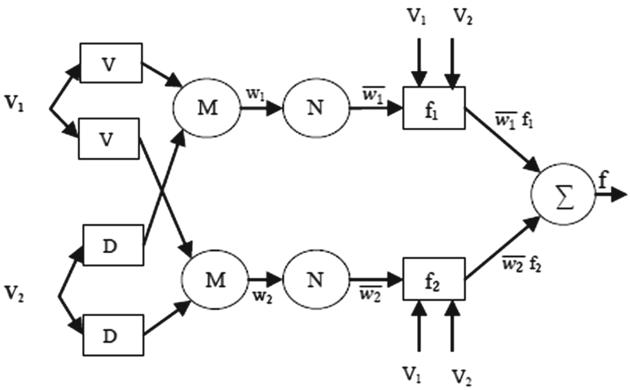


Fig. 2. Basic architecture of ANFIS [14]

It can be observed that ANFIS architecture consists of two adaptive layers, namely the first layer and the fourth layer. The three modifiable parameters a_i, b_i, c_i are so-called premise parameter in first layer and in the fourth layer, there are also three modifiable parameters p_i, q_i, r_i pertaining to the first order polynomial. These parameters are so-called consequent parameters [16].

3.4 Learning Algorithm of ANFIS

The basic idea of Neuro-fuzzy adaptive learning techniques is to provide a method for the fuzzy modeling procedure to learn information about a data set, in order to automatically compute the membership function parameters that best allow the associated fuzzy inference system (FIS) to track the given input/output data [14,16].

The membership function parameters are tuned using a combination of least squares estimation and back-propagation algorithm for membership function parameter estimation. These parameters associated with the membership functions will change through the learning process similar to that of a neural network. Their adjustment is facilitated by a gradient vector, which provides a measure of how well the FIS is modeling the input/output data for a given set of parameters. Once the gradient vector is obtained, any of several optimization routines could be applied in order to adjust the parameters so as to reduce error between the actual and desired outputs. This allows the fuzzy system to learn from the data it is modeling. The approach has the advantage over the pure fuzzy paradigm that the need for the human operator to tune the system by adjusting the bounds of the membership functions [18–20]. To design ANFIS controller, first it has to be trained using input/output data. So, here data is generated using PID controller.

3.5 Controller Design Objectives

Since maglev train system is very nonlinear and unstable system, a controller is designed to stabilize it despite of any disturbances which affects the system. In this paper, the Adaptive NeuroFuzzy Inference System (ANFIS) which is a hybrid neuro-fuzzy controller is designed. ANFIS is a controller which has the ability to learn and adapt. The suspension control is always the technical difficulty of EMS Maglev train. ANFIS controller algorithm should meet the following requirements:

– Good adaptability for the uncertainty suspension system model
– Good robustness for the interference of the external system.

3.6 Sample Generated Data

We have used ANFIS controller with two inputs and one output. The data is generated error, error derivative (change of error) and the control signal using PID controller.

Training vector has two input parameters that of error E and the rate of change of error EC. In this simulation, the PID controller is applied to the magnetic levitation train system to get input and output data. The sampling time is set to 0.1 s. 1000 sample data is selected as the ANFIS controller training data as shown in Table 2 The simulink structure of magnetic levitation train controller based on adaptive neural network simulation is shown in Fig. 3.

Table 2. Generated data using PID

error	errorderiv	control signal
1	0	-72000
0.988756	-0.209415116	-36081.04789
0.937753	-0.618734304	-16120.37842
0.850472	-0.979919494	-4296.599869
0.741668	-1.088047024	2745.402569
0.624763	-1.169045816	6742.562558
0.509366	-1.153971941	8744.148387
0.401517	-1.078490552	9446.172708
0.304538	-0.969785574	9327.959348
0.21985	-0.846884786	8721.653212
0.147616	-0.722333785	7854.811132
0.087227	-0.603893431	6879.724237
0.037633	-0.495938463	5894.721591
-0.00242	-0.400528291	4959.923811
-0.03424	-0.318197331	4108.897101
-0.05909	-0.248523865	3357.208366
-0.07814	-0.190530246	2708.629869
-0.09244	-0.142957455	2159.570329
-0.10289	-0.104447641	1702.17989
-0.11025	-0.073660437	1326.474761
-0.11519	-0.049342475	1021.746651
-0.11822	-0.030364566	777.4584003
-0.1198	-0.015737167	583.7771224
-0.12026	-0.004611794	431.8572424
-0.11988	0.003726162	313.9557979
-0.1189	0.009869415	223.4394926
-0.11747	0.014302997	154.7257311
-0.11573	0.017419012	103.1869965
-0.11377	0.019530868	65.03845638
-0.11168	0.020886397	37.22180792
-0.10952	0.021679543	17.29346262
-0.10731	0.022060519	3.3217344
-0.1051	0.022144453	-6.204655785
-0.10289	0.022018632	-12.45600116
-0.10072	0.02174848	-16.32918516
-0.09858	0.021382429	-18.50422601
-0.09648	0.020955833	-19.49091202
-0.09444	0.020494099	-19.66677227
-0.09243	0.020015138	-19.3076312
-0.09048	0.019531289	-18.61192642
-0.08858	0.019050799	-17.71986014
-0.08672	0.018578951	-16.72832737
-0.08491	0.018118923	-15.70243345
-0.08314	0.01767242	-14.68428761
-0.08141	0.017240155	-13.69964439
-0.07973	0.016822181	-12.76286173
-0.07809	0.016418142	-11.88055636
-0.07649	0.016027444	-11.05426104
-0.07492	0.015649376	-10.28232534
-0.07339	0.015283181	-9.561249632
-0.0719	0.014928116	-8.88659933
-0.07044	0.014583474	-8.253612715
-0.06902	0.014248605	-7.657588295
-0.06763	0.01392292	-7.094116474
-0.06627	0.013605896	-6.559203626
-0.06494	0.013297067	-6.049323878
-0.06364	0.012996025	-5.561424108
-0.06237	0.012702408	-5.092900299
-0.06112	0.0124159	-4.641557863
-0.05991	0.012136221	-4.205564516
-0.05872	0.011863121	-3.783401301
-0.05756	0.011596376	-3.373815293
-0.05643	0.011335786	-2.975775999
-0.05532	0.011081167	-2.588436481
-0.05424	0.010832351	-2.211099555
-0.05318	0.010589181	-1.843188989
-0.05215	0.010351512	-1.484225363
-0.05113	0.010119206	-1.133806128
-0.05014	0.009892131	-0.791589343
-0.04918	0.009670164	-0.457280567
-0.04823	0.009453185	-0.130622394
-0.04731	0.009241078	0.188613801
-0.0464	0.009033732	0.500634325
-0.04552	0.008831037	0.805628168
-0.04466	0.00863289	1.103770847
-0.04381	0.008439188	1.39522733
-0.04299	0.00824983	1.680154231
-0.04218	0.008064719	1.958701458
-0.04139	0.00788376	2.231013426
-0.04062	0.007706861	2.497229948
-0.03987	0.007533929	2.757486881
-0.03913	0.007364877	3.011916586
-0.03841	0.007199617	3.260648266
-0.03771	0.007038064	3.503808196
-0.03702	0.006880136	3.741519891
-0.03635	0.006725751	3.973904225
-0.03569	0.00657483	4.201079519
-0.03505	0.006427296	4.423161604
-0.03442	0.006283071	4.640263872
-0.03381	0.006142083	4.852497319
-0.03321	0.006004258	5.059970584
-0.03262	0.005869526	5.262789982
-0.03205	0.005737817	5.461059542
-0.03148	0.005609064	5.654881041
-0.03094	0.005483199	5.844354041
-0.0304	0.005360159	6.029575924

After training data, next step is establishment of ANFIS network structure. For ANFIS model, first step is to determine the output and input variables membership functions. The system uses triangular and gaussian shaped memberships function to represent the first input membership function.

In next step, for adjusting the parameters of ANFIS fuzzy model the hybrid learning algorithm is chosen. ANFIS learning methods integrated use of gradient descent learning method and least square method (LSE).

After selecting initial membership function, Sugeno fuzzy model identification methods and hybrid learning algorithm are selected. By using the grid partition method fuzzy rules and modified membership function are identified. Closing value of the error (Error Tolerance) is set to 0 and the training times (epoch) is set to 100.

In final step, the FIS output can be exported to workspace for the ANFIS controller to make its rules The ANFIS controller is attached to the plant and simulated.

If the operating input-output data are outside their training data range, estimator will not operate accurately. As a result, the training data set should possess sufficient operational range including the maximum and minimum values for input output variables.

3.7 Design of ANFIS Controller

The ANFIS controller is designed using Matlab/Simulink software. In this Paper, ANFIS with two input and one output is used. Inputs to the ANFIS controller are error and change in error which are defined as follow:

$$\begin{aligned} e(s) &= X_{ref} - X_{act} \\ \triangledown e(s) &= \tfrac{d}{dt} e(s) \end{aligned} \tag{9}$$

Where,

X_{ref} is the reference position (levitation) of the train system
X_{act} is the actual position (levitation) of the train system

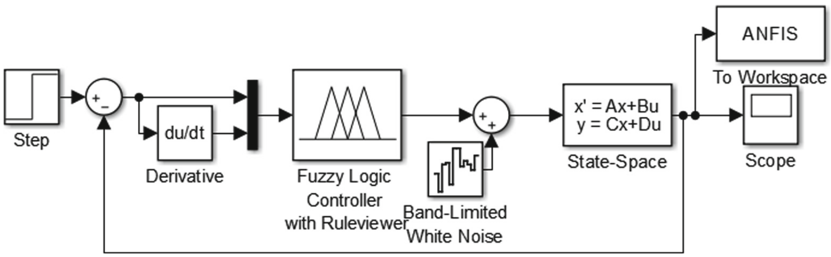


Fig. 3. Simulation Model of Magnetic levitation train system based on ANFIS

As shown in Fig. 3, Band-Limited White Noise is used as input disturbance to test the robustness and adaptability of the controllers.

In this paper, four types of ANFIS controllers have designed as shown in Table 3.

Table 3. Input-output membership functions

Controller type	Input membership function	Output membership function
ANFIS1	Triangular	Constant
ANFIS2	Triangular	Linear
ANFIS3	Gaussian	Constant
ANFIS4	Gaussian	Linear

All types of member function of error and change of error are automatically generated base on the data trained and the current value of the input signals. Membership function of the error and change of error are shown in Figs. 4 and 5 respectively for ANFIS1 and ANFIS2. And membership function of the error and change of error are shown in Figs. 6 and 7 respectively for ANFIS3 and ANFIS4.

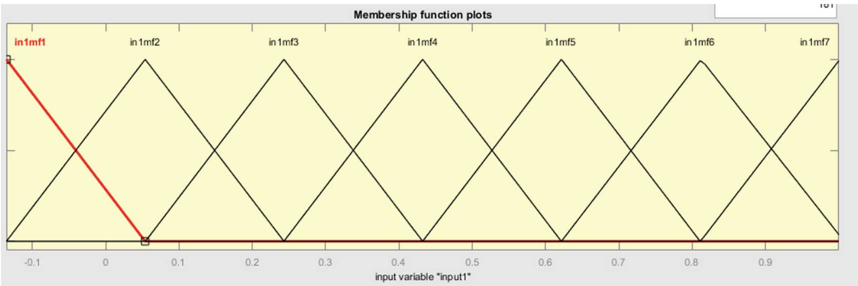


Fig. 4. Error membership function

4 Result and Discussion

The simulation results of the controllers have been compared and discussed their performance. The Maglev system is simulated using the Matlab/Simulink environment. The dynamic performance of ANFIS and PID controllers have been discussed under different scenarios.

The graph in Fig. 8 shows the overall step response of the magnetic levitation train system for controllers considering no disturbance and Fig. 9 shows step response when there is an input disturbance. From Fig. 8 and Table 4, it can

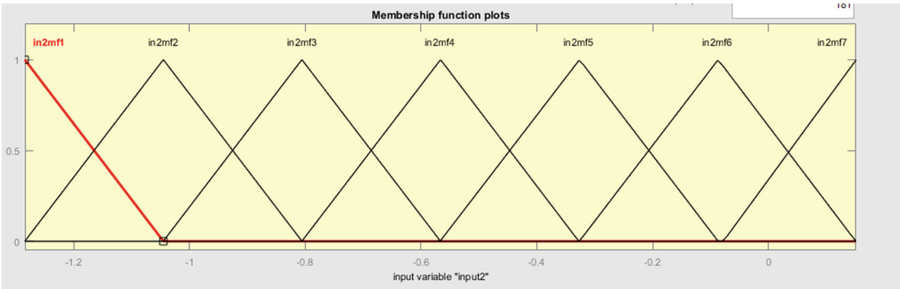


Fig. 5. Membership function for change of error

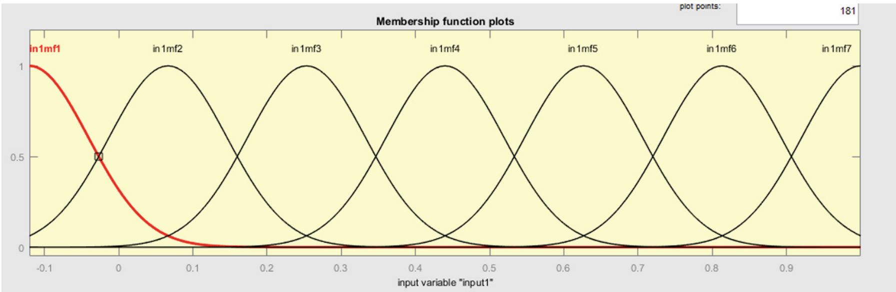


Fig. 6. Gaussian error membership function

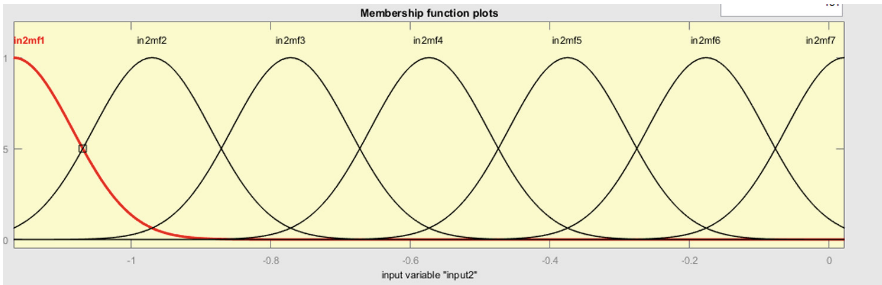


Fig. 7. Gaussian membership function for change of error

observe that ANFIS1 and ANFIS3 have small settling time, overshoot, and peak response but they have rise time which is greater than ANFIS2 and ANFIS4. This means that the response of ANFIS1 and ANFIS3 is slower than ANFIS2 and ANFIS4 until it reaches 90% but it settles to the desired state with in short settling time. Comparing ANFIS1 and ANFIS3, ANFIS3 has small overshoot, settling time and peak response and it has good performance than ANFIS1. Comparing ANFIS2 and ANFIS4, they have almost the same response as shown in Fig. 8 except ANFIS2 has smaller settling time than ANFIS4.

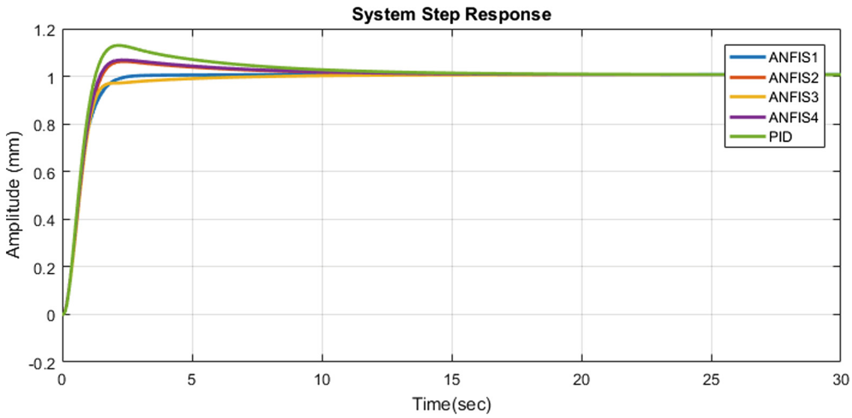


Fig. 8. ANFIS and PID controllers step response without disturbance

Table 4. Performance of ANFIS and PID controllers without disturbance

Controllers	Setting time (s)	Rise time (s)	Overshoot (%)	Peak response (mm)	Steady state response (mm)
PID	6.784	0.830	11.798	1.130	1.00
ANFIS1	1.667	1.127	0.268	1.008	1.00
ANFIS2	3.621	0.958	5.851	1.012	1.00
ANFIS3	1.552	1.014	−0.737	1.007	1.00
ANFIS4	4.4425	0.936	5.851	1.010	1.00

The performance of ANFIS controller is shown in Table 5. From Table 5, we can observe that all ANFIS controller have almost the same steady state response. But ANFIS3 has small overshoot, settling time and peak response which is good performance than the others. And ANFIS4 has bad performance than the others. The settling time and rise time of ANFIS2 is smaller than ANFIS1 but ANFIS2 has higher overshoot and peak response than ANFIS1. Since the overshoot measures how much the response exceeds the target point and if the response has large overshoot, it is bad performance which bring system instability. This implies ANFIS1 has better performance than ANFIS2 with large disturbance is added to maglev train system as load disturbance.

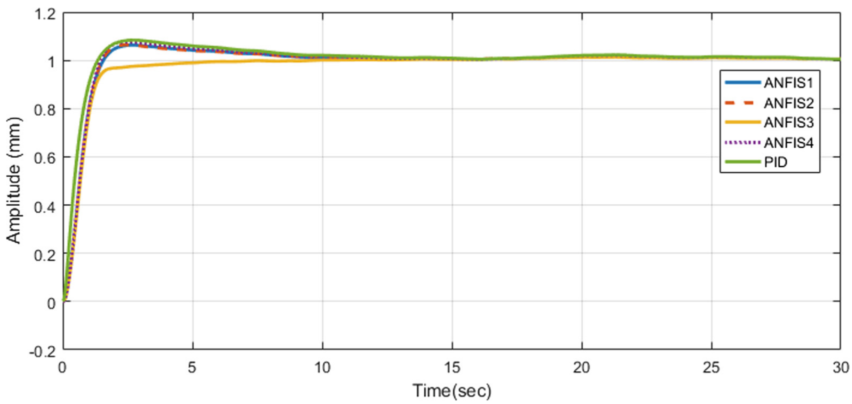


Fig. 9. ANFIS and PID controllers step response with White Noise input disturbance

Table 5. Performance of ANFIS and PID controllers with white noise disturbance

Controllers	Setting time (s)	Rise time (s)	Overshoot (%)	Peak response (mm)	Steady state response (mm)
PID	7.780	1.078	13.068	1.148	1
ANFIS1	4.08	0.988	4.737	1.062	1
ANFIS2	3.851	0.958	5.851	1.064	1
ANFIS3	1.552	1.035	−0.817	1.008	1
ANFIS4	4.77	0.957	5.851	1.075	1

5 Conclusion

ANFIS controllers have been designed with different types of input/output membership functions. Based on the input/output types of membership functions, four types of ANFIS controllers have designed and compared their performance with PID controller.

When there no disturbance which affect the train system, from Fig. 8 and Table 4 shows that All ANFIS controllers have better performance than PID. Comparing between ANFIS controllers, ANFIS3 has better performance than the other ANFIS controllers. ANFIS3 has designed with Gaussian input membership function and constant output function. This shows that Gaussian membership function has better representation of the input data than triangular and constant type membership function has better representation of output data than linear. When there is input disturbance which is considered as load variation only in this paper and modeled as band limited white noise, as shown in Fig. 9 and Table 5, ANFIS3 has better performance than the others. And ANFIS4 has bad performance than the other ANFIS controllers.

When there is process disturbance which brings system parameter perturbation and this disturbance is modeled as random number in this paper, from matlab simulation results shows that ANFIS3 has better performance than the other ANFIS controllers.

The overall performance of the controllers under input disturbance, process disturbance and without disturbance, it has been observed that ANFIS controllers have better performance than PID. One of the main advantages of the ANFIS scheme is that, it is computationally efficient, increases the dynamic performance and provides good stabilization when there is a sudden fluctuation in the levitation of the train. This shows the excellent response of the proposed control scheme as it has the learning capability using the neural networks.

When we have compared ANFIS controllers with PID, we have shown that ANFIS has minimum settling time, Overshoot and the peak response than PID controller. From this observation, the speed of convergence of the response increased and Superior performance, robustness, and efficiency of the proposed method have been proved through extensive simulation results. The proposed ANFIS controller is a powerful approach to retrieve the deceptiveness of a maglev system. The results show that this approach is robustness and suitable for optimizing various control problems if the controller is trained well. Finally, The simulation results show that the designed ANFIS controllers meets the requirements for the robust stability and adaptability performance. And among ANFIS controllers, we have shown that ANFIS3 controller has good robustness and adaptability than the other ANFIS controllers.

References

1. Sharkawy, A.B., Abo-Ismail, A.A.: Intelligent control of magnetic levitation system. J. Eng. Sci. Assiut Univ. **37**(4), 909–924 (2009)
2. Liu, Z., Long, Z., Li, X.: Maglev train overview. In: Liu, Z., Long, Z., Li, X. (eds.) Maglev Trains, pp. 1–28. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-45673-6_1
3. Tandan, G.K., Sen, P.K., Sahu, G., Sharma, R., Bohidar, S.: A review on development and analysis of maglev train. Int. J. Res. Advent Technol. **3**(12), 14–17 (2015)
4. Bajuri, M.F.: Modelling magnetic levitation (maglev) train. Ph.D. thesis, UMP (2012)
5. Magnetic levitation train system. http://techdatacare.blogspot.com/2011/12/magnetic-levitation.html. Accessed 1 Sept 2017
6. Cabral, T., Chavarette, F.: Dynamics and control design via LQR and SDRE methods for a maglev system. Int. J. Pure Appl. Math. **101**(2), 289–300 (2015)
7. Choudhary, S.K.: Robust feedback control analysis of magnetic levitation system. WSEAS Trans. Syst. **13**(27), 285–291 (2014)
8. Pati, A., Pal, V.C., Negi, R.: Design of a 2-DoF control and disturbance estimator for a magnetic levitation system. Eng. Technol. Appl. Sci. Res. **7**(1), 1369 (2016)
9. Ahmad, I., Javaid, M.A.: Nonlinear model & controller design for magnetic levitation system. In: Recent Advances in Signal Processing, Robotics and Automation, pp. 324–328 (2010)
10. Sun, Y., Qiang, H., Lin, G., Ren, J., Li, W.: Dynamic modeling and control of nonlinear electromagnetic suspension systems. Chem. Eng. Trans. **46**, 1039–1044 (2015)

11. Al-Hmouz, A., Shen, J., Al-Hmouz, R., Yan, J.: Modeling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for mobile learning. IEEE Trans. Learn. Technol. **5**(3), 226–237 (2012)
12. Panda, G., Panda, S., Ardil, C.: Hybrid neuro fuzzy approach for automatic generation control of two-area interconnected power system. Int. J. Comput. Intell. **5**(1), 80–84 (2009)
13. Kaur, A., Kaur, A.: Comparison of fuzzy logic and neuro-fuzzy algorithms for air conditioning system. Int. J. Soft Comput. Eng. **2**(1), 417–420 (2012)
14. Walia, N., Singh, H., Sharma, A.: ANFIS: adaptive neuro-fuzzy inference system - a survey. Int. J. Comput. Appl. **123**(13), 32–38 (2015)
15. Vieira, J., Dias, F.M., Mota, A.: Neuro-fuzzy systems: a survey. In: 5th WSEAS NNA International Conference on Neural Networks and Applications, Udine, Italia (2004)
16. Allaoua, B., Laoufi, A., Gasbaoui, B., Abderrahmani, A.: Neuro-fuzzy DC motor speed control using particle swarm optimization. Leonardo Electron. J. Pract. Technol. **15**, 1–18 (2009)
17. Kusagur, A., Kodad, S., Ram, B.V.S.: Modeling, design & simulation of an adaptive neuro-fuzzy inference system (ANFIS) for speed control of induction motor. Int. J. Comput. Appl. **6**(12), 29–44 (2010)
18. Sivakumar, R., Sahana, C., Savitha, P.: Design of ANFIS based estimation and control for mimo systems. Int. J. Eng. Res. Appl. **2**(3), 2803–2809 (2012)
19. Yousef, H.A., Khalfan, A.K., Albadi, M.H., Hosseinzadeh, N.: Load frequency control of a multi-area power system: an adaptive fuzzy logic approach. IEEE Trans. Power Syst. **29**(4), 1822–1830 (2014)
20. Rashid, U., Jamil, M., Gilani, S.O., Niazi, I.K.: LQR based training of adaptive neuro-fuzzy controller. In: Bassis, S., Esposito, A., Morabito, F.C., Pasero, E. (eds.) Advances in Neural Networks. SIST, vol. 54, pp. 311–322. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-33747-0_31

Wireless and Mobile Computing Track

Optimization of Electrical Tilt for Addis Ababa LTE Deployment Scenario

Taddege Assefa Beyene[1(✉)] and Beneyam Berehanu Haile[2(✉)]

[1] Ethio Telecom, Addis Ababa, Ethiopia
taddege.assefa@ethiotelecom.et
[2] Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia
beneyamb.haile@aait.edu.et

Abstract. Demand for mobile broadband connectivity is significantly increasing and mobile network operators are continuously improving their network by introducing new technologies to accommodate the demand. After mobile network is operational for the first time, to maximize network performance, it is common to perform further optimization of initially planned and configured network parameters. One important category of such network parameters that incur significant network performance impact is antenna parameters including antenna tilt. Optimization of antenna tilt plays a major role in improving network coverage, capacity, load balancing and mitigating interference of mobile networks. To deliver mobile data services with better quality of experience, long term evolution (LTE) network has been operational in Addis Ababa, Ethiopia since 2015. In this paper, we evaluate performance of this Addis Ababa LTE network from antenna tilt perspective. Based on the evaluation, we also perform optimization of the antenna tilt by applying educated trial and error optimization approach that is focused to address worst performance areas. The performance evaluation and optimization of the network are undertaken using WinProp simulation tool. Results show that there is still a room to optimize antenna tilt further for better performance.

Keywords: LTE · Electrical tilt · SINR · Throughput · Network simulation · Addis Ababa · Ethiopia

1 Introduction

To address mainly better mobile data rate and network capacity demand, standardization of long term evolution (LTE) has been launched in November 2004 by the third generation partnership project (3GPP) [1]. Specific goals include improving efficiency, lowering costs, improving services, making use of new spectrum opportunities, and better integration with other open standards. LTE is selected as the next generation broadband wireless technology for 3GPP and 3GPP2. The LTE standard supports both frequency division duplex (FDD) and time division duplex (TDD). After Rel 8, Rel 9 was a relatively small update on top of Rel 8, and Rel 10, also known as LTE-Advanced, provided a major step in terms of data rates and capacity with carrier

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 137–145, 2019.
https://doi.org/10.1007/978-3-030-26630-1_11

aggregation, support for heterogeneous network (HetNet), higher-order Multi-Input-Multi-Output (MIMO) up to eight antennas in downlink and four antennas in uplink.

LTE air interface uses orthogonal frequency division multiple access (OFDMA) for downlink transmission to achieve high peak data rates in high spectrum bandwidth and single carrier frequency division multiple access (SC-FDMA) for uplink transmission, a technology that provides advantages in power efficiency. LTE supports advanced adaptive MIMO, balance average/peak throughput, and coverage/cell-edge bit rate. Compared to 3G, significant reduction in delay over air interface can be supported in LTE, and it is suitable for real-time applications.

LTE Rel 8 was frozen in December 2008 and it has been the basis for first wave of LTE equipment. Since then several operators have been planning and deploying LTE network to satisfy data demands of their mobile users that use various innovative mobile data services. As a result, global number of LTE commercial networks and subscriptions has reached 681 and 3.2 billion, respectively by the end of quarter 1 of 2018 [2]. The number of subscriptions accounts for 38.5% of total mobile subscriptions. It has also been forecasted that LTE will account for more than 50% of all global mobile subscriptions by 2020 and more than 60% by 2022.

Ethiopian incumbent operator ethio telecom has also deployed and provided LTE based data service since 2015 but the service is limited in the Ethiopian capital, Addis Ababa. Thus, LTE has only a small share not more than a few hundred thousand out of ethio telecom mobile subscribers that has reached 39.54 million mobile subscribers as of February 2019 [3]. Due to increasing penetration of social media, video streaming and other innovative data services in the Ethiopian mobile market, mobile data demand has been significantly increasing [4]. As a result, beside optimizing existing limited LTE network in Addis Ababa, it is important to expand it to meet demands of data customers in major cities and other parts of the country.

In mobile radio network planning and optimization tasks, antenna parameters including tilt, azimuth and height are very important parameters to meet target network capacity, coverage and quality of service [5–7]. Various researchers analyze impacts of antenna tilt and its optimization to achieve quality LTE network, for instance authors of [8–11]. Authors in [8] have performed comparative analysis of mechanical and electrical down tilt on performance of LTE using a snapshot LTE simulator with 3D antenna modeling and they show that electrical down tilt has significant LTE performance impact, mainly in case of interference limited system. In [9] and [10], they have expanded the analysis including other antenna parameters and from perspective of self-optimization network. In [11], antenna tilt impact on performance of LTE network has also been evaluated considering realistic antenna radiation patterns.

In this paper, we present performance evaluation of existing Addis Ababa LTE network and then perform antenna tilt optimization work by applying educated trial and error optimization approach. For the work, we select a hotspot LTE network scenario and perform network simulation using WinProp software [14, 15]. Signal-to-interference-plus-noise-ratio (SINR) and throughput metrics are used for performance comparison. Results reveal that there is still room to improve existing LTE network performance by reconfiguring antenna tilt parameters. For instance, signal–to-interference-plus-noise-ratio (SINR) and throughput results show that we are able to achieve 12.5% and 4.4% peak SINR and throughput gains, respectively, with further optimization of existing electrical tilt configuration of the LTE network. We also believe that this work presents contribution to LTE optimization and expansion efforts of ethio telecom.

Remaining parts of the paper are organized as follows. Section 2 presents Addis Ababa LTE deployment scenario, network modelling and assumptions. Section 3 discusses antenna tilt optimization approach. In Sect. 4, results and performance comparison are presented. Finally, Sect. 5 provides concluding remarks and future works.

2 Deployment Scenario, Network Modeling and Assumptions

2.1 Addis Ababa LTE Deployment Scenario

Currently, there are 331 three-sector LTE evolved node Bs (eNBs) that are deployed in Addis Ababa to provide LTE data services. All sectors configuration is single carrier and apply 1800 MHz frequency band that is shared with existing 2G system.

Two types of antenna solutions are adopted in Addis Ababa LTE network: green-field and rooftop. In the green-field solution, antennas are mounted on towers and operate on dual mode within a frequency range of 1710–2200 MHz. One mode is for 1800 MHz band where LTE and 2G operate and the other one is for 2100 MHz band where 3G operates. In the rooftop solution where the antennas are installed on top of buildings and can operate in frequency ranges of 790–960 MHz and 1710–2180 MHz. The former is used for 2G while the latter is used for both LTE and 3G.

For this study, we select hotspot Bole area (Addis Ababa, Ethiopia) covered by 34 eNBs (102 cells). The selected area is shown in Fig. 1.

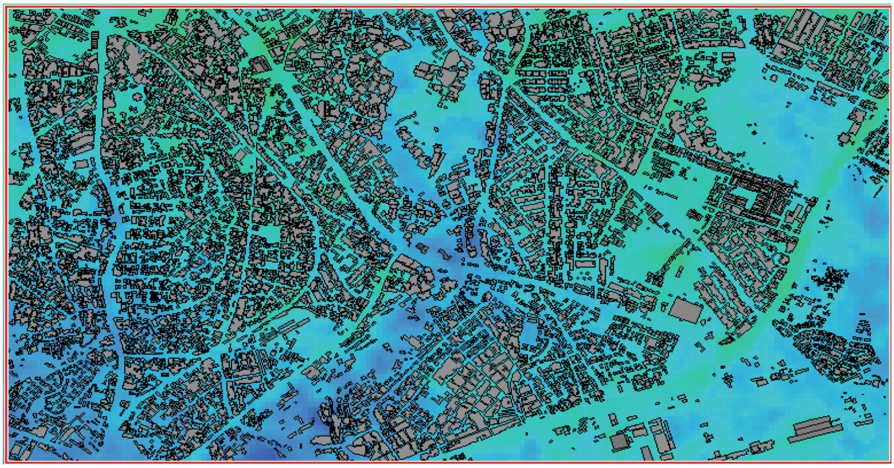


Fig. 1. Selected Bole area

Other configuration of the LTE radio network including antenna tilt are used for our work but all not directly presented here for confidentiality purpose. Only important parameters for completeness of this work are discussed.

2.2 Network Modelling and Assumption

One of the most useful performance metric of the LTE network is the SINR and is defined as

$$SINR = \frac{Pr}{\Sigma P' + N}$$

Where Pr is signal power, P′ is interference power and N is noise power. Another performance metric is throughput. The SINR metric is mapped to the throughput. The network performance evaluation and comparison is depicted using these metrics.

The simulation tool used for this study is the WinProp simulation tool. This tool is commercial and has been used in different research activities by senior researchers and academicians [12, 13]. The mapping in SINR to throughput is assumed in the WinProp simulation tool. The details about the tool is found in [14, 15]. Summary of basic network parameters and assumptions is depicted in Table 1. It has various modules such as the AMan (Antenna Manager) tool used for developing the antenna pattern of a practical antenna. In this study the antenna pattern is developed using AMan software module for each tilt of the antenna in both the horizontal and vertical planes and then convert them to a 3D pattern by extrapolating the pattern in both planes.

The propagation model adopted for the network simulation is the dominant path loss model. This model is used for modeling the propagation because it can be applied for urban scenario while majorly it concentrates only on the dominant paths and does not calculate the paths with small energy contributions. Hence, it doesn't consider all the details of the database, requiring less time to pre-process it. The accuracy of the model in very complex environments (with high shadowing of the direct ray) is higher than other models like ray tracing.

Table 1. Basic simulation parameters/assumptions

Parameters	Values
Number of eNBs	34 (102 cells)
Air interface	LTE FDD
Carrier frequency/bandwidth	1800/20 MHz
Simulation	System level simulator (WinProp)
Propagation model	Dominant path loss
Transmitter power	43 dBm
Antenna type	Aggison ADU451819 & ATR451704v01
UE distribution	Uniform

3 Tilt Optimization Approach

The selected Bole area existing network parameters configuration are used to simulate the network. Keeping other parameters setting, electrical antenna tilt parameter of the cells with worst SINR and throughput performance are altered and then simulated iteratively. The simulation iteration is done until optimal performance is achieved. Here, optimal value is to mean the best value obtained by this work. There could be other better possible values in other iterations. Figure 2 shows the steps of the optimization approach. The adjustment as in Table 2 shows the tilt setting to obtain best network performance improvement.

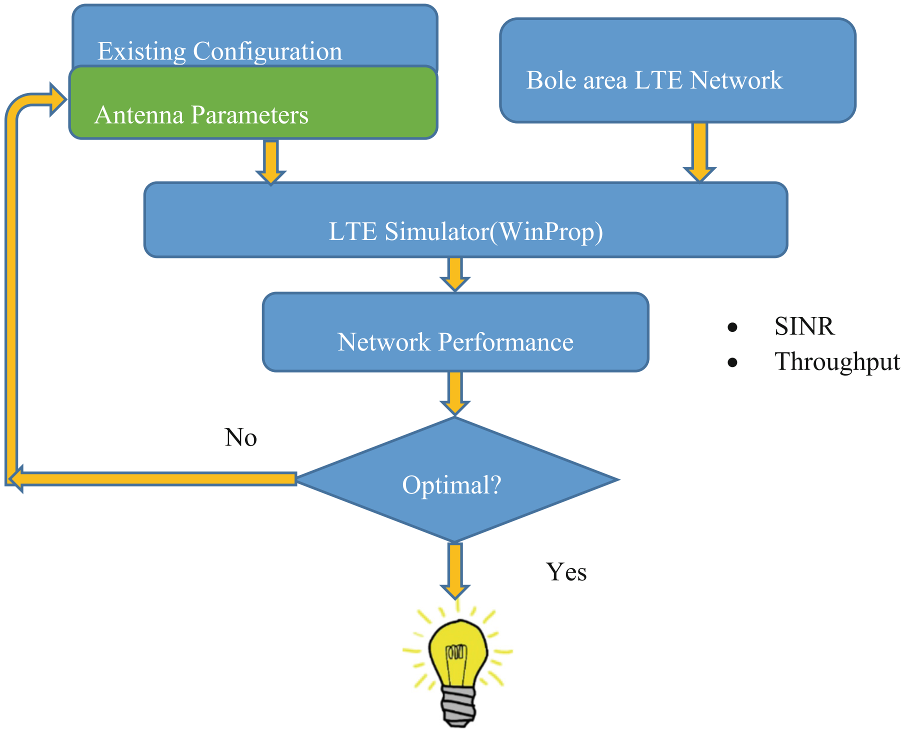


Fig. 2. Optimization approach

Table 2. Electrical down tilt adjustment

Site	34			4			6			8			5		
Antenna	Ant 1	Ant 2	Ant 3	Ant 1	Ant 2	Ant 3	Ant 1	Ant 2	Ant 3	Ant 1	Ant 2	Ant 3	Ant 1	Ant 2	Ant 3
Existing EDT	2	2	0	5	5	7	3	7	0	1	2	4	2	4	3
New EDT	2	4	1	4	5	5	4	5	0	3	2	1	3	5	3

4 Results and Performance Comparison

The SINR performance plot shown in Fig. 3 compares the performance for both the existing network configuration and after electrical tilt has been adjusted. The maximum achievable SINR for the existing network configuration is 16 dB and it is 18 dB after the electrical tilt adjustment as stated in Table 2. The highest SINR performance gain is observed at 70%.

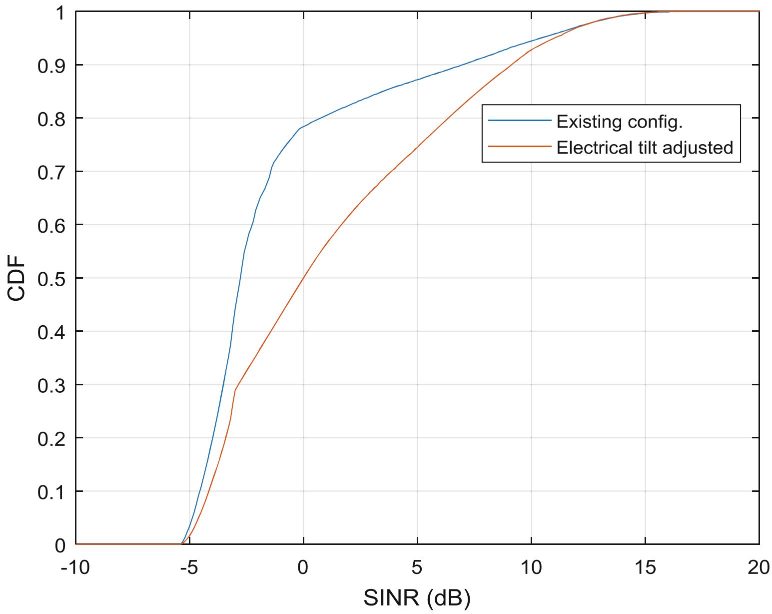


Fig. 3. SINR performance comparison

The SINR percentile shown in Fig. 4 shows a loss where as there is a good gain at 90 percentiles. The loss for the 10 %tile and 50 %tile is due to the fact that the SINR had negative values.

Figure 5 shows the throughput gain percentile. At the 10 %tile (1), there is no significant improvement. It shows better improvement at the 50th %tile (2) whereas there is best throughput gain at 90 percentiles (3) compared to 10 and 50 percentiles.

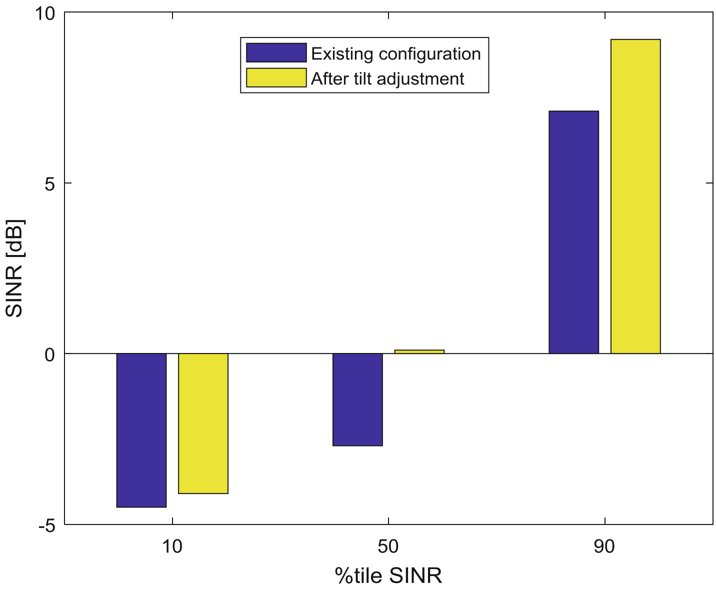


Fig. 4. SINR percentile; 10%, 50% and 90%

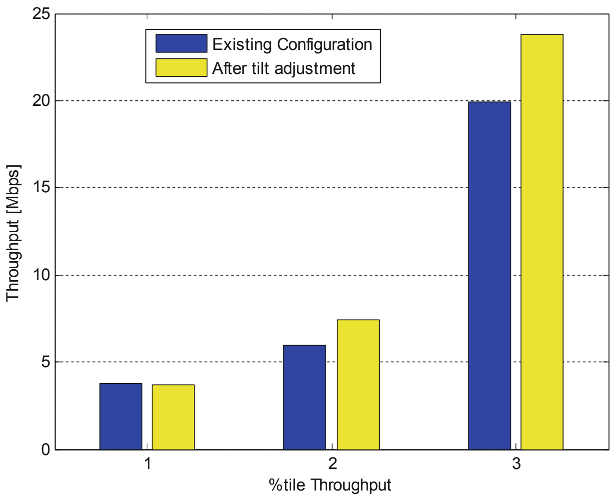


Fig. 5. Throughput percentile; 10%, 50% and 90%

5 Conclusions

In this paper, we have reviewed Addis Ababa LTE network and deployed antenna solutions. Moreover, antenna electrical down tilt parameter has been studied to investigate their impact on LTE network performance, particularly for an Addis

Ababa LTE deployment scenario. Furthermore, antenna tilt optimization is performed using educated trial and error approach.

Results show that antenna electrical down tilt can be further optimized to improve network performance. Such kind of antenna optimization are of paramount advantage for the operator for it does not require additional investments while improving network performance and in turn satisfying customers.

Future works for this study include study of the antenna down tilt using dynamic simulators since in reality the traffic and environment are changing. The research can be extended to the study of the future of antenna technology such as active antenna. Full or partial automation of antenna optimization process applying various self-optimization techniques is also another potential research area.

Acknowledgment. This work is performed under telecom MSc program jointly implemented by ethio telecom and Addis Ababa Institute of Technology. Network data is collected from staffs of engineering department of ethio telecom.

References

1. 3gpp. http://www.3gpp.org/technologies/keywords-acronyms/98-lte. Accessed 18 Sept 2018
2. GSA Report: Status of the LTE Ecosystem, Global Mobile Suppliers Association, August 2018
3. Ethio telecom. http://www.ethiotelecom.et/. Accessed 05 Feb 2019
4. Haile, B.B., Bulti, D.A., Zerihun, B.M.: On the relevance of capacity enhancing 5G technologies for Ethiopia. In: 10th Ethiopian ICT Annual Conference (EICTC-2017), June 2017
5. Hampel, G., et al.: Tradeoff between coverage and capacity in dynamic optimization of 3G cellular networks. In: 58th IEEE Vehicular Technology Conference (VTC 2003-Fall) (2003)
6. Lozano, M., et al.: UMTS optimum cell load balancing for inhomogeneous traffic patterns. In: 60th IEEE Vehicular Technology Conference (VTC2004-Fall), Los Angeles, CA, USA (2004)
7. Pettersen, M., et al.: Automatic antenna tilt control for capacity enhancement in UMTS FDD. In: 60th IEEE Vehicular Technology Conference (VTC2004-Fall), Los Angeles, CA, USA (2004)
8. Yilmaz, O.N.C., Hämäläinen, S., Hämäläinen, J.: Comparison of remote electrical and mechanical antenna downtilt performance for 3GPP LTE. In: 70th IEEE Vehicular Technology Conference Fall, Anchorage, AK, USA (2009)
9. Yilmaz, O.N.C., Hämäläinen, S., Hämäläinen, J.: Analysis of antenna parameter optimization space for 3GPP LTE. In: 70th IEEE Vehicular Technology Conference Fall, Anchorage, AK, USA (2009)
10. Yilmaz, O.N.C., Hämäläinen, S., Hämäläinen, J.: Self-optimization of remote electrical tilt. In: 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey (2010)
11. Ruise,A.C.: LTE Mobile network performance with antenna tilt considering real radiation patterns. M.Sc. thesis, KTH Royal Institute of Technology, May 2015
12. Haile, B.B., Mutafungwa, E., Hämäläinen, J.: LTE-advanced enhancements for self-backhauled LTE-U small cells: an addis Ababa case study. In: IEEE AFRICON 2015, 14–17 September 2015, Addis Ababa, Ethiopia (2015)

13. Haile, B.B., Mutafungwa, E., Hämäläinen, J.: Coordinated multi-point transmission for relaxation of self-backhauling bottlenecks in heterogeneous networks. EURASIP J. Wirel. Commun. Netw. **2015**, 78 (2015)
14. Altair Engineering: WinProp-Propagation Modeling. https://altairhyperworks.com/product/FEKO/WinProp-Propagation-Modeling
15. https://www.slideshare.net/ProjectENhANCE/winprop-propagation-modeling-and-network-planning-tool. Accessed 06 Aug 2018

Low Complexity Radio Mobile Speed Estimation Using Count of Received Envelop Peaks and Troughs

Atli Lemma(✉)

Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia
atli.lemma@outlook.com

Abstract. Low complexity mobile speed estimation is required for power control, mobility and resource management, effective handoff, and outage probability computation of cellular mobile communication systems. It is also useful for mobility determination and energy-efficient packet routing in wireless sensor and mobile ad-hoc networks. In this paper, we present a low complexity mobile speed estimation model using count of peaks and troughs of the received signal envelop. Our simulation result shows that the model has a maximum error of 0.25 m/s. The model has two advantages. First, it does not require measurement of the received signal power; it only counts envelop peaks and troughs. Second, the model is independent of dc offset inherent in the radio receivers. However, the model has one limitation- it does not give the crossing component of a mobile's velocity.

Keywords: Mobile speed estimation · Envelop peaks and troughs · Resource management

1 Introduction

Accurate and low complexity mobile speed estimation is required for power control, radio resource and mobility management such as dynamic channel assignment, effective handoff, and outage probability computations in cellular mobile communication systems [1–11]. Specifically, the long term evolution (LTE) network is designed to provide high data rates under various mobility conditions [1]. Thus, low complexity and high accuracy mobile speed estimation is required in these systems. In addition, speed estimation is needed for mobility determination and energy-efficient packet routing in wireless sensor and mobile ad-hoc networks respectively [1, 9]. Moreover, adaptive transmission in mobile systems depends on knowledge of mobile speed [8]. In all these applications, it is sufficient to determine the magnitude of the mobile velocity.

A number of mobile speed estimation techniques can be found in literature. The first is based on the level crossing rate of the received signal envelop. The scheme is efficient and robust to signal variation [5]. However, since it requires envelop power which in turn depends on mobile location and shadowing, it requires dynamic adjustment of the level and hence it is difficult to use this model in practice [5, 8]. The second technique is based on the zero crossing rates of the inphase or quadrature

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 146–153, 2019.
https://doi.org/10.1007/978-3-030-26630-1_12

components of the received radio signal. Like the level crossing technique, this model is efficient and robust [5, 8]. But, it requires mobiles or base stations to use coherent detectors [5, 8]. The third technique employs the auto-covariance of the received signal envelop [5, 8]. This technique requires measurement of envelop power, which is difficult in cellular systems [5, 8]. The fourth method is based on diversity combiner and has high complexity [5]. The fifth technique uses the rate of maxima of the received signal envelop [1]. We have found that this method does not require measurement of mean envelop power [5]. Nevertheless, its performance degrades with prevalence of shadowing [5]. In addition, when compared with our technique, it has higher complexity. Finally, a mobile speed estimation technique is given in [11] that is based on average slope duration of the received signal envelop. The technique does not require measurement of mean envelop power [11]. Also, since it does not require variable observation window, it is convenient to implement in practice [11]. However, the model has high complexity. In our work, we have developed a mobile speed estimation model that has all the advantages of the previous models plus it has low complexity. This paper is organized as follows. In Sect. 2, the received envelop model is given. Here, we presented a modified form of Jake's the sum-of-sinusoids model. In Sect. 3, we derived the speed estimation model using the count of received envelop peaks and troughs. In Sect. 4, we presented our simulation results and discussions. Finally, in Sect. 5, we give important conclusions.

2 Received Envelop Model

2.1 Modified Sum of Sinusoids

In a multipath radio environment, there is no scenario specific information and obstacles are not uniformly spaced as often assumed [13, 14]. Moreover, multipath has different fading distributions for different scenarios [14]. Let us consider a frequency selective channel characterized by multipath fading. A mobile node receives multiple attenuated versions of the transmitted signal, *s(t)*. For the sake of simplicity, let us neglect noise; for a fairly big signal-to-noise ratio (SNR), the model is independent of the noise. Thus, the received signal can be written as [12–14]:

$$r(t) = A_t e^{-j\omega_t t} \sum_{n=1}^{N} c_n e^{j(\omega_{rm} t cos[\theta_n] + \varphi_n)} \tag{1}$$

Here, A_t is the transmit signal strength, ω_t is the transmit frequency, c_n is the multipath attenuation coefficient, ω_{rm} is the received maximum frequency with respect to the transmit frequency, θ_n is the angle of arrival of the n^{th} path and is the angle between the velocity vector and the radial direction between the mobile node and the base station, and φ_n is the phase of n^{th} path received signal.

The inphase and quadrature components of the received signal envelop in (1) are given by

$$\left.\begin{aligned} r_i(t) &= A_t cos(\omega_t t)\sum_{n=1}^{N} c_n cos[\omega_{rm} t cos(\theta_n) + \varphi_n] \\ r_q(t) &= A_t sin(\omega_t t)\sum_{n=1}^{N} c_n sin[\omega_{rm} t cos(\theta_n) + \varphi_n] \end{aligned}\right\} \tag{2}$$

The maximum radian frequency of the received signal can be found by applying Doppler equation given by [12, 14]:

$$\omega_{rm} = \omega_t\left(1 \pm \frac{v}{c}\right) \tag{3}$$

2.2 Phase of Arrival

For multipath and Doppler fading mobile communication channel, the propagation delay of each path can be chosen from an exponentially distributed delay variable [16–17]. The probability density function (PDF) of such a variable, τ, is given by [19–20]:

$$f_\tau(\tau) = \frac{1}{T_{av}} e^{-\frac{\tau}{T_{av}}} \tag{4}$$

Here, T_{av} is a constant that can be determined as follows. First, the minimum average propagation delay can be estimated, for example, with a 3 m minimum separation of the mobile from the base station. So, for free space τ_{min} = 10 ns. Then, the maximum delay, on the other hand, can be estimated from a cell diameter D and microwave propagation speed c = 3e+8 m/s using:

$$\tau_{max} = \frac{D}{2c} \tag{5}$$

The probability distribution function over all possibilities of the propagation delay is unity. That is:

$$\int_{\tau_{min}}^{\tau_{max}} \frac{1}{T_{av}} e^{-\frac{\tau}{T_{av}}} d\tau = 1 \tag{6}$$

Carrying out the integration, we get the following:

$$e^{-\frac{\tau_{min}}{T_{av}}} - e^{-\frac{\tau_{max}}{T_{av}}} = 1 \tag{7}$$

The equation above is transcendental and we resorted to numerical methods to obtain T_{av}. For a known value of the average propagation delay, the cumulative distribution function (CDF) of τ is [14]:

$$F_\tau(\tau) = \int_{-\infty}^{\tau} f_\tau(\delta)d\delta \tag{8}$$

After computing the integral and simplifying it further in the interval $\tau_{min} < \tau < \tau_{max}$ we get:

$$F_\tau(\tau) = 1 - e^{-\frac{\tau}{T_{av}}} \tag{9}$$

Thus, the delay can be obtained from the CDF in (9) and the phase of the received radio signal can be found from:

$$\varphi = 2\pi n\tau \tag{10}$$

3 Speed Estimation Model

Studies showed that the count of envelop peaks and troughs contains rich information about the speed of a mobile. Also, empirical measurements show that the adjacent envelop peaks are half wavelength apart. That is, as the mobile moves faster, the number of envelop peaks and troughs per unit time increases. We have simulated the envelop at two different speeds as shown in Fig. 1. Thus, the count of envelop peaks and troughs are a function of the mobile's radial speed. Thus, it should be possible to estimate the mobile speed from count of the envelop peaks and troughs. If we take the time derivative of the signal envelop, then the zeros of this rate will indicate the locations of envelop peaks and troughs. So, the counts of the zeros of this derivative is equal to the count of the envelop peaks and troughs. The signal envelop is Rayleigh distributed and the count of the envelop peak and trough is given by [12, 13]:

$$R_r(t) = 2\int_{-\infty}^{+\infty} |r'(t)| f_{rr'}(r = 0, r')dr' \tag{11}$$

Where the received signal envelop is a random process with a probability density function $f_r(r, r')$. Now, making the transformation $W(t) = r^2(t)$, we get:

$$W^{'}(t) = 2r(t)r'(t) \tag{12}$$

Therefore, the count of the envelop peak and trough can be written as [9, 11]:

$$R_{W'}(t) = 2\int_{-\infty}^{+\infty} \left|W^{''}(t)\right| f_{rr'}\left(W^{'} = 0, W^{''}(t)\right)dW^{''} \tag{13}$$

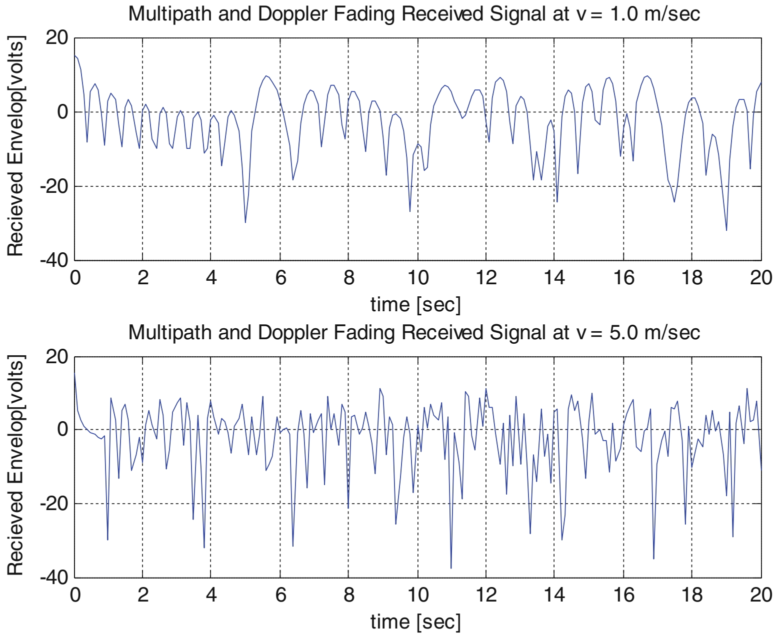


Fig. 1. Rayleigh fading envelop at two different mobile speeds

Thus, the count of zeros of the time derivative of the envelop $R_{r'}(t)$and the count of zeros of the time derivative of the envelop squared $R_{W'}(t)$are equal [11]. That is:

$$R_{r'}(t) = R_{W'}(t) \tag{14}$$

The average number of envelop peaks and troughs per second can be found from (13) and (14) to be [11]:

$$R_{W''}(t) = \frac{1}{2\pi}\sqrt{\frac{b_2}{b_0}}I_W \tag{15}$$

I_W is the indefinite integral of (13) and is numerically computed to be 4.2757 [8]. On the other hand, the field moments are given by [12, 13]:

$$b_2 = 2b_0[f_{max}(t)]^2 \tag{16}$$

The maximum frequency of the received signal is given by the Doppler equation [11–15]:

$$f_{max}(t) = f_t\frac{v(t)}{c} \tag{17}$$

Now, from (15), (16), and (17), we obtain:

$$R_{W'}(t) = \frac{v(t)}{c\sqrt{2}} f_t I_W \tag{18}$$

Finally, from (14) and (18), we get the estimate of the radial speed of the mobile to be:

$$\hat{v}(t) = 1.323 \frac{c}{f_t} R_{r'}(t) \tag{19}$$

The Algorithm:
Step 1: Generate the reference mobile speed which is to be estimated. We have generated the reference mobile speed using $v = \sqrt{t}$.
Step 2: Generate large samples of the received signal envelop for a speed estimation interval using (2) and the reference speed obtained in step 1.
Step 3: Compute the count of the received envelop peaks and troughs from the sample set.
Step 4: Find the estimate of the mobile speed using (19), increment the time.

4 Results and Discussions

In our work, we simulated the signal envelop by implementing the modified form of Jakes Sum-of-Sinusoids presented in Sect. 2. The simulation parameters are given in Table 1. The distance dependent path loss does not affect the performance of this model and hence it is neglected. We have used envelop averages and neglected the effect of noise on the model.

Table 1. Simulation parameter.

Parameter	Value	Description
c	3.00e + 8 m/s	Speed of propagation
f_t	5.00e + 9 Hz	Transmit frequency
D	5.00e + 3 m	Diameter of cell site
τ_{min}	1.00e − 8 s	Minimum envelop delay
τ_{max}	4.30e − 7 s	Maximum envelop delay

We have given the reference and estimated mobile speeds as shown in Fig. 2. The estimated speed is close to the reference speed with a maximum error of 0.25 m/s. However, the moving average of the estimated speed is very close to the reference speed. This signifies that better speed estimates can be obtained by slightly increasing the complexity with the use of a moving average filter. For example, the worst error for the peak and tough count model is 0.25 m/s (at v = 2.5 m/s). Furthermore, the error range is from 0 m/s to 0.25 m/s for the peak and trough count model.

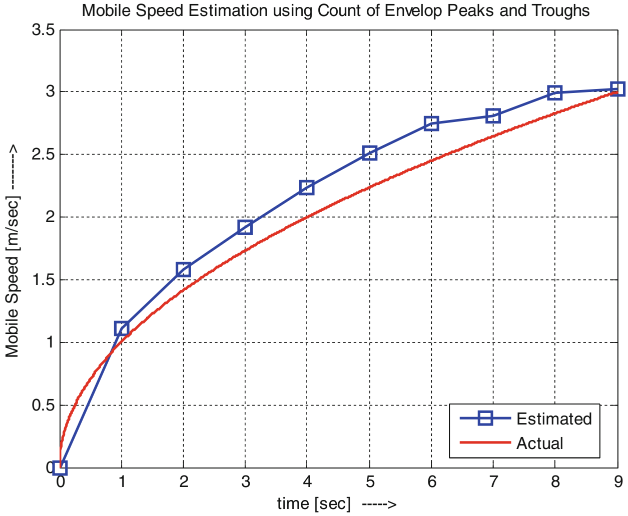


Fig. 2. Speed estimates using count of envelop peaks and troughs

5 Conclusions

Accurate and low complexity mobile speed estimation is required in wireless communications systems. We have developed a simple and practical mobile speed estimation model that has two advantages. First, it does not require measurement of the received signal power. Second, the model is independent of dc offset inherent in the radio receivers. However, our model has one limitation- it does not give the crossing component of a mobile's velocity. The model is developed to estimate the radial speed of a mobile with respect to another node which may or may not be mobile. The model has high accuracy and low complexity that makes it attractive for time sensitive speed estimation, such as mobile handoff.

References

1. Osfiko, M.O., Okubanjo, A.A., Abolade, O.R., Oyetola, O.K., Oyedeji, A.O., Sanusi, O.I.: Evaluating the effect of mobility speed on the performance of three handover algorithms in long term evolution networks. J. Appl. Sci. Environ. Mang. **22**(4), 503–506 (2018)
2. Merwadey, A., Guvenc, I.: Handover Count Based Velocity Estimation and Mobility State Detection in Dense HetNets (2016)
3. Zheng, Y.R., Xiao, C.: Mobile speed estimation for broadband wireless communications over Rician fading channels. IEEE Trans. Wirel. Commun. **8**(1), 1–5 (2009)

4. Abdi, A.: A New Velocity Estimator for Cellular Systems Based on Higher Order Crossings
5. Patachaianand, R., Sandrasegaran, K.: User speed estimation techniques for UMTS. Electron. Lett. **43**(19), 1036–1037 (2007)
6. Iera, A., Molinaro, A., Marano, S.: Handoff management with mobility estimation in hierarchical systems. IEEE Trans. Veh. Technol. **51**(5), 915–934 (2002)
7. Abdi, A., Kaveh, M.: Level crossing rate in terms of the characteristic functions: a new approach for calculating the fading rate in diversity systems. IEEE Trans. Commun. **50**(9), 1397–1400 (2002)
8. Abdi, A., Zhang, H., Tepedelenlioglu, C.: A unified approach to the performance analysis of speed estimation techniques in mobile communication. IEEE Trans. Commun. **56**(1), 126–135 (2008)
9. Li, Z., Sun, L., Ifeachor, E.C.: Rang-based relative velocity estimations for networked mobile devices. IEEE Trans. Veh. Technol. **58**(4), 2095–2099 (2009)
10. Juncker, C., Toft, P., Morch, N.: Speed Estimation for WCDMA on the Channel Envelop Derivative
11. Zhao, L., Mark, J.W.: Mobile speed estimation based on average slope duration. IEEE Trans. Commun. **52**(12), 2066–2069 (2004)
12. Stuber, G.L.: Principle of Mobile Communications, 4th edn. Springer, Atlanta (2017)
13. Jakes, W.C.: Microwave Mobile Communications, 2nd edn. CRC/IEEE Press (1995)
14. Jeruchim, M.C., Balaban, P., Shanmugan, K.: Simulations of Communication Systems: Modeling, Methodology, and Techniques, 2nd edn. (2000)
15. Xiao, C., Rosa, Y., Beaulieu, N.C.: Novel sum-of-sinusoids simulation model for Rayleigh and Rician fading channels. IEEE Trans. Wirel. Commun. **5**(12), 3667–3679 (2006)

Performance Analysis of Vertical Sectorization for LTE Realistic Deployment Scenario

Seifu Girma Zeleke(✉) and Beneyam Berehanu Haile

Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia
{seifu.girma, beneyamb.haile}@aait.edu.et

Abstract. Mobile network capacity demand has been continuous increasing due to increasing penetration of data intensive services. To meet the increasing capacity demand, vertical sectorization has been considered one important capacity enhancing technology through vertical cell splitting into inner and outer cells by applying active antenna system. Planning vertically sectorized network is a new concept and need to be investigated well in order to characterize its impact on cell edge performance and capacity. Although this method presented by different authors, still various issues needs to be further investigated. Analyzing this method in realistic network environment, vendor antenna pattern and effect of spacing between vertical pattern with considering cell edge performance are rarely reported. In this paper, impact of this technology with performance metrics of 5% tile, 50% tile and 90% of CDF of SINR and their modified shannon mapping in realistic environment are examined. Propagation computation is performed using deterministic path loss model and network simulation is made using matlab. According to simulation results the identified 3×2 configuration out performs others 3×2 studied scenarios. And this configurations can enhance system performance up to 51.8% compared to the existing 3×1 network.

Keywords: Vertical sectorization · Active antenna system · Antenna pattern

1 Introduction

The introduction of new application, new devices and services has driven the network demand growth at an exponential rate. Forecasts also indicate that global mobile data and internet traffic will show an increase from the current 11,183 PB/month to 48,270 PB/month by 2021 [1]. Similarly, significant mobile traffic growth has also been seen in emerging market like Ethiopia due to significant growth of mobile internet penetration and the increasing social media and other multimedia service usage, particularly in the major cities like Addis Ababa, Ethiopia [2]. This increase of traffic demand has motivated the development of different solutions for increasing network capacity. Network densification technique using vertically sectorized cells has been recently considered due to its capability enhancing system performance without introducing additional sites unlike small cell densification. This is done by splitting the cell vertically, creating inner and outer layouts by active antenna system. Both beams

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 154–163, 2019.
https://doi.org/10.1007/978-3-030-26630-1_13

use the same communication resources [3–6]. A graphical representation of vertical sectorization is shown in Fig. 1 where the outer and the inner beams forms the outer and the inner sectors respectively. The inner cell is closed to the BS and typically covers a small portion of the cell surface.

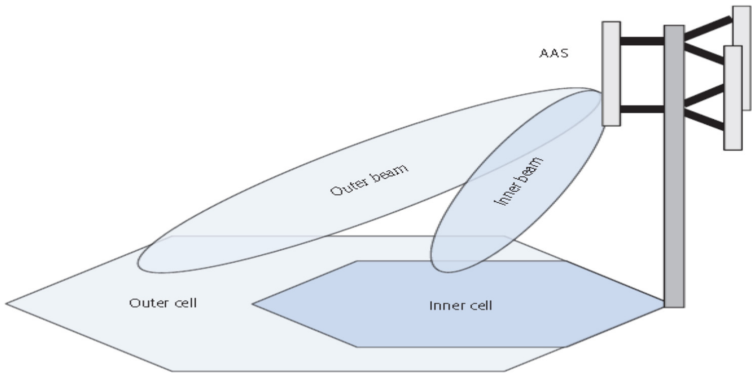


Fig. 1. Sector layout of vertical sectorization system [3].

Mobile operators like Ethio telecom are expected to optimize their mobile networks with these emerging technologies to accommodate the continuous data demand growth. However optimizing network with vertical sectorization is a new concept and needs further investigation to characterize its impact for different antenna configuration. Figure 2 shows a traffic data collected around bole, Addis Ababa and covers an area of about 1.62 km^2.

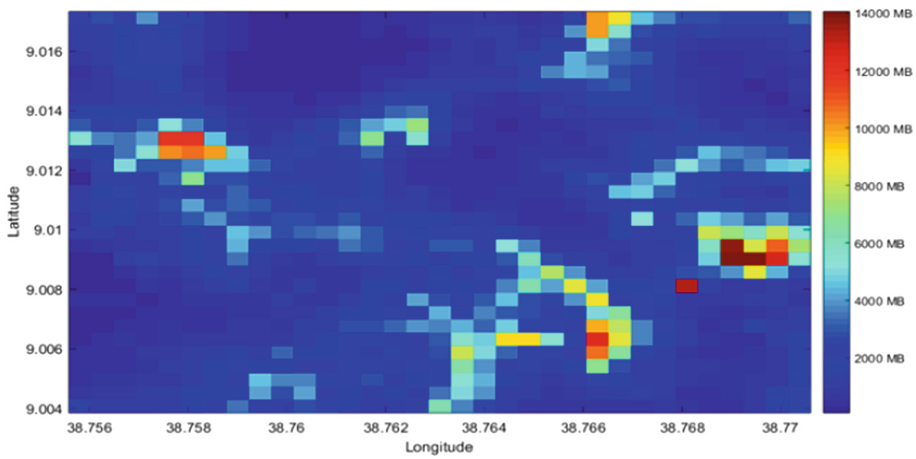


Fig. 2. Traffic data collected during June, 10, 2018 around bole, Addis Ababa

Researchers have shown the effectiveness of vertical sectorization on system performance [5–12]. Performance of vertical sectorization in LTE system for single user is tested in [4]. Simulation results indicated that the co-channel interference between the inner and outer cells is the main source of throughput impact. [5] and [6] investigate the performance impact of this technology for various vertical half-power beam widths (HBPW) and electrical tilt angles to observe the optimization space of the overlap between two vertical sectors and the gain in terms of cell capacity. They use 3GPP case simulation parameters and antenna pattern proposed by [7]. The authors [8–10] optimizing the performance of vertical sectorization with site configuration of antenna down tilt, vertical and horizontal beam width and transmit power in LTE-A downlink systems. The authors in [11] analyzed the performance impact of vertical sectorization in realistic environment using different antenna beam width but for super cell case. The performance of vertical sectorization with respect to different inner beam down tilt and power ratio is also examined in [12]. However, it considers synthetic environments. Performance impact of small cell deployment in unlicensed band for the case of Addis Ababa is analyzed in [13]. But, the author's do not consider vertical sectorization in their work. Antenna configuration, UE distribution cell association, inters site distance and building height can affect the performance of this technology [3–6] and [12]. Most of the above work carried out their simulation in unrealistic environment and antenna pattern. This results in neglecting the influence of side lobes. Furthermore, investigation of this technology for varying vertical beams bore sights angle difference considering both cell edge performance and capacity is rarely reported.

The purpose of this study is to analyze the performance impact of vertical sectorization for different site configuration and to compare with the existing 3×1 network in realistic network environment and vendor antenna pattern by varying outer beams electrical tilt angle.

To do this, first we identified the studied antenna configuration. Then we model user distribution based on traffic data. The received power of each user is obtained using the deterministic dominant path loss model in win prop propagation tool of studied area based on terrain and building maps. Finally network simulation is implemented using Matlab.

The remaining part is organized as follows. Section 2 provides deployment scenario, system model and assumptions used in the paper for simulation purpose. Simulation results and discussions are described in Sect. 3. Finally, Sect. 4 concludes the work.

2 System Model, Deployment Scenario, and Assumptions

2.1 System Model

In this work we consider a down link LTE mobile network 6 macro cells in the studied area. Cell association is made based on user's received power. The received power is computed using deterministic dominant path loss model based on building and terrain maps of the deployment area. Performances of the different site configurations are studied based on 5% tile, 50% tile and 90% tile of CDF of the SINR. And the performance of the identified site configuration is compared with the existing network by corresponding modified Shannon mapping of CDF of SINR and throughput. Throughput is expressed using modified Shannon formula [14] as:

$$\mathrm{TP} \; = BW_{eff} N_{PRB} B_{PRB} \log_2(1 + \frac{SINR}{SINR_{eff}}) \quad (1)$$

Where N_{PRB} is the number of PRBs, B_{PRB} is the bandwidth per PRB, BW_{eff} is the adjusted bandwidth to fit with LTE the system bandwidth efficiency and $SINR_{eff}$ is the adjusted $SINR$ implementation efficiency.

Antenna Models: In this paper we used huawei antenna patterns of the following models: ADU451819 for the existing 3 × 1 reference network, ADU451900 for 3 × 2 the outer cell with different remote electrical tilt and A19451800 for 3 × 2 inner cell. Sample antenna pattern used is shown in Fig. 3 below.

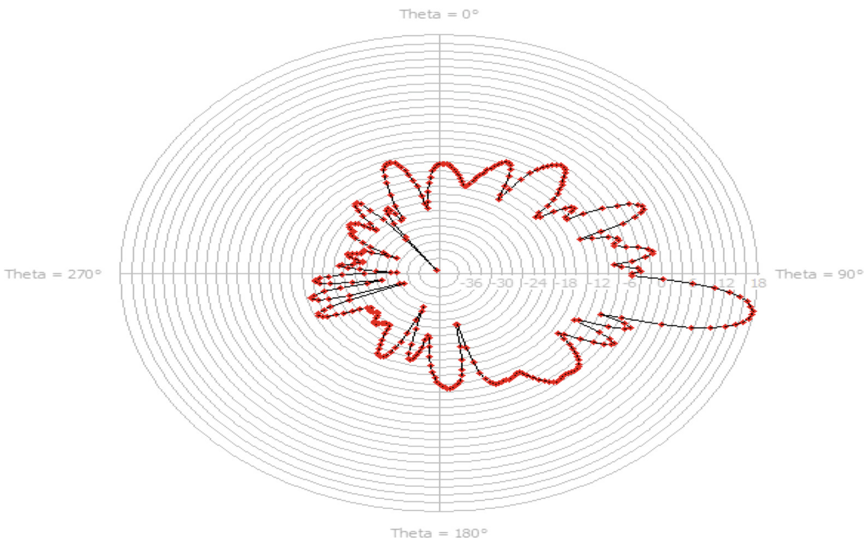


Fig. 3. Sample vertical antenna patteren used in simulation (RET 10 and V.HPBW 7.5 Deg.)

2.2 Deployment Scenario

To study the performance impact of vertical sectorization, a realistic network environment in Addis Ababa Ethiopia that exemplifies an urban scenario with dense hotspots is selected. It is specifically found at around Dembele, a place where high data traffic is recorded. It covers an area of 1.1 km × 1.0 km, building of various heights up to 70 m and terrain with topography range from 2000–3000 m. The area consists of tri-sector of 6 macro sites with real location of eNodeBs for existing Addis Ababa LTE network. For vertical sectorization study purpose those 3 × 1 sectors are replaced by 3 × 2 vertically sectorized cells. The location of macro sites are shown in Fig. 4 below.

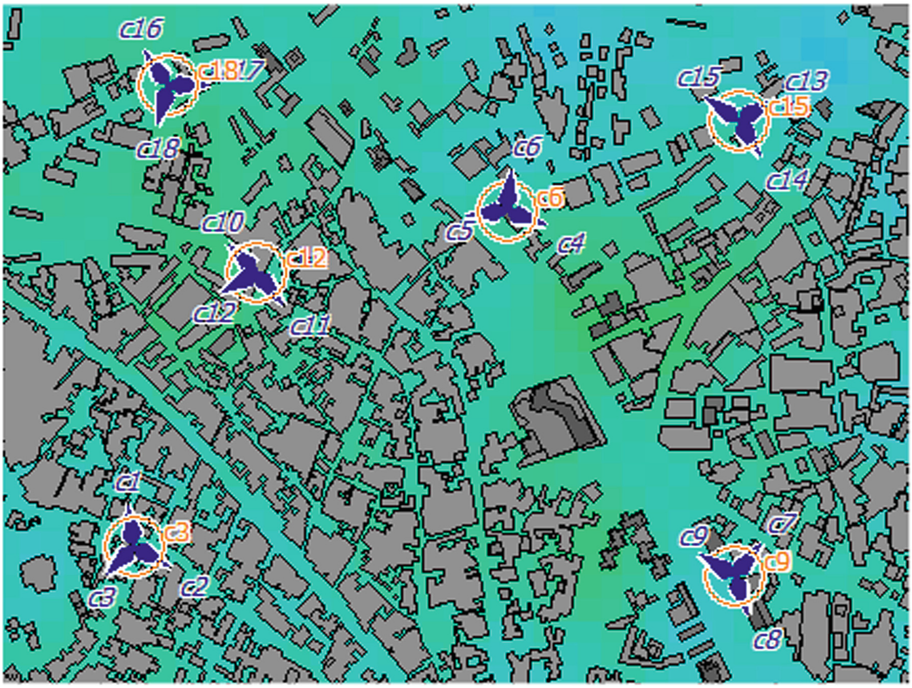


Fig. 4. Performance studied area

2.3 Simulation Parameters and Assumptions

In this work we have studied antenna configurations and simulation parameters shown in Tables 1 and 2 below respectively. The existing 3 × 1 configuration comprises of 3 sectors in each macro sites and antenna pattern describes in Sect. 2.1. For the case of 3 × 2 vertically sectorized cells we considered 10° RET for all inner beams and different RET for outer beams. The studied configurations are selected to observe the optimization space between the inner and the outer cells.

Table 1. Studied antenna configuration

Scenarios/ configuration (3 × 2)	V.HPBW out. beam (Deg.)	V.HPBW inn. beam (Deg.)	RET inn. beam (Deg.)	RET out. beam (Deg.)	Mech. tilt (Deg.)	Antenna height (m)	RET angle difference (Deg.)
1	4.5	7.5	10	0	10	40	10
2	4.5	4.5	10	2	10	40	8
3	4.5	4.5	10	4	10	40	6
4	4.5	4.5	10	6	10	40	4
5	4.5	4.5	Same as existing 3 × 1 configuration				

Most of system simulation parameters are taken from 3GPP case 3 except network layout and traffic distribution. Real building and terrain maps is used as network layout and modeled user distribution based on recorded traffic is used in place of traffic distribution.

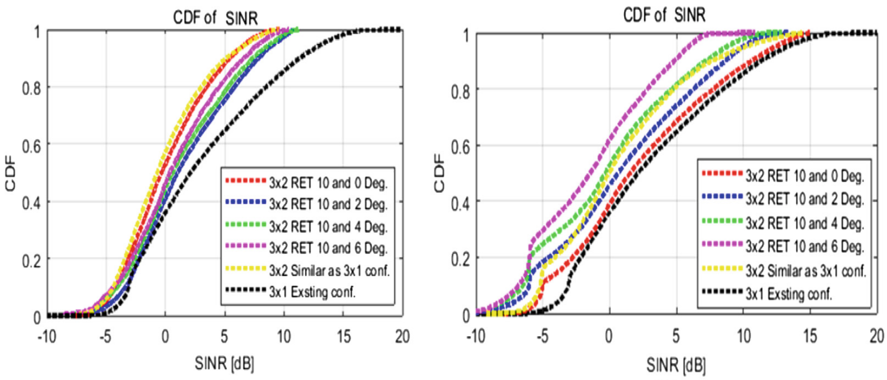
Table 2. Simulation parameters and assumptions

Parameters	Values/Assumptions
Downlink transmit power	46 dBm
System frequency	1800 MHz
Antenna technique	2×2 MIMO
Frequency reuse factor	1
Base station height	40 m
Horizontal HPBW	63°
Number of UEs/Vertical sector	10
UE height	1.5 m
Simulation area	1.1×1.0 km^2
Receive antenna gain	0 dBi
BW_{eff}	0.60
$SINR_{eff}$	1.8
Thermal floor	−174 dBm/Hz
Traffic distribution	Hotspot
Cell association	RSRP

3 Results and Discussions

In this section, the impact of vertical sectorization in LTE downlink cellular networks is evaluated. CDF of SINR and throughput performance are analyzed for different antenna configuration in realistic network environment. Simulations are carried out to select antenna configuration that can give the best performance among the selected site configuration. Then throughput performance of the identified configuration is compared with the exsting LTE network configuartion. The performance matrices and antenna configurations used in the simulation are described in Sect. 2.3.

Performances of the outer and inner beams with respect to CDF of SINR and the associated UE load distribution are shown in Fig. 5 and Table 3. From Fig. 5(a), CDF of SINR increasing with increasing spacing between RET tilt except for the exsting site configuration. Among the studied configuration, RET of 10 and 0° (configuration 1) out performs the other configurations. For example, SINR at 50% tile, configuration 1 has 4 dB performance difference with that of Configuration 4. And also, this configuration has a 2 dB and 2.2 dB performance difference with that of angle diffrence of 4 (configuration 2) configuration 3 at 50% tile of CDF. For inner cells case as shown in Fig. 5(b), angle difference of 8° (configuration 2) gives the best performance. However in both cases the exsting 3×1 configuration outperforms all the studied configurations interms of SINR with the corresponding load distribution as shown in Table 3.



(a) (b)

Fig. 5. Comparisions of CDF of SINR, (a) Outer cells users (b) Inner cells users

Table 3. User load distribution for inner and outer cell for different configuration

Scenarios/ configuration	RET tilt (Deg.)		Mechanical tilt. (Deg.)	UE-load distribution outer cell (%)
	Inner cell	Outer cell		
1	10	0	10	79.8
2	10	2	10	79.2
3	10	4	10	83.2
4	10	6	10	88.8
5	Same as the exsting 3×2 configuration		10	71%

The combined effect of both inner and outer beams users CDF of SINR are compared using the exsting 3×1 network as reference configuration as shown in Fig. 6. From Fig. 6, configuration 1 having angle difference 10° (scenario 1) gives the best SINR results compared to other 3×2 scenaros with respect to all the three performance metrices. In all cases, CDF of SINR of the exsting 3×1 configuration outperforms other studied configuration. This is mainly because when vertical sectorization is activated additional co-channel interferences are introduced beside vertical sectorization doubling the resource elements.

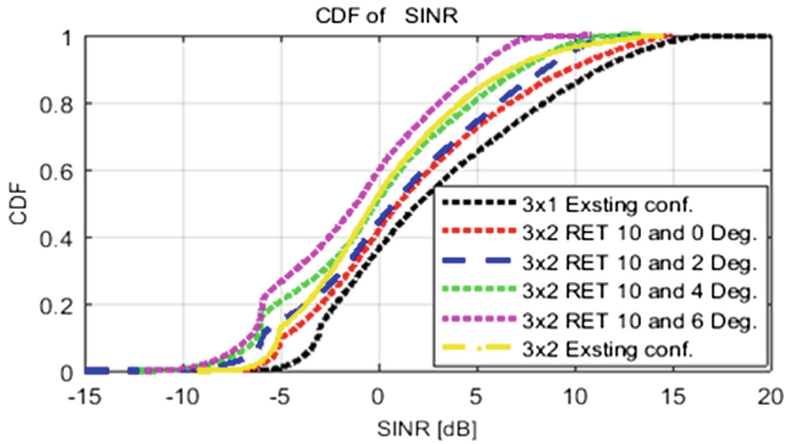


Fig. 6. Comparisions of CDF of SINR for diffrent cells configuration

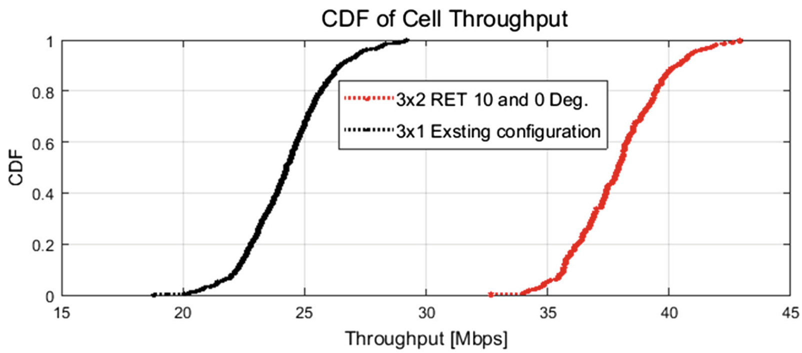


Fig. 7. CDF of cell throughput

Figure 7 show CDF throughput comparisons of the exsting 3×1 and the identified 3×2 vertically sectorized cells configuration. As it can be seen in Fig. 7 although vertical sectorization reduce the SINR performance of each user as shown in the Fig. 6 above, it improves CDF of cell throughput. This is because of vertical sectorization doubling the resource elements.

Figure 8 show performances of average cell and user throughput of the exsting 3×1 configuration and the identified 3×2 configuration. As we can see from the figure both average user and cell throughput of vertically sectorized cells out performs the exsting configuration. For example there is 1.36 Mbps incremental in average user throughput and 51.8% Mbps performance gain compared to the exsting network with and without vertical sectorization (Table 4).

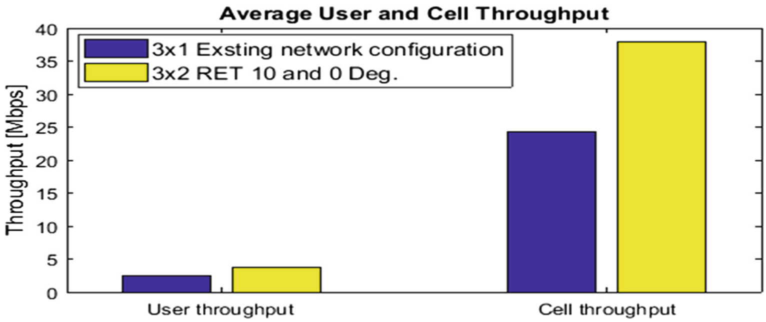


Fig. 8 Comparisions of average user and cell throughput

Table 4. Performance gain of the identified 3 × 2 vertically sectorized configuration

Down link average cell throughput with out vertical sectorization	Down link average cell throughput with vertical sectorization	Capacity gain by vertical sectorization
24.89 Mbps	37.78 Mbps	51.8%

4 Conclusion

In this paper, we have analyzed the impact of vertical sectorization in realistic environment of Addis Ababa, for diffrent scenarios to identify site configuration which can give best performance interms of 5% tile, 50% tile and 90% tile considering the exsting network as a refernce. Throughput performance of the identified configuration is compared with the exsting LTE network. Based on obtained results vertical sectorization can inhance system capacity upto 51.8% besides reduction of users SINR. The performance gain is obtained due to the introduction of additional resource elements. We also note that, when the angle separation between the inner and outer beams become narrow and narrow a high performance degradation is observed on both the capacity and cell edge performance. Increasing spacing between vertical RET keeping other parameters constant enhance both capacity and cell edge performance. Effect of vertical sectorization in high rise building, multi objective optimization and cell association in vertical sectorization can be an important future work.

References

1. Cisco white paper, Cisco Visual Networking Index: Forecast and Methodology, 2016–2021. Cisco Public (2017)
2. Adam, L.: Understanding what is happening in ICT in Ethiopia. A supply and demand side analysis of the ICT sector, Evidence for ICT Policy, Action Policy Paper (2012)
3. Weiliang, X., Qimei, C., Fengyi, Y., Qi, B., Yifei, Y.: Experimental investigation on a vertical sectorization system with active antenna. IEEE Commun. Mag. **54**, 89–97 (2016)

4. Fengyi, Y., Jianmin, Z., Weiliang, X., Xuetian, Z.: Field trial results for vertical sectorization in LTE network using active antenna system. In: IEEE International Conference on Communications (ICC), pp. 2508–2512. IEEE, Sydney (2014)
5. Osman, N., Yilmaz, S., Jyri, H.: System level analysis of vertical sectorization for 3GPP LTE. In: 6th International Symposium on Wireless Communication Systems, pp. 453–457. IEEE, Tuscany (2009)
6. Youqi, F., Jian, W., Zhuyan, Z., Liyun, D., Hongwen, Y.: Analysis of vertical sectorization for HSPA on a system level: capacity and coverage. In: IEEE Vehicular Technology Conference (VTC Fall), pp. 1–5. IEEE, Quebec (2012)
7. Derege, K., Wegmann, B., Viering, I., Klein, A.: Mathematical model for vertical sectorization (VS) in AAS based LTE deployment. In: 11th International Symposium on Wireless Communication Systems, pp. 100–105. IEEE, Barcelona (2014)
8. Joydeep, A., Salam, A.: Optimizing vertical sectorization for high-rises. In: International Conference on Computing, Networking and Communications (ICNC), pp. 1–5. IEEE, Garden Grove (2015)
9. Siew, N., Zhong, L., Sai, W., Yong, C.: Optimizing radio network parameters for vertical sectorization via Taguchi's method. IEEE Trans. Veh. Technol. **65**, 860–869 (2015)
10. Jinping, N., Geoffrey, Y., Li, F., Wei, W., Weike, N., Xun, L.: Downtilts optimization and power allocation for vertical sectorization in AAS-based LTE-A downlink systems. In: IEEE 86th Vehicular Technology Conference (VTC-Fall), pp. 1–5. IEEE, Toronto (2017)
11. Muhammad, H., Usman, S., Jukka, L.: Analysis of vertical and horizontal sectorization in suburban environment using 3D ray tracing. In: 23rd International Conference on Telecommunications (ICT), pp. 1–6. IEEE, Thessaloniki (2016)
12. Fangchao, Z., et al.: A system level evaluation of vertical sectorization for active antenna system. In: 2015 IEEE/CIC International Conference on Communications, pp. 126–131. IEEE, China (2015)
13. Beneyam, B., Biruk, E., Bekele, M.: Planning of small cells in unlicensed band for Addis Ababa deployment scenario, Wireless World Research Forum Meeting, Durban (2018)
14. Preben, M., et al.: LTE system capacity compared to Shannon bounds. In: IEEE VTC Spring 2007, pp. 1234–1238. IEEE, Dublin (2007)

Quality Evaluation for Indoor Mobile Data Customers in Addis Ababa Business Area Using Data from Network Management System, Walk Test, Crowdsourcing and Subjective Survey

Abera Reesom Bisrat[1(✉)], Beneyam Berehanu Haile[2], Edward Mutafungwa[3], and Jyri Hämäläinen[3]

[1] Ethio telecom, Addis Ababa, Ethiopia
abera.reesom@ethiotelecom.et
[2] Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia
beneyamb.haile@aait.edu.et
[3] Department of Communications and Networks, Aalto University, Espoo, Finland
{edward.mutafungwa, jyri.hamalainen}@aalto.fi

Abstract. Introduction and penetration of high-speed and innovative mobile data services including social media and video services have been increasing significantly across the globe. This growth is driven by deployment and availability of capable and affordable mobile networks and smartphones. These trends have also been observed in Ethiopia as the country's sole incumbent operator ethio telecom has expanded coverage and capacity of its mobile data network, especially for universal mobile telecommunications system (UMTS) network. The successful provision of these innovative services requires mobile operators to have continuous monitoring of the quality of their networks and customer satisfaction. To that end, this monitoring allows the operator to make timely network optimization, marketing and other relevant decisions that enhance service quality. Traditionally measurement and analysis of network quality of service (QoS) is performed using data from network management system (NMS) and drive/walk test (DT/WT). In recent times, the usage QoS data from crowdsourcing (CS) is also emerging as useful monitoring tool. However, all these methods do not fully capture users' quality of experience (QoE), which is a function of not only network QoS but also other factors including subjective users' technical literacy and opinion. As a result, operators have recently provided attention not only for QoS but also QoE. Yet, most quality related analyses are performed using data from only subjective survey, NMS, DT/WT, or CS. It is not straightforward to find quality investigation based on data from all the methods. More importantly, such quality investigation has not been made in the context of Ethiopia's mobile market. In this paper, we perform QoS and QoE evaluation for UMTS mobile data service based on quality data from all NMS, WT, CS and subjective survey for a business area in Addis Ababa. The evaluation is made using downlink user throughput and mean opinion score (MOS) metrics. The obtained 50th percentile throughput results from NMS, WT and CS are good enough for web browsing and offline video streaming but not

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 164–175, 2019.
https://doi.org/10.1007/978-3-030-26630-1_14

for high definition (HD) video calls and live HD video streaming. Furthermore, CS provides optimistic results compared to NMS and WT and its 90th percentile throughput is very good even for live HD video streaming. Moreover, the obtained MOS results from subjective survey almost fit to the QoS results, thus indicating that the network QoS is a significant contributor to users' perceived QoE.

Keywords: Quality of experience · Quality of service · UMTS · Throughput · Mobile operators · Data service · Customer satisfaction · Mean opinion score · Network management system · Walk test · Crowdsourcing · Subjective survey

1 Introduction

Because of advancement of mobile networks, mobile device technologies and innovative data services, it has been observed that mobile data usage and traffic are significantly increasing [1, 2]. As a result, mobile operators are being challenged to successfully serve the increasing diverse mobile data services with satisfactory users' quality of experience (QoE). To meet the increasing demand, they are expected to continuously monitor and optimize their network and operations while expanding network capacity and coverage on a need basis.

Seeing Ethiopian mobile market, the country's incumbent operator ethio telecom with its Long Term Evolution (LTE) network in Addis Ababa and Universal Mobile Telecommunications System (UMTS) network in all urban areas of the country has been serving increasing number of active mobile data subscriptions that reach around 20 million as of the beginning of 2019 [3]. Due to the increased penetration of users, social media and video services, mobile data traffic has also been challenging the mobile networks. As Ethiopian mobile data market is not yet saturated in terms of data subscription, data price has recently seen up to 50% reduction and people income is expected to increase, mobile data demand is expected to continue increasing significantly. For instance, Addis Ababa mobile data demand is predicted to be around 20 petabytes per month in 2021, see Fig. 1 from [4].

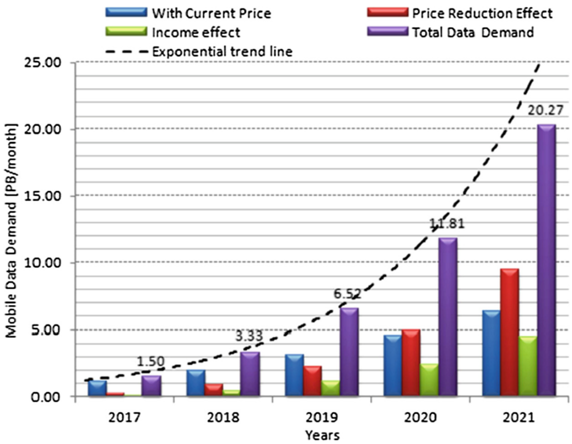


Fig. 1. Mobile data demand forecast for Addis Ababa [4]

To achieve and maintain successful operational business, mobile operators need to continuously meet QoE for their users [5]. User QoE for a given service is a function of quality of service (QoS), terminal capability, subscription, technical literacy, subjective opinion and other factors. To monitor and quantify QoS, operators traditionally collect and analyze data from their network management system (NMS) and drive/walk test (DT/WT) and recently from crowdsourcing (CS) systems complementing each other having their own pros and cons [6–8]. On the other hand, QoE is measured using mean opinion square (MOS) results from subjective surveys [2, 9]. For instance, ethio telecom continuously analyzes quality related key performance indicators (KPIs) and key quality indicators (KQIs) from its NMSs such as Huawei Nastar and undertakes DT/WT quality checks based on relevant user complaints but has not yet applied CS approach. Furthermore, to understand its customers' QoE, it has been performing subjective survey interviewing its individual and enterprise customers across the main cities of the nation twice a year. These quality related tasks are done by marketing departments of ethio telecom in collaboration with other stakeholders. We have noted that there is a need to reinforce quality handling process of departments under network and customer service divisions in ethio telecom while enhancing quality monitoring, quantifying and analyzing mechanisms so that significant number of customers' complaints, particularly for UMTS data users are successfully addressed [10].

Most QoE related researches are based on data from either NMS, DT/WT, CS or subjective survey or a combination of them. For instance, authors in [6] and [11] analyze QoE based on data from NMS and feedback survey while [7] presents QoE analysis based on quality data from DT, CS and subjective survey. Only subjective survey based quality analysis is also conducted in [2, 12, 13]. Thus, it is difficult to find QoE study using data from all NMS, DT/WT, CS and Subjective Survey. Furthermore, such a study has not been undertaken considering the Ethiopian mobile data market context. Yet, [14] presents QoS analysis for UMTS voice service in Addis Ababa using NMS and DT data while [15] compares DT and CS techniques based on collected data from Addis Ababa LTE networks.

In this work, we present QoE/QoS evaluation for UMTS data users based on data from all NMS, WT, CS and subjective survey. The evaluation is performed for a business area in Addis Ababa using downlink user throughput and MOS metrics. Furthermore, obtained throughput results are compared to required minimum throughput for browsing, offline video streaming, high definition (HD) video call and live HD video streaming, adapted from [16]. Results show that NMS, WT and CS provide 50th percentile throughput values that is good enough for browsing and offline video streaming but not for HD video call and live video streaming. For live HD video streaming, only the result from CS is optimum at the 90th percentile. Furthermore, we see that obtained MOS results from subjective survey aligned well with the QoS results. This shows that unsatisfied QoS is the main contributors to mobile users' dissatisfactions.

The remaining parts of the paper are organized as follows. Sections 2 and 3 provide description of quality measurement methods and data collection mechanism for target business area. Then Sect. 4 presents obtained results and interpretations. Finally, Sect. 5 presents concluding remarks of the work.

2 Quality Measurement Methods

2.1 Network Management System

QoS is measured from the network side using various KPIs and KQIs that are computed based on counters collected from mobile network elements and measurement reports from user equipment and application server [17]. Such quality measurement is operationally efficient to easily gather 24/7 quality data with low cost but its data spatial resolution is limited to cell-level.

For this work, geographic throughput data is collected from ethio telecom NMS called Nastar. Nastar is a comprehensive Huawei network monitoring and analysis platform that provides various UMTS quality analysis based on counters from base stations and reports from users [18, 19]. Besides data service quality data, Nastar also provides voice related quality data. The author in [14] presents voice QoS to evaluate voice quality of service in Addis Ababa using the Nastar platform and position of the Nastar is shown in Fig. 2.

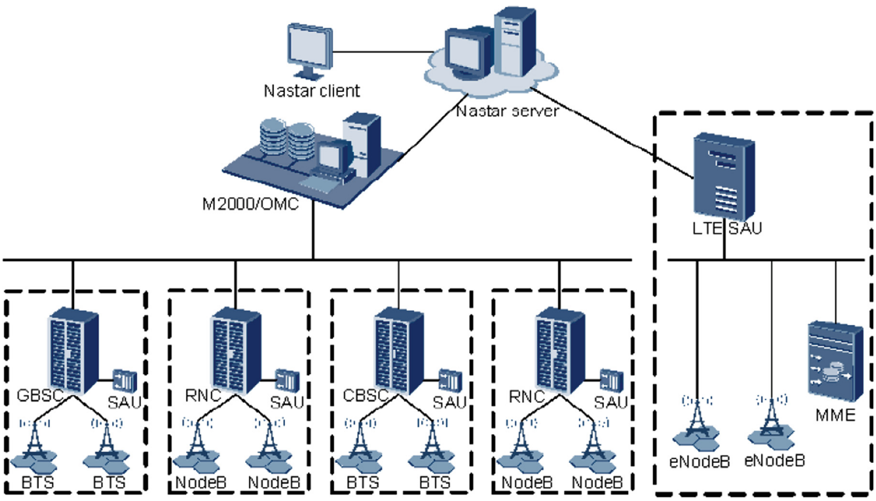


Fig. 2. Position of the Nastar in the network [14]

2.2 Drive/Walk Test

QoS can also be measured from user side and traditionally it is performed using DT where a team of experts with specialized and/or standard terminal devices in a car drives around target test area within coverage of network [7]. If target area is indoor, rather than driving the team performs WT moving around target indoor environment. Usually, target area for DT or WT is preplanned based on customer complaints and other network quality problems. Although DT or WT is providing very localized user side quality understanding, it is spatially and temporally limited and operationally costly due to required experts, devices and vehicle for DT, mainly when there is a need to cover large area.

For our study, we perform indoor WT for the target business area using most popular WT tool used by mobile operators called Nemo Handy [20]. We recall that there are quality related researches based on data from Nemo Handy based tests, see [14, 20–22].

2.3 Crowdsourcing

CS is another recently emerging user side QoS measurement method where mobile users report their QoS measured using a client application to a centralized server at which further overall QoS analysis is made [23]. The client application designed and developed for QoS measurement purpose is expected to be installed and used by users who might get various incentives for doing so. CS provides relatively inexpensive approach while its spatial and temporal coverage is limited by distribution of users who use the client application. Like NMS and DT/WT, accuracy of CS is very important and accuracy performance of couple of popular CS methods and Nemo Handy based DT are studied for Addis Ababa LTE network in [15].

The CS tool we apply for our work is called RTR-NetTest that is developed under the Austrian Regulatory Authority for Broadcasting and Telecommunications [24]. References [24–26] are exemplary works on how to use the tool for different QoS analysis.

2.4 Subjective Survey

Subjective survey is a standard method to articulate and quantify QoE by formulating relevant QoE questions, collecting their answers, undertaking interviews and/or focused group discussion [17]. To improve accuracy of subjective survey, besides formulating right questions, it needs proper sampling.

For this work, we formulate QoE questions related to UMTS data services that are targeted to mobile data customers in the selected business area. The survey questionnaire consists of 15 quality related questions and is prepared by filtering current quality issues and learning experience from [2, 12, 13]. To collect answers, the questionnaire is distributed to the target participants using hard copies and Google Form. Of the 15 quality related questions, the following four questions are sorted out for analysis.

- *How satisfied or dissatisfied are you with the overall quality of 3G data service?*
- *How satisfied or dissatisfied are you when you download files or video/music?*
- *How satisfied or dissatisfied are you when you access social media networks?*
- *How satisfied or dissatisfied are you with the quality of videos (in YouTube or video chatting)?*

3 Data Collection for Business Area of Addis Ababa

The indoor quality evaluation is performed for 20 enterprise buildings including ethio telecom headquarter within one of the business areas of Addis Ababa that is depicted in Fig. 3. The throughput data from the NMS, WT and CS are collected daily within office hours (9 am – 4 pm) in a study period of June and July months of 2018. Furthermore, answers of quality questionnaire from mobile data customers in the selected business area are also collected within the study period.

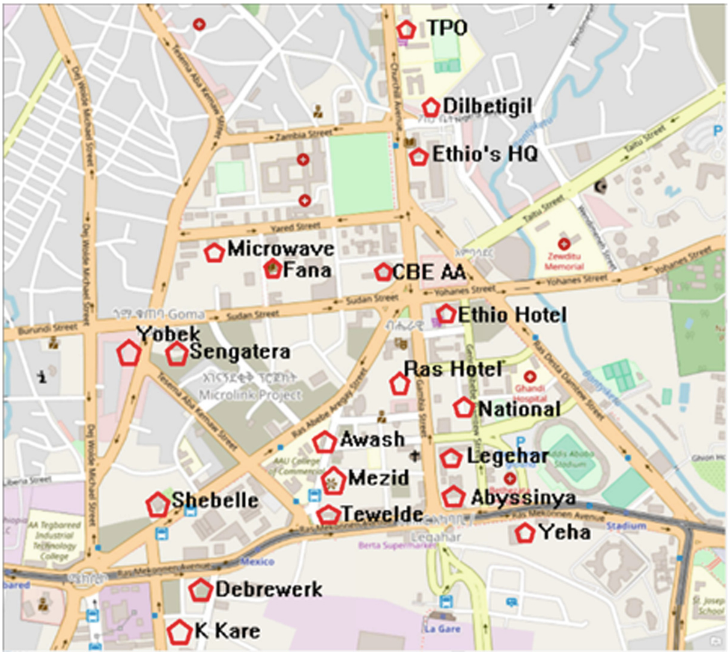


Fig. 3. Business area of quality evaluation

The NMS provides a 50 m geographic resolution for the downlink throughput and other metrics data for the whole Addis Ababa. Then we do data processing to localize this data to the buildings of the business area for further quality analysis. Accuracy of the localization effort has been validated by cross checking geographic coordinates of the buildings from the NMS with independently collected coordinates.

The indoor WT throughput data is collected with Nemo Handy on floors of the buildings within office hours of the study period. The applied Nemo Handy is configured well to generate required downlink throughput data.

The throughput data using the CS RTR is collected with the help of mobile data customers within the buildings to download and install RTR client in their smart phones and perform measurement test during office hours within the study period. The measurements are undertaken by 31 participants within and around their offices by locking their smartphones to 3G only. They use 21 unique smart phone models that are listed in Table 1 with their capabilities of supporting a maximum download speed for the UMTS network. Correct collections of the data from the RTR–NetTest server is made using technology type, the device model and timestamp on the screenshots of the tests that has been shared to us.

Table 1. List of mobile devices

Mobile devices	Maximum download speed (Mbps)
Max Download (Mbps), Che2-UL00, Galaxy S4 LTE-A, Galaxy S5, HTC One (HTC One_M8), HTC One E9PLUS dual sim, Huawei G620-L72, HUAWEI P8max, iPhone 6 Plus, Lenovo PB2-650M, MI NOTE LTE, SM-G965F, SM-J320W8 and TECNO-W5	42.2
Galaxy Grand DuoS, Galaxy Note 3, HUAWEI TIT-AL00, SM-C710F, SM-G530H, and T1-701u	21.1
iPhone 4s	14.4

To collect QoE data, the questionnaire is distributed to 129 staffs of enterprises in hard copies or Google doc on their need basis. Answers to questionnaire are collected within the study period. The WT and NMS data are collected near the offices of the questionnaire participants.

4 Results and Interpretations

To measure QoS relative to popularly used applications, obtained throughput results from NMS, WT and CS are presented by comparing them to the required minimum value for browsing, offline video streaming, HD video call and live HD video streaming. We use the required minimum throughput values which we have learnt from [16] for the aforementioned popular services and the values are listed in Table 2. The comparison is made by dividing obtained throughput results by the required minimum throughput.

Table 2. Minimum required download throughput [16]

Service type	Minimum requirement download speed (Mbps)
Web browsing	0.3
Offline video streaming (e.g. YouTube videos)	0.5
HD video calls (e.g. Skype)	1.5
Live HD video streaming (e.g. iPlayer)	3.2

4.1 QoS Results Compared to Browsing and Offline Video Streaming

Figures 4 and 5 show 10%-ile, 50%-ile and 90%-ile throughput results of NMS, WT and CS compared to web browsing and offline video streaming, respectively. As can be seen in the figures for both services, the 50%-ile and 90%-ile throughput ratio results are more than 1 indicating the majority and peak users achieve a very good QoS. On the other hand, the 10%-ile ratio results are below 1 indicating unsatisfactory QoS is experienced by at least 10% of unfavorable indoor users.

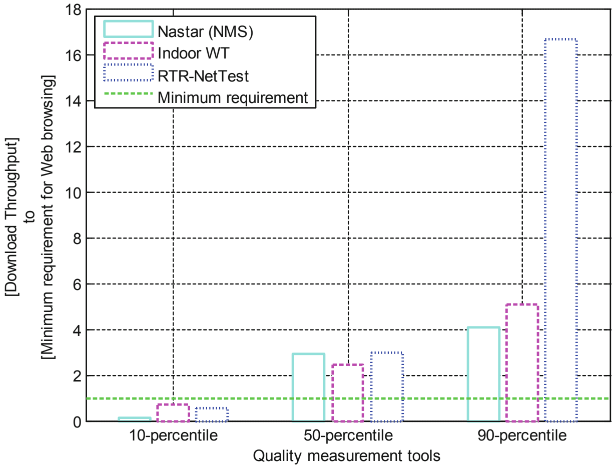


Fig. 4. Throughput results of NMS, WT & CS relative to browsing service

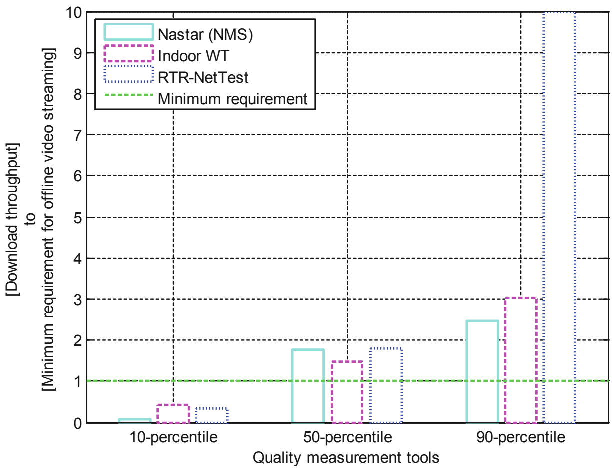


Fig. 5. Throughput results of NMS, WT & CS relative to offline video streaming

4.2 QoS Results Compared to HD Video Call and Live Streaming

The 10%-ile, 50%-ile and 90%-ile throughput ratio results of NMS, WT and CS compared to HD video calls and live HD video streaming are shown in Figs. 6 and 7, respectively. As can be seen from the figures, only the 90%-ile throughput ratio results from CS and WT are more than 1 for the video call while only the former is greater than 1 for the live video streaming. On the other hand, for both services, the 10%-ile and 50%-ile ratio results are far below 1. Thus, only the most optimistic CS results indicate only peak users can have satisfactory QoS for HD video call and live video streaming while majority of users cannot achieve satisfactory services according to any of the QoS measurement methods.

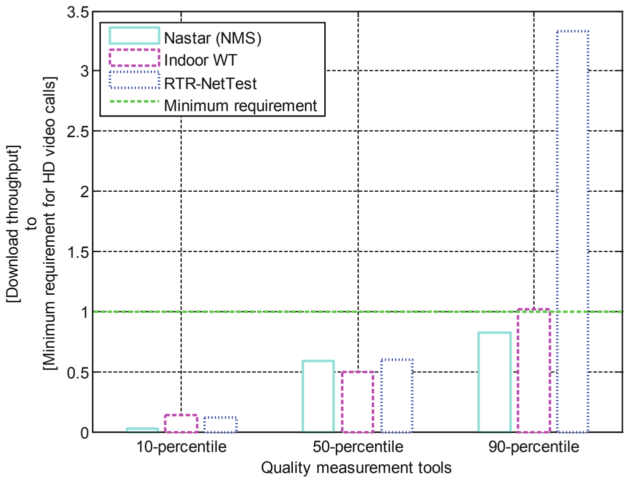


Fig. 6. Throughput results of NMS, WT & CS relative to HD video call

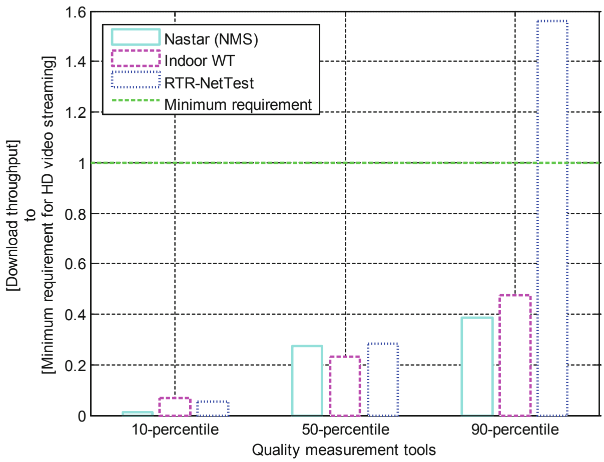


Fig. 7. Throughput results of NMS, WT & CS relative to live HD video streaming

4.3 Subjective Survey: Perception of Participants

MOS results of quality perceptions of survey participants on download speed, social media usage, video streaming usage and overall data usage are depicted in Fig. 8. The download speed, social media and video streaming perception questions are selected as they are the most popular use cases among staffs of the enterprises and related with the four services used for QoE comparison in the previous subsections.

In [9], it is mentioned that MOS value of greater than 3.5 is recommended to obtain a fair user experience for downlink bandwidth, uplink bandwidth and network delays. In addition to this, it is described in [27] that a perceived video quality is achieved when the MOS value is in the range of 3.5 and 4.0. In our case, we have taken a MOS value of 3.5 as a reference.

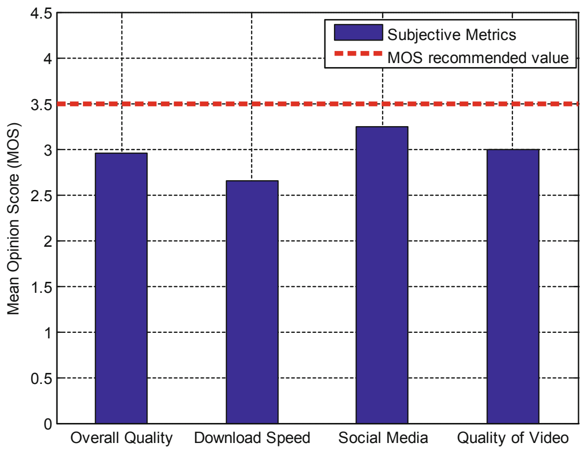


Fig. 8. MOS results of the perception survey

As can be seen in Fig. 8, the MOS values for download speed, social media, video streaming and overall data quality are 2.65, 3.25, 3.00 and 2.96, respectively. All MOS values are less than 3.5 which is the minimum value to claim that users have customer satisfaction, according to [9] and [27]. Thus, users on average are not satisfied with the quality of the overall UMTS data service and also specifically for the download speed, social media and video streaming usages.

5 Conclusion

To achieve successful operational business, mobile operators should meet quality of experience (QoE) expected by their customers. QoE is a function of network quality of service, users' literacy, terminal capability, subscription, opinion and other factors. To meet the required QoE, operators need to monitor, quantify and analyze QoE and QoS that are achieved by users. To that end, they undertake QoS and QoE analysis based on quality data measured and collected from network management system, drive/walk test, crowdsourcing and subjective survey, each with its pros and cons. Although various literature presents QoE and QoS analysis based on one or more of the quality measurement methods for various data usage context, it is not common to find indoor quality analysis based on data from all four measurement methods.

In this paper, we have presented QoS and QoE analysis for UMTS indoor data users for an Addis Ababa business area considering 20 enterprise buildings based on quality data from all the four methods. QoS is measured in terms of download speed and obtained results indicate that achieved quality for majority users is good enough for web browsing and offline video streaming (e.g. YouTube) but not for high definition video calls and live streaming services. The QoE results also show that users are not satisfied with UMTS download speed, social media and video streaming services. This indicates QoS has significant contribution for unsatisfied UMTS data users. Furthermore, crowdsourcing provides the most optimistic download throughput results that its accuracy should be checked well although it presents cost effective 24/7 QoS measurement methods for operators.

Acknowledgment. This research work is undertaken by the MSc. Telecom Engineering program jointly developed and implemented by ethio telecom and Addis Ababa Institute of Technology. To attain relevant research data, we are supported by staffs of ethio telecom (Departments of Service Management Center, Engineering and Customer Service Front Office).

References

1. Cisco: Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, White Paper (2017)
2. HOT TELECOM: 4 Keys to Shaping Quality of Experience (2015)
3. Ethio telecom: 2011 EC first half business performance report (2019)
4. Haile, B.B., Bulti, D.A., Zerihun. B.M.: On the relevance of capacity enhancing 5G technologies for Ethiopia. In: CONFERENCE 2017, EICTC, Addis Ababa (2017)
5. Lakhtaria, K.I.: Enhancing QoS and QoE in IMS enabled next generation networks. Int. J. Appl. Graph Theor. Wirel. Ad Hoc Netw. Sens. Netw. **2**(2), 61–71 (2010)
6. Schatz, R., Egger, S.: Vienna surfing – assessing mobile broadband quality in the field. In: Proceedings of the First ACM SIGCOMM Workshop on Measurements Up the Stack, Toronto, pp. 19–24 (2011)
7. Lin, H.Y., Hsieh, S.H., Chen, K.H., Wu, T.C.: Large-scale mobile internet performance measurement in Taiwan. In: 2014 6th International Conference on Ubiquitous and Future Networks, Shanghai, pp. 516–519 (2014)
8. Streijl, R.C., Winkler, S., Hands, D.S.: Mean Opinion Score (MOS) revisited: methods and applications, limitations and alternatives. Multimed. Syst. **22**, 213–227 (2016)
9. Pedro, C., Raimund, S.: Quality of experience in cloud services: survey and measurements. Comput. Netw. **68**, 149–165 (2018)
10. Universal Mobile Telecommunications System (UMTS): 3G Telecom Management: Principles and high level requirements, ETSI TS 132 101 V3.4.0 (2001)
11. Casas, P., Gardlo, B., Seufert, M., Wamser, F., Schatz R.: Taming QoE in cellular networks: from subjective lab studies to measurements in the Field. In: 2015 11th International Conference on Network and Service Management, Barcelona, pp. 237–245 (2015)

12. Reason to Complain Research, Ofcom Fieldwork (2016)
13. Customer Satisfaction Survey – Mobile Results, CICRA 16/10 (2016)
14. Ayele, J.: Quality of service evaluation of voice over UMTS network: the case of Addis Ababa. M.Sc. dissertation, Department of Electrical and Computer Engineering, Addis Ababa Institute of Technology, Addis Ababa (2017)
15. Alemayehu, B.: Evaluation of user side quality of service measurement techniques for Addis Ababa LTE data service. M.Sc. dissertation, Department of Electrical and Computer Engineering, Addis Ababa Institute of Technology, Addis Ababa (2018)
16. Ken's Tech Tips. https://kenstechtips.com/index.php/download-speeds-2g-3g-and-4g-actual-meaning. Accessed 01 Aug 2018
17. Rodriguez, D.Z., Rosa, R., Nunes, R.D., Affonso, E.T.: Assessment of quality-of-experience in telecommunication services. Int. J. Dig. Inf. Wirel. Commun. **6**(4), 241–259 (2016)
18. Wang, Y., Zhou, W., Zhang, P.: QoE Management in Wireless Networks. Springer, Heidelberg (2017)
19. Massam, P.: Intelligent CEM Making the NPS Connection, A Huawei White Paper, Huawei (2015)
20. Gupta, S., Garg, R.: Taxonomy of tools and techniques for network monitoring and quality assurance in 3G networks. Int. J. Comput. Appl. **120**(21), 34–41 (2015)
21. Afroz, F., Subramanian, R., Heidary, R., Sandrasegaran, K., Ahmed, S.: SINR, RSRP, RSSI and RSRQ measurements in long term evolution networks. Int. J. Wirel. Mob. Netw. **7**(4), 113–123 (2015)
22. Palazon, J.A., Sepulcre, M., Gozalvez, J., Orozco, J., Lopez, O.: Heterogeneous wireless connectivity for fixed and mobile sensing applications in industrial environments. In: Proceedings of the 16th IEEE International Conference on Emerging Technologies and Factory Automation, Toulouse, pp. 1–8 (2011)
23. BEREC Preliminary report in view of a common position on monitoring mobile coverage. BEREC BoR (17), 186 (2017)
24. Lipenbergs, E., Bobrovs, V., Ivanovs, G.: Investigation of service quality of measurement reference points for the internet services on mobile networks. Latvian J. Phys. Tech. Sci. **5**, 64–73 (2016)
25. Kousias, K., Lutu, A., Midoglu, C., Argyriou, A., Alay, O., Riegler, M.: The same, only different: contrasting mobile operator behavior from crowdsourced dataset. In: IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, Montreal, pp. 1–6 (2017)
26. Khatouni, A.S., et al.: Speedtest-like measurements in 3G/4G networks: the MONROE experience. In: IEEE 2017 29th International Teletraffic Congress (ITC 29), Genoa, pp. 169–177 (2017)
27. Elkotob, M., Grandlund, D., Andersson, K., Ahlund, C.: Multimedia QoE optimized management using prediction and statistical learning. In: 35th Annual IEEE Conference on Local Computer Networks, Denver, pp. 324–327 (2010)

Performance Evaluation of 6-Sector Site and Small Cell for Addis Ababa UMTS Network

Tirufire Aberra Fulle[1](✉) and Beneyam Berehanu Haile[2](✉)

[1] Ethiotelecom, Addis Ababa, Ethiopia
tirufire.aberra@ethiotelecom.et
[2] Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia
beneyamb.haile@aait.edu.et

Abstract. The demand for high mobile data rate has been increasing due to introduction of various innovative mobile data services to improve our life. To accommodate the increasing data demand with satisfactory quality of service operators must continuously improve their network capacity. To that end, besides deploying fourth generation network, already operational third generation (3G) network capacity can be improved by applying some capacity improvement techniques including six-sectorization, densification, multiple input multiple output and using multiple carriers. Although these capacity improving techniques have been well investigated theoretically and practically, their performance is not studied well to address capacity challenges of most widely used 3G data network in Ethiopia. In this paper, we select six-sectorization and small cell based on relevance and device availability trends and investigate their performance in addressing capacity challenges of selected Addis Ababa 3G deployment scenario considering realistic capacity demand, propagation environment and legacy network topology. Propagation computation and network simulation are made using WinProp and dominant path model based on building and topography maps are used for propagation computation. Performance results show that deployment of six-sector or/and small cells provide significant performance gain compared to already existing full 3 sector cells. For instance, six-sector deployment provides a 284.23% average downlink throughput gain. Furthermore, deployment of small cells with 5% penetration presents a 114.67%, average throughput gains over the existing full 3-sector.

Keywords: UMTS · 3G · Capacity challenges · Capacity enhancement techniques · Six sectorization · Small cell · Addis Ababa

1 Introduction

Universal Mobile Telecommunication System (UMTS) is the most widely deployed 3G mobile system that presents a range of data, voice, text, and multimedia-based mobile services [1]. To improve quality of data services the first UMTS standard Release 99 (R99) packet domain has been evolved to High-Speed Packet Access (HSPA) and

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 176–187, 2019.
https://doi.org/10.1007/978-3-030-26630-1_15

HSPA+ in the consecutive Third-Generation Partnership Project (3GPP) releases applying various advanced enhancements including multiple carriers, higher order modulation, higher order, Multiple Input Multiple Output (MIMO) and other features. To accommodate demand of Third Generation (3G) data users, most mobile operators across the globe have already deployed HSPA/HSPA+ network and Ethiopian incumbent ethio telecom has also deployed them in almost all cities and other parts of the country (Ethiopian Telecommunications Corporation [2]: click to the Ethiopian Millennium: special issue, September 2009).

The 3G technology share in 2017 and forecasted 3G share in 2025 both across the globe and in Sub-Saharan countries is shown in Fig. 1 (GSM Association [3]: The Mobile Economy 2018). As can be seen from the figure, 3G has been and will be significantly contributing to accommodating mobile services. Its role is more significant in Sub-Sahara countries including Ethiopia.

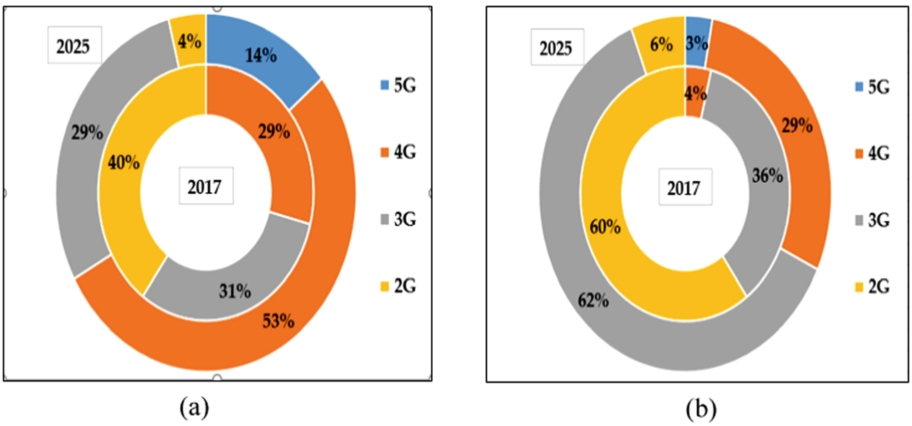


Fig. 1. Technology share in the (a) world and (b) Sub-Saharan Africa [13]

As quality mobile data demand continuously increasing due of innovative data services, besides preparing and deploying 4G network, mobile operators shall continuously optimize their existing 3G network. Currently, there is a growing demand for data services and the adoption of mobile data service is increasing rapidly ([4]). Expanding mobile network to support these increasing data demand and introducing new relevant technologies is an important work for mobile operators even before it creates capacity limitations.

There has been some progress in recent years to meet the increasing data demand in Addis Ababa [5]. Nonetheless, several challenges remain in terms of how to enhance and expand these networks to meet the growing demands of an increasingly digital economy and society. To improve capacity of UMTS network without rework, proper selection, study and evaluation of capacity improvement techniques is important.

The paper presented in [6] show that before the deployment of a network or before expansion work, different options should be evaluated using different metrics. Yet, these techniques are not evaluated well in the context of Ethiopia considering existing network topology, realistic capacity demand and accurate propagation environment so that viable technics can be selected and implemented.

In this paper, first analyzed network traffic and connection failures data of existing Addis Ababa 3G network from network management system and we have seen that there is considerable amount of capacity challenges that need to be addressed. Secondly, six-sectorization and small cell technologies are selected based on relevance in terms of user device availability trend and then their performances are investigated for selected Addis Ababa deployment scenario. Note that currently only 20% of user devices support MIMO or multicarrier operations [7].

Propagation computation are made using deterministic dominant path model based on building and terrain map information of the selected Addis Ababa deployment scenario [8]. Network simulation are made using WinProp implementation [9] and performance evaluations are undertaken using Carrier-to-Interference-Plus-Noise-Ratio (CINR) and throughput metrics. Results show that both six sectorization, small cell and their hybrid usage provide significant CINR and throughput gains compared to already existing 3-sector macro only network. For instance, full 6 sector and femto with 5% penetration improved the CINR by 2.9 and 0.3 dB respectively.

The paper is organized as follows. In Sect. 2, we provide a brief background on selected capacity improvement techniques and Sect. 3, baseline and enhancement scenarios are described. In Sect. 4, we describe the deployment scenario and simulation assumptions. Finally, in Sect. 5 and 6, we present detail simulation results and conclusion.

2 Background on 3G Capacity Improvement Techniques

To address UMTS network capacity challenges, there are some capacity improvement techniques including six sectorization, small cell, and advanced MIMO that are not yet implemented in existing WCDMA/HSPA/HSPA+ network. These capacity enhancement techniques are studied well from theoretical and practical perspectives [10–12]. Capacity enhancement techniques such as adding a new site or spectrum are challenging for operators since new site acquisition is becoming difficult and a spectrum is scarce resources. Moreover, in Ethiopia, user devices availability which supports multiple frequency bands for UMTS is another challenge. As a result, upgrading to 6-sector and using small cells are more feasible enhancement options.

2.1 Six Sector

Sectorization is one important capacity enhancing method by using multiple directional antennas unlike Omni directional antenna case. Six-sector site can be implemented using 6 directional antennas or nowadays applying dual beam technology with only three antennas. The latter is advantageous in terms of reducing load on the tower and interference leakage [13].

Different research has showed that six-sector site can be used to improve network capacity, coverage, and quality [13–19]. In [14], a case study of hotspots 6-sectored with maximum DL power of 20 W, showed that there was a 23% increase in coverage and 1.8x capacity increase. In paper [15], further demonstrated simultaneous users increase when the number of sectors increases for 384 kbps. For example, for bit energy to noise ratio (Eb/No) value of 4 dB, 88 users are simultaneously connected in 6-sectors but for 3-sectors, 45 users are simultaneously connected. Furthermore, a study in [16] showed that not only higher order sectorization but also the beam width of the antenna affects capacity and coverage. The sectorization gain can vary from 1 for Omni to 5.07 for 6-sector with 33-degree orientation in DL.

Moreover, a study conducted in central London area of Telefonica UK's network with 51 sites is presented in [17]. For inter-site distances of 500 m, 6-sector with 34-degree beamwidth has 1.82x capacity gain over 3-sector at 62-degree beamwidth. The research also shows that the gain increases as the number of users increases. For example, network simulation showed that 6-sector has 1.41x, 1.71x, 1.85x and 1.87x capacity gain over 3-sector with 500, 1000, 2000 and uniform traffic. Additionally, it is shown that six-sectorization with dual-beam increases site capacity by 50% to 85% and improves the coverage performance by 3 dB [13]. For instances, AIS has progressively used different method to improve the service but it is 6-sector cell with dual beam technology that showed remarkable performance up to 70% capacity gain [18]. Similarly, Globe's 2100 MHz UMTS network downlink data traffic increase by 50%, uplink data traffic by 74.73% and voice traffic by 31.69% [19].

2.2 Introduction of Small Cells

The second technique selected that can improve the capacity of UMTS network is deploying small cells. Small cells are low-power radio access nodes which can be used in densely populated areas and indoor hotspots where adding spectrums or macro base stations may not visible. Small cell has easy site acquisition and flexible deployment to improve the user experience [20].

Femtocells bring benefits such as increasing cell density at realistic costs, enhancing indoors coverage, improving spectrum usage efficiency, reducing the load from the macro layer, improving service for outdoor users and giving a better SINR distribution (MIMO works better). Some femtocell deployment challenges have already been overcome. For example, the interference issues have now been resolved by introducing interference mitigation techniques which enable co-channel operation between the femtocell and macrocell. Study done in [21], shows that the use of adaptive UL attenuation for the uplink and downlink transmit power calibration of femtocell to limit interference to the macro layer has shown both capacity and coverage performance gains. Another challenge of small cells is backhaul availability. Backhaul network provides connectivity between the core network, and the small cells to deliver the service with the desired QoS level [22–24]. Currently, there are different options to choose from like using self-backhauling using LTE macrocell ([25]).

3 Baseline and Enhancement Scenarios

3.1 Baseline Scenarios

Four Carrier per Cell Full 3-Sector: It has been deployed according to real network configure position, antenna azimuths, tilt and a total of 4 carriers are used except two NodeBs which has 3 and 2 carriers per cell and named as four-carrier per cell full 3-sector (4-SC full 3-sector) (Fig. 2).

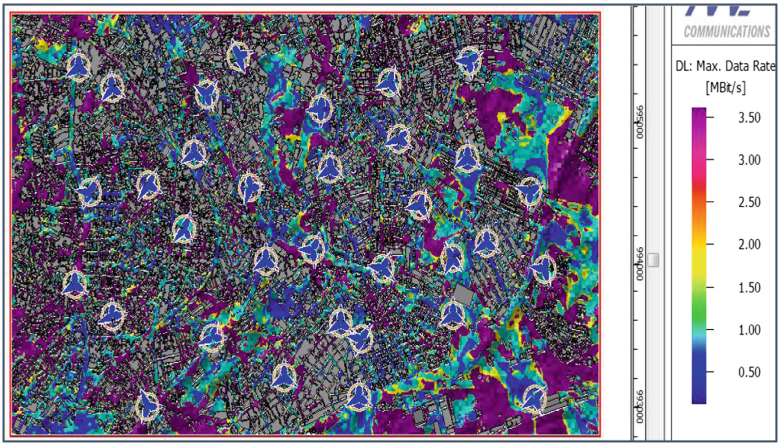


Fig. 2. Data rate plot of full 3-sector with four single carriers per cell

One Carrier per Cell Full 3 Sector: In this scenario, macrocells have been deployed according to real network configured positions, antenna azimuths and tilt but one carrier per cell is used named as one carrier per cell full 3-sector (1-SC full 3-sector) and four carriers are reused.

3.2 Enhancement Scenarios

Full Six Sector: All Node Bs are 6-sector with same positions as the baseline scenario but with different antenna azimuth and tilt (5-degree). One carrier per cell is used and a total of 4 carriers are reused (Fig. 3).

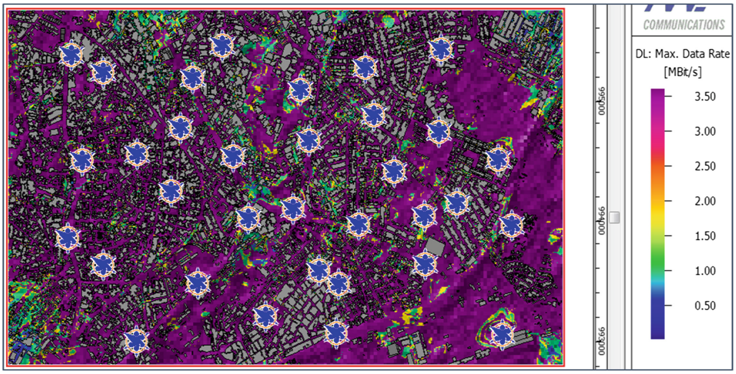


Fig. 3. Data rate plot of full 6-sector

Hybrid: Containing both 3 and 6-sectored Node Bs with same site positions as the first scenario but with different antenna azimuth and tilt. The placement of 6-sector Node Bs is based on data rate from experiment 1(1-SC full 3-sector). 3-sector Node Bs are upgraded to 6-sectored if have low data rate or are within a layer of worst cells. One carrier per cell is used and a total of 4 carriers are reused (Fig. 4).

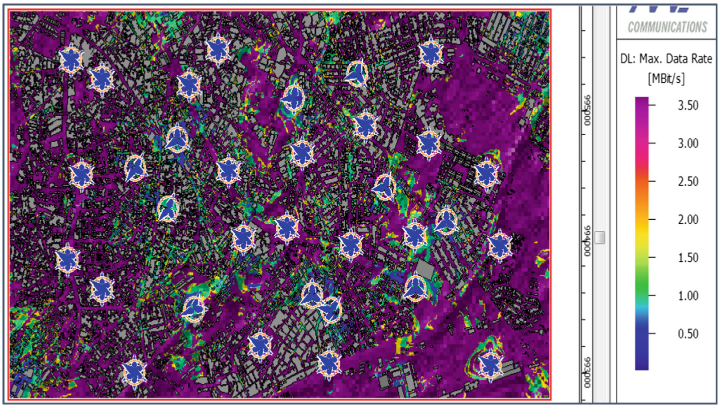


Fig. 4. Data rate plot of hybrid

Three Sectors with 0.05 Femto Penetration: Three sectors with 0.05 femto penetration using orthogonal deployment between macrocells and femtocells. The placement of femtocell is based on data rate values from the first experiment which is 1-SC full 3-sector (circle marked by orange are Omni directional small cells) (Fig. 5).

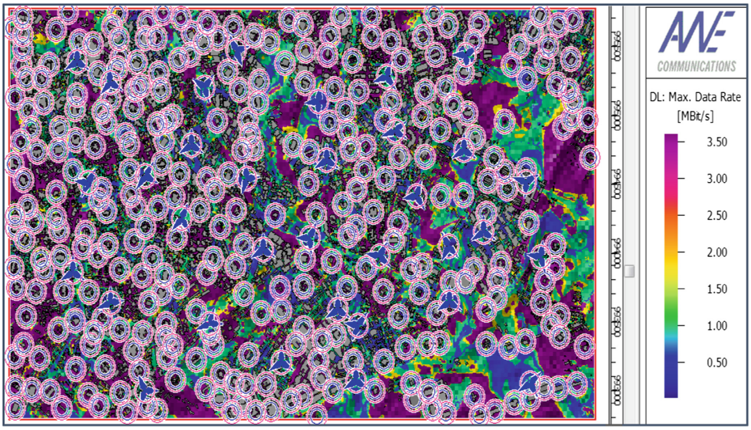


Fig. 5. Data rate plot of three sectors with 0.05 femto penetration (Color figure online)

Hybrid with 0.005 Femto Penetration: Hybrid with 0.005 femto penetration using orthogonal deployment. The placement of femtocell is based on data rate values from the third experiment that is hybrid. Specific parameters related to each scenario is shown in Table 1 (Fig. 6).

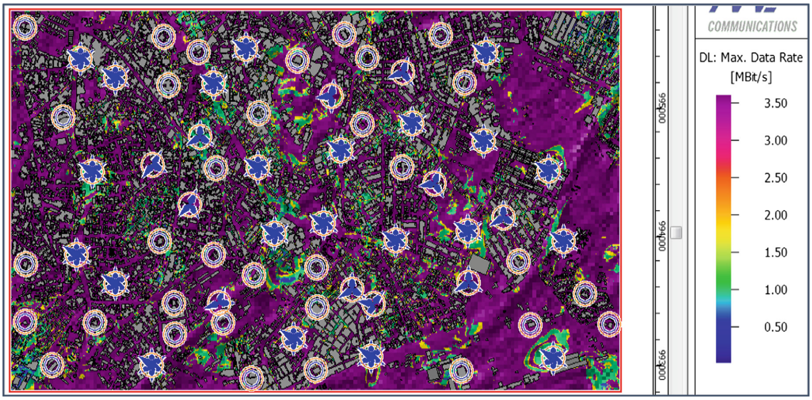


Fig. 6. Data rate plot of hybrid with 0.005 femto penetration

Table 1. Parameters for baseline and enhancement scenarios

Scenarios	Number of macrocells	Number of Femtocells
Single carrier per cell Full 3-Sector	99	–
Four carrier per cell Full 3-Sector	396	–
Full 6-Sector	198	–
Hybrid (both 3 and 6-sectored)	165	–
Femto with 5% penetration	99	308
Hybrid with 0.55% femto penetration	165	31

4 Deployment Scenario and Simulation Assumptions

4.1 Deployment Scenario

Addis Ababa is the capital city of Ethiopia that shares most urban population [26]. World Bank study shows, in Ethiopia, an increase in urban population as % of total population from 18.6% in 2013 to 20.4% in 2017. For this study, 15Sq Km is used in bole area which has 2,694.1 population density per sq. m and also densely populated buildings [27]. In this regard, network deployment in this area will have higher propagation losses which presents the worst-case scenario. Currently, 33 macro sites with 3 sector per site are deployed in the specified location. Figure 7 shows topographic with building databases.

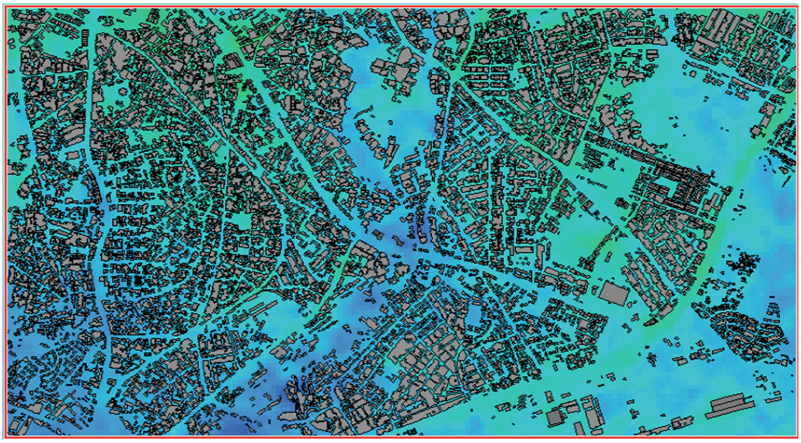


Fig. 7. Topographic database with building data on it

4.2 Simulation Parameters and Assumptions

General simulation assumptions and parameters are shown in Table 2.

Table 2. General simulation assumptions and parameters

Parameter	Values/Assumptions
Air interface	Five 5 MHz carrier, 3.84 Mchips/s, Duplexing scheme FDD
Prediction height	1.5 m
Propagation	Dominant Path Model
Macro Node B (MNB)	Power 43 dBm, antenna gain 17.5 dBi, 4 dB noise figure, cable loss 2 dB/100 m, Coupling loss 10, Radiating cable loss 2 dB
Antenna	Aggison ADU451819 for 3-sector MNB [28]

(*continued*)

Table 2. (*continued*)

Parameter	Values/Assumptions
Femto	Omnidirectional antenna with a power of 23 dBm
UE	Omnidirectional 1 dB body losses, 6 dB noise figure
Femtocell height	From 2–7 m
User distribution	Uniform
Buildings	Heights 3–56 m, penetration loss: 10 dB

5 Simulation Results and Discussions

The CDF of different deployment scenarios is presented Figs. 8 and 9. As it can be seen, full 6-sectored deployment significantly improves the CINR by delivering 2.9 dB, 2.2 dB and 0.5 dB gains at 10, 50 and 90-percentile respectively compared to 1-SC full 3-sector. Further, full 6-sector deployment gives 2.8 dB, 2.1 dB and 0.4 dB gains compared to 4-SC full 3-sector at 10, 50 and 90-percentiles respectively which indicates 4-SC full 3-sector has 0.1 dB gain compares to 1-SC full 3-sector. Hybrid deployment has also shown remarkable performances at 10, 50 and 90-percentiles with 2.2 dB, 2.1 dB, and 0.5 dB gains correspondingly. With the addition of femto at 0.55% penetration to hybrid, the CINR increases by 0.4 dB at 10-percentile compared to hybrid. Moreover, femto with 5% penetration improved the performance by 0.3 dB, 0.3 dB and 0.2 dB relative to 1-SC 3-sector case at 10, 50 and 90-percentile.

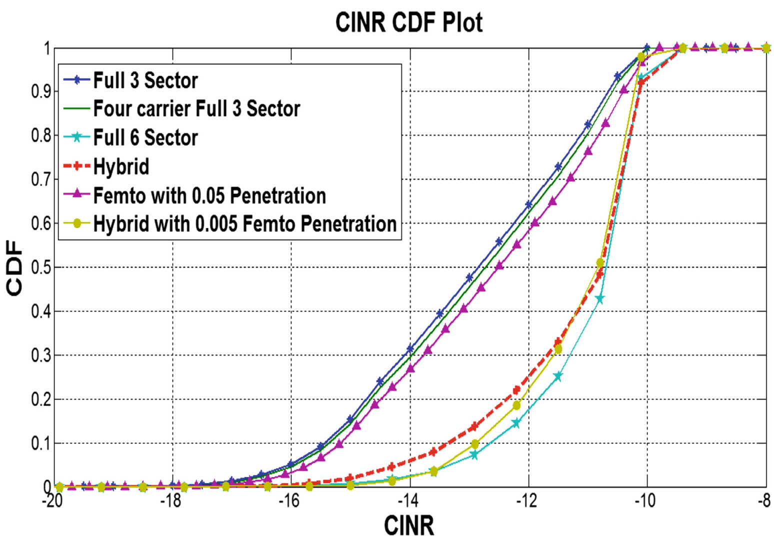


Fig. 8. CINR plot for the 6 scenarios tested

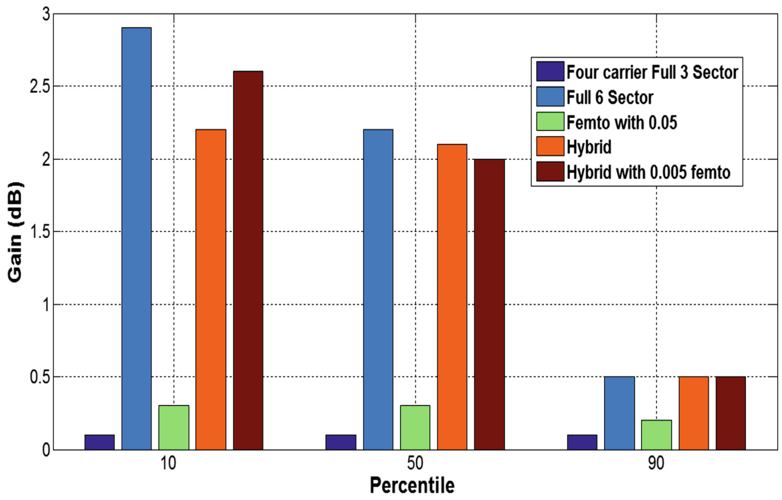


Fig. 9. Performance compared to one carrier full three-sector

CDF of DL maximum throughputs are presented in Fig. 10 and performance gains when compared to full 3-sector deployments at 10, 50 and 90-percentiles are shown in Fig. 11. The results show that full 6-sectored deployment delivers 284.23%, 280.00%, and 118.81% gain at 10, 50 and 90-percentile respectively when compared to 1-SC full three-sector. Moreover, full 6-sector compared with 4-SC Full 3-sector, improves the throughput by 2.44x, 2.72x and 1.17x at 10, 50 and 90-percentile respectively. In addition, femto 0.05 penetration, hybrid, and hybrid with 0.005 femto improved the DL throughput by 114.67%, 231.05%, and 278.73% compared to 1-SC full 3-sector respectively at 10-percentile. While in at 50-percentile hybrid, femto with 0.05 penetration and hybrid with 0.005 femto deliver 232.26%, 104.86%, and 230.02% gains compared to 1-SC full 3-sector.

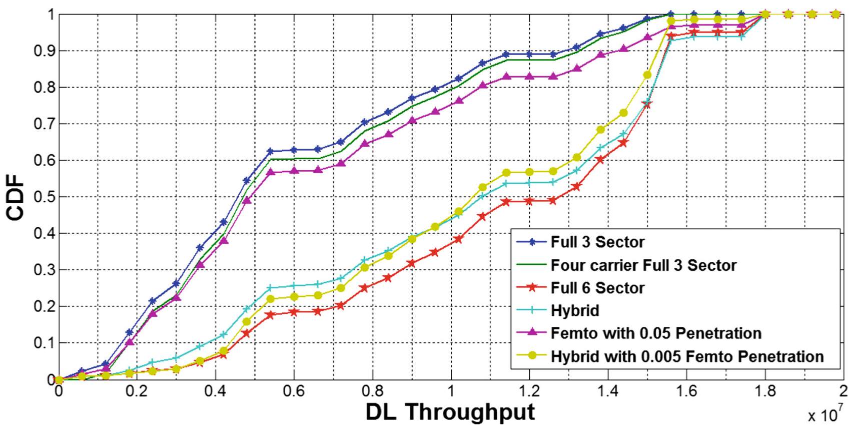


Fig. 10. CDF of DL throughput (Mbps)

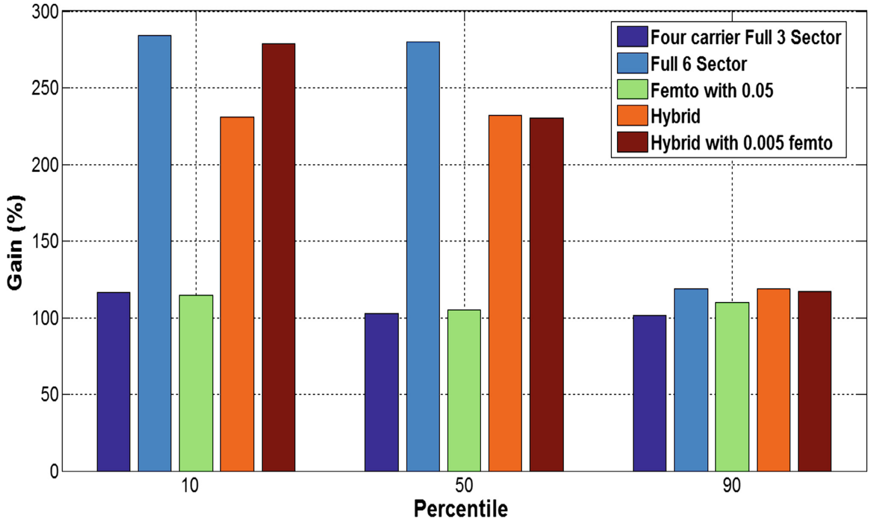


Fig. 11. DL throughput percentile compared to one carrier full 3-sector

6 Conclusions and Future Work

This paper has first studied network challenges from network management system data. We observed that, there is significant amount of network failures to due to limited channel element and power resources. To overcome these resource challenges, we have evaluated various different improvement techniques namely MIMO, carrier aggregation, adding another band, macro densification, six sector cell and small cell. Then due to their practicability and relevance six sector site and small cell technologies are selected.

The Simulation study highlighted the potential benefits of using 6-sectorization and Femtocells. Two different deployment strategies for 6 sectors, full and hybrid, have been studied and potential benefits are shown. Full 6 sector has a remarkable performance improvement compared to all studied scenarios. At 10-percentile, Full 6 sectored deployment delivers 2.84x DL throughput and femto with 5% penetration has 1.14x compared to 1-SC full three sector. Future works is needed to select and evaluate different backhaul options.

Acknowledgment. The research is performed jointly by ethio telecom and AAiT with the support of Aalto University.

References

1. Holma, H., Toskala, A.: WCDMA for UMTS HSPA Evolution and LTE, 5th edn. John Wiley & Sons, Ltd. (2010)
2. Ethiopian Telecommunications Corporation (ETC): click to the Ethiopian Millennium: special issue, September 2009
3. GSM Association (GSMA): The Mobile Economy 2018
4. Haile, B.B., Bulti, D.A., Zerihun, B.M.: On the relevance of capacity enhancing 5G technologies for Ethiopia. In: 10th Ethiopian ICT Annual Conference (EICTC-2017) (2017)

5. Ministry of Finance and Economic Development: Growth and Transformation Plan Annual Progress Report for F.Y. 2012/13, Ministry of Finance and Economic Development, Addis Ababa, February 2014
6. Real wireless: Techniques for increasing the capacity of wireless broadband networks: UK, 2012–2030, April 2012
7. Altair: WinProp - Propagation Modeling. https://altairhyperworks.com/product/FEKO/WinProp-Propagation-Modeling, Accessed 29 Aug 2018
8. Altair: About. https://altair.com/about/. Accessed 5 Aug 2018
9. Huawei: Service & experience quality analyst. Software
10. Holma, H., Toskala, A., Tapia, P.: HSPA + Evolution To Release 12 Performance and Optimization. John Wiley & Sons (2014)
11. Holma, H., Toskala, A.: HSDPA/HSUPA for UMTS, 5th edn. John Wiley & Sons, Ltd. (2010)
12. Chapman, T., Larsson, E., von Wrycza, P., Dahlman, E., Parkvall, S., Skold, J.: HSPA Evolution the Fundamentals for Mobile Broadband. Elsevier Ltd (2015)
13. Sorrells, P.: Twin beam technology adds immediate capacity without additional antennas, White Paper, CommScope (2013)
14. Hagerman, B., Imbeni, D., Barta, J., Pollard, A., Wohlmuth, R., Cosimini, P.: WCDMA 6-sector deployment-case study of a real installed UMTS-FDD network. In: IEEE 63rd Vehicular Technology, Melbourne, Australia (2006)
15. Fazlul, A., Mohammad, M., Baitul, M., Ka, M., Fokhray, M.: Performance analysis of UMTS cellular network using sectorization based on capacity and coverage. Int. J. Adv. Comput. Sci. Appl. **2**(6), 48–55 (2011)
16. Wacker, A., Laiho-Steffens, J., Sipila, K., Heiska, K.: The impact of the base station sectorization on WCDMA radio network performance. In: IEEE 50th Vehicular Technology, Amsterdam, Netherlands (1999)
17. Joyce, R., Morris, D., Brown, S., Vyas, D., Zhang, L.: Higher order horizontal sectorization gains for 6, 9, 12 and 15 sectored cell sites in a 3GPP/HSPA + network. IEEE Trans. Veh. Technol. **65**(5), 3440–3449 (2016)
18. Ashish, P.: HOW TO OPERATE: Philippines Globe: New network configuration, new MBB gains, ISSUE 80, Globe, October 2016
19. Huawei: Thailand AIS: A Shift from Embarrassing to Wonderful. http://carrier.huawei.com/en/success-stories/wireless-network/03-ubiquitious-mbb/thailand%20ais. Accessed 14 July 2018
20. Small Cell Forum: Small cells - what's the big idea? White Paper, UK, February 2012
21. Yavuz, M., et al.: Interference management and performance analysis of UMTS/HSPA + femtocells. IEEE Commun. Mag. **47**, 102–109 (2009)
22. Ericson: It all comes back to Backhaul, White Paper, February 2012
23. Zhang, J., la Roche, G.D.: Femtocells: Technologies and Deployment. John Wiley & Sons (2010)
24. Next Generation Mobile Networks (NGMN): Small Cell Backhaul Requirements, White Paper, NGMN Alliance, June 2012
25. Amin, P., Kibret, N.S., Mutafungwa, E., Haile, B.B., Hamalainen, J., Nurminen, J.K.: Performance study for off-grid selfbackhauled small cells in dense informal settlements. In: Proceedings of the 25th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (IEEE PIMRC 2014), pp. 1652–1657, September 2014
26. The World Bank: World Development Indicators. http://databank.worldbank.org/data/reports.aspx?source=2&series=SP.URB.TOTL.IN.ZS&country=#selectedDimension_WDI_Ctry. Accessed 03 Mar 2018
27. Addis Ababa city government, population. http://www.addisababa.gov.et/es/web/guest/bole-sub-city. Accessed 03 Mar 2018
28. Huawei Technologies Co., Ltd.: Antenna Specification, ADU451819 datasheet (2014)

Techno-Economic Investigation of LTE-Advanced Deployment for Africa: Case of Addis Ababa City

Gizachew Hailegebriel Mako[1(✉)], Beneyam Berehanu Haile[2(✉)], and Edward Mutafungwa[3(✉)]

[1] Ethio Telecom, Addis Ababa, Ethiopia
gizachew.hgebriel@ethiotelecom.et
[2] Addis Ababa Institute of Technology, Addis Ababa, Ethiopia
beneyamb.haile@aait.edu.et
[3] Department of Communications and Networks, Aalto University, Espoo, Finland
edward.mutafungwa@aalto.fi

Abstract. The proliferation of innovative data services and subsequent increase in mobile data traffic continues to be a prevailing trend across the globe. For instance, according to Ericsson, mobile data traffic is forecast to reach 71 Exabyte per month in 2022 from 8.8 Exabyte per month in 2016. Moreover, it is noted that Sub-Saharan Africa region including Ethiopia exhibit fastest growth rates in terms of mobile data demand. To accommodate this increasing demand, beside Long Term Evolution (LTE), mobile operators are optimizing their network and deploying LTE-Advanced (LTE-A) and LTE-A Pro technologies. Having operational LTE network in capital Addis Ababa, the Ethiopian incumbent sole operator ethio telecom is still at the early preparation phase of upgrading their network to include LTE-A capabilities. To that end, there is a need of understanding techno-economically viable LTE-A deployment scenarios to meet the local demand. In this paper, relevant LTE-A deployment scenarios are developed for Addis Ababa using the scenario-planning method considering the local context. To observe viability of the developed scenarios, we also perform techno-economic analysis using modified TERA model assuming 6 years study period and 10% discount rate. The findings of our study note that the net present value and payback period results show that LTE-A scenario that applies Progressive Deployment with Out-of-band small cells is the most viable scenario for Addis Ababa.

Keywords: LTE · LTE-Advanced · Scenario planning · Carrier Aggregation · Small cell · Techno-economics · TEA · TERA · Addis Ababa · Africa

1 Introduction

Mobile data demand and corresponding traffic due to introduction of various data services have been significantly increasing across the globe [28]. Various sources have also forecasted that the growth continue exponentially mainly due to emerging data

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 188–202, 2019.
https://doi.org/10.1007/978-3-030-26630-1_16

intensive quality video services including augmented reality, virtual reality and ultra-high definition live video streaming. For instance, Ericsson report has forecasted that in 2022 global mobile traffic generated by mobile data users will reach 71 Exabyte per month from 8.8 Exabyte per month in 2016 [28]. Same report also indicated that similar significant data traffic growth will be seen in Sub-Saharan Africa with largest growth rate. Mainly due to increasing new data users, mobile based social media and video service penetrations. The Ethiopian market also experiencing significant mobile data traffic growth, mainly in the capital Addis Ababa [3, 29, 30].

To successfully deliver the increasing mobile data demand with satisfactory quality of service, mobile operators expanding their network capacity by deploying LTE-A and LTE-A Pro technologies [24]. These technologies enhance network capacity by exploiting different sub-technologies such as Carrier Aggregation (CA), enhanced Multiple Input Multiple Output (eMIMO) and Heterogeneous Networks (HetNet).

Ethio telecom, has partially deployed LTE in some parts of Addis Ababa. To meet the increasing capacity demand, besides expanding the LTE network coverage, there is a need in ethio telecom to deploy LTE-A in Addis Ababa to provide further capacity enhancements [3]. However, there is no concrete understanding on which LTE-A scenario is most viable for the Addis Ababa context, from both technical and economical perspectives. Addis Ababa has its own context in terms of existing network, user demand distribution, spectrum availability and other factors.

The aim of this paper is to identify most likely LTE-A deployment scenarios for Addis Ababa and perform techno-economic analysis on them to analyze their viability, assuming a six year study period and 10% discount rate. Scenario planning method is applied to identify suitable LTE-A scenarios while modified TERA model is used to analyze techno-economic viability of the scenarios [27]. Based on two main uncertainties, full or progressive deployment and out-of-band or in-band small cells, four potential scenarios are devolved: *Full LTE-A Deployment with Out-of-band Small Cells (SC1), Progressive LTE-A Deployment with Out-of-band Small Cells (SC2), Full LTE-A Deployment with In-band Small Cells (SC3)* and *Progressive LTE-A Deployment with In-band Small Cells (SC4).* Techno-economic evaluation results, i.e. Net Present Value (NPV) and Payback Period (PP), show that the LTE-A scenario with *Progressive LTE-A Deployment with Out-of-band Small Cells* is the most viable option for Addis Ababa.

The remaining part of the paper is organized as follows. Background on LTE-A technologies is provided in Sect. 2 and identification of potential LTE-A scenarios for Addis Ababa is discussed in Sect. 3. Then description of methodology, assumptions and parameter values for the techno-economic analysis are presented in Sect. 4 which is followed by presentation of results of the techno-economic analysis in Sect. 5. Finally, concluding remarks are provided in Sect. 6.

2 Background on LTE-A Technologies

Base line LTE is evolved to LTE-A that is standardized in 3rd Generation Partnership Project (3GPP) since Release 10 and it presents further enhancements to LTE in terms of network capacity, spectral efficiency, energy efficiency, interference mitigation for dense network, as well as, cost-per-bit [1, 2]. To achieve the enhancements LTE-A exploits four key technology components: CA, eMIMO, densification and Coordinated Multipoint (CoMP) transmission & reception.

2.1 Carrier Aggregation (CA)

CA is one important LTE-A technology component that enhance network capacity by increasing system bandwidth by combining component carriers (CCs) from same or different frequency bands [1]. The CCs can have any of LTE Release 8 bandwidths: 1.4, 3, 5, 10, 15, 20 MHz and in principle a maximum of five CCs can be aggregated to achieve a maximum of 100 MHz, as can be seen in Fig. 1 [7–9]. For practical usages considering bandwidth availability for operators, various CA configurations are standardized in 3GPP Release 10 and beyond within intra-band and inter-band aggregation types [7–9]. In Frequency Division Duplexing (FDD) mode, aggregated CCs in downlink and uplink can be different and UL CCs ca be less than or equal to number of DL CCs [5, 8]. In case of Time Division Duplexing (TDD), number of CCs as well as the bandwidths of each CC shall normally be same [8].

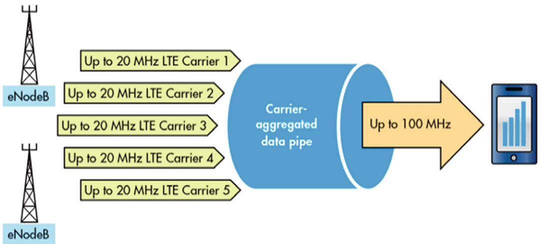


Fig. 1. Aggregation of LTE CCs to get higher bandwidth [11]

2.2 Enhanced MIMO (eMIMO)

To improve network spectral efficiency, one key technology of LTE-A is eMIMO that enables LTE-A to apply up to 8 $\times$ 8 MIMO in 3GPP Release 10 to enjoy more MIMO multiplexing, diversity and array gains compared to LTE Release 8 that uses up to 4 $\times$ 4 MIMO [6]. Of course, an achieved gain depends on radio environment and applied transmission mode [6]. Different transmission models are defined for LTE-A eMIMO in Release 10 (and beyond) that could be for single user MIMO, multi user MIMO and dynamic hybrid of the two.

2.3 Heterogonous Network (HetNet)

It has been well noted even before LTE-A that traditional macro cell only deployment is not effective to meet the increasing spatiotemporally heterogeneous data demand, that emanates mostly form indoor [10]. One important technology to meet the heterogeneous data demand is densification where low-power nodes (such as Pico cells, femto cells, relays) are deployed under umbrella macro network, resulting HetNet [27].

This brings considerable architectural change that affects network efficiencies in terms of capacity, coverage, energy, cost and other factors [10, 15, 16]. The low-power nodes can be planned and deployed to enhance data capacity, coverage, cost efficiency and energy efficiency while meeting the heterogeneous demand considering its spatiotemporal distribution.

Although densification has already started even before 3GPP LTE release 8, the LTE-A Release 10 and beyond incorporate efficient interference management techniques between cells. This enables LTE-A to enjoy much better spectral efficient densification.

2.4 Coordinated Multi Point (CoMP)

To address inconsistent quality of experience across mobile network challenge, LTE-A can use a technology called CoMP that mitigate inter-cell interference (main reason of inconsistent experience) by coordinating neighboring cells in various levels. CoMP can be performed among macro cells and between macro and small cells and both in uplink and downlink [6].

3 Scenario Planning for Addis Ababa

3.1 Scenario Planning Method

To formulate potential LTE-A scenarios for Addis Ababa, we adopt scenario-planning method that is described in [4, 12]. The method consists of four important steps: *Trend analysis*, *Uncertainty identification*, *Forming correlation matrix* and *Formulating scenario matrix*. Accordingly, we first perform trend analysis for LTE-A considering the technology, best experiences, expert opinions, vendor reports and Addis Ababa demands. The resulted trend is then compiled using PESTLE (*Political, Economic, Social, Technological, Legal and Environmental*) [25] and summarized as presented in Table 1.

Table 1. Summary of trends (PESTLE [25])

PESTLE analysis	Trends
Political (P)	Unutilized/free bands are available in 2G, 3G and LTE networks. Also acquiring new bands is possible
Economical (E)	Possible to deploy LTE-A on the existing mobile sites, mobile tariff reduction performed, less trend on CAPEX and OPEX cost vs. energy consumption analysis, high cost of LTE-A capable devices, mobile services takes high share from ethio telecom revenue
Social (S)	User awareness on new technologies, less effort on local content creation, digital growth helps to boost social capital, customization of phones
Technological (T)	Availability of LTE-A capable devices, mobile penetration rate increasing, underutilized backhaul and spectrum, indoor traffic, data traffic increasing
Legal (L)	Placement of macro and small sites
Environmental (En)	Alternative energy sources used for remote BSs, possible technological solutions to promote Green ICT

Then, from the key trends, several uncertainties including *deployment strategy* and *band usage of small cells* compared to macro cells are identified and their correlation is analyzed in the correlation matrix forming step, see [27] for details. Finally, we select the aforementioned two uncertainties and develop a scenario matrix that provides four potential LTE-A deployment scenarios for Addis Ababa. Other uncertainties are used for further definition of the selected scenarios.

3.2 Potential LTE-A Deployment Scenarios for Addis Ababa

The formulated LTE-A scenario matrix based on the two uncertainties is shown in Fig. 2. The obtained four potential scenarios are defined further based on other uncertainties including suitability of LTE-A features for local context, national spectrum status and small cell backhauling. The potential scenarios are briefly defined as follows for the Addis Ababa demography's, i.e. Dense Urban (DU), Urban (U) and Sub-Urban (SU):

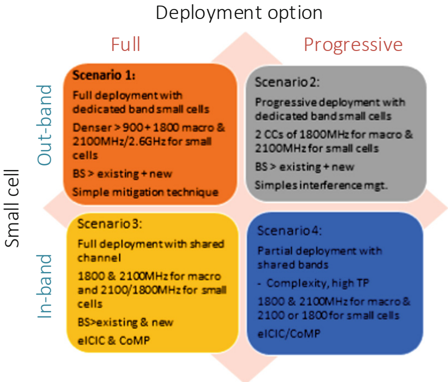


Fig. 2. LTE-A scenario matrix

SC1: This is a scenario that applies *Full deployment of LTE-A with Out-of-band small cells*. For the U areas non-contiguous CCs, intra-band CA, of *900* and *1800* MHz with 30 MHz aggregated bandwidth (BW_{agg}). For DU, contiguous CCs, i.e. inter-band CA, of *1800* MHz for macro cells allocated to get an BW_{agg} of 30 MHz. Regarding small cells, bands different from the macro cell are allocated. Based on these, *2100* MHz for *small scale* and 2.6 GHz *large scale* small cells assigned. For SU, *1800* MHz used for macro coverage and *2100* MHz for *hotspot* areas that helps to obtain enhanced experience for the UEs on the overlapping areas.

SC2: This is a scenario that applies *Progressive deployment of LTE-A with Out-of-band small cells* for DU and U areas. For these scenario contiguous CCs of *1800* MHz used for macro cells that can provide 30 MHz BW_{agg}. Whereas, CCs of *2100* MHz for *small scale* and/or *2.6 GHz* for *large scale* small cells used.

SC3: This is a scenario that applies *Full deployment of LTE-A with In-band small cells*. Similar bands will be used for the macro and small cells. For the possible interference, these scenario necessities interference management techniques such as eICIC and CoMP. Contiguous CCs of *1800* MHz for macro and small cells in DU areas whereas non-contiguous CCs from *1800* and *2100* MHz allocated for macro cells and *2100* MHz for small cells in U areas. In SU, *1800* MHz for macro coverage and other CCs of *1800* MHz for hot spot areas to provide increased bandwidth for users in the overlapping areas.

SC4: This is a scenario that applies Progressive deployments of LTE-A with In-band small cells for DU and U areas. Similar to SC3 the allocated bands shared between macro and small cells and requires interference management techniques. For these, CCs of *1800* MHz used for macro cell and available CCs of *1800* MHz for *small scale* and/or *2.6* GHz for *large scale* assigned for small cells.

Due to availability of significant unused bands in Addis Ababa even in the most popularly used *1800* MHz, *2100* MHz and *900* MHz bands, see Fig. 3, we select SC1 and SC2 for further techno-economic investigation.

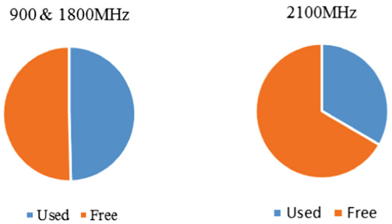


Fig. 3. Ethio telecom band usage [3]

For the selected LTE-A deployment scenarios, the network layout designated in Figs. 4 and 5 relying on the applied bands propagation characteristics.

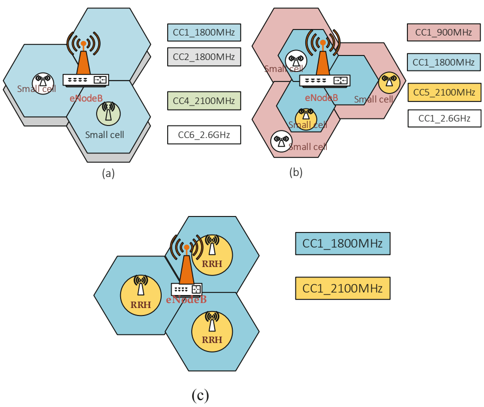


Fig. 4. Full LTE-A deployment with Out-of-band small cells: (a & b) DU/U and (c) SU

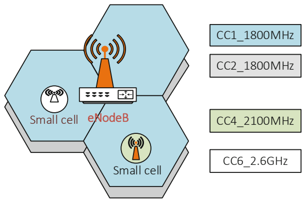


Fig. 5. Progressive LTE-A deployment with Out-of-band small cells

4 Techno-Economic Methodology and Assumptions

For the techno-economic analysis, we adopt techno-economic models from TERA framework that is described in [20, 21] and tune it to suit our study. The modification is to include important input parameters for the study and resulted modified TERA model is show in Fig. 6. The most essential parts of the TERA model are described briefly in the remainder of Sect. 4. In the tool market analysis, dimensioning, value of network elements and spectrum are taken as major parameters together with economic inputs, discussed in the following sub-section.

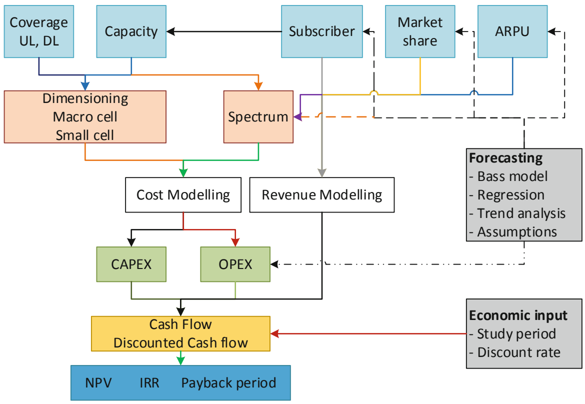


Fig. 6. Modified TERA model

4.1 Market Analysis and Economic Input

For the market analysis, data driven analysis, forecasting, reports and other relevant assumptions are considered [27]. LTE-A data plan/dimensioning performed taking the existing LTE monthly data plan and taking recommended minimum data requirements for intensive apps and services. Regarding LTE-A capable device penetration, we relied on reports taken from [3]. Candidate subscriber and Average Revenue per User (ARPU) estimated, taking ten years back data record from ethio telecom and forecasted for the study period. While market share of LTE-A are determined based on [27–30]. In addition, study period and discount rate defined for the economic input.

4.2 LTE-A Radio Network Dimensioning

LTE-A planning differs from other wireless network planning is mainly in the allocation of Physical Resource Block (PRB) and bandwidth gain because of CA and MIMO [15]. Different parameters are taken in to consideration for dimensioning process [3, 24, 27].

Macro Cell Dimensioning. Depending on adaptive modulation scheme, coding, available bandwidth and required data plan per user capacity driven eNodeBs determined [14]. Whereas, Allowed Propagation Loss (APL) produced using path loss to estimate the coverage radius of eNodeB regarding coverage dimensioning [14, 27]. Separate selected link budget calculations are performed for DU, U and SU for the selected scenarios. For this, excel based link budget tool used for coverage dimensioning while COST-231 Hata applied as a propagation model, Eq. (1), [14].

$$PL = -46.3 - 33.9Log(f) - 13.82Log(hBS) - a(hUE) - [44.9 - 6.55Log(hBS)]Log(d) - C_O \quad (1)$$

In (1), the *blue* and *red* colors represent the path loss constant and path loss exponent, respectively, and *a(hUE)* is UE antenna gain function and defined for urban (Eq. (2)) and sub-urban (Eq. (3)) areas as the following:

$$a(hUE) = 3.2(log 11.75hUE)^2 - 4.97, f \geq 400\,\text{MHz} \quad (2)$$

$$a(hUE) = 1.1(logf - 0.7)hUE - (1.56logf - 0.8) \quad (3)$$

whereby, f = carrier aggregation in MHz, *hBS* BS antenna height in meters, *hUE* UE antenna height in meters, d is the cell radius in km and the term *Co* (0 dB for SU and 3 dB for DU/U) are used to account for different trains [24]. The obtained number of eNodeBs in the LTE-A radio network dimensioning are summarized in Table 2.

Table 2. Summary of results of coverage and capacity dimensioning

Scenario	Capacity	Coverage
SC1	701	420
SC2	580	328

Small Cell Dimensioning. In this study, hot spot areas are identified and organized in clusters for indoor SCs based on the geographic area [27]. To select the area, existing LTE KPI report and building density of Addis Ababa are used as a key dimensioning parameter for indoor small cells [4, 27]. While for the outdoor SCs, beside LTE KPI report, building density, optimal placement of cells among concentrated traffic hotspots analyzed, Table 3 [17, 27, 3GPP TR 36.932].

Table 3. Impact of placement on effectiveness of metro cells [17]

MeBS distance from MaBS	% gain in median user throughput
200 m	2%
400 m	70%
700 m	467%

Based on the above points, assumptions and recommendations the small cell dimensioning result summarized in the following table for the selected scenarios, Table 4.

Table 4. SCs for full and progressive deployments

Scenario	Outdoor SC	Indoor SC
SC1	200	1240
SC2	100	709

Spectrum Dimensioning. The methodology adopted from ITU to predict future spectrum requirement for ethio telecom for the macro and small cells [27]. To check whether the existing spectrum support LTE-A or not, i.e. the amount of spectrum required, Eq. (4) [18, 19] is used taking the forecasted number of end-users for the study period, required bandwidth for the macro and small cells and other market analysis and dimensioning results [27].

$$BW_R = Cell_C(Mbps)/\eta\left(\frac{\frac{bps}{Hz}}{cell}\right) \tag{4}$$

Whereby, BW_R is required bandwidth, $Cell_C$ is cell capacity (calculated based on Eqs. (5) and (6)) and η is spectral efficiency, Eq. (7).

$$Cell_C = S * \left(8192\left(\frac{Mb}{GB}\right)\right) * \frac{D_P\left(\frac{GB}{\frac{sub}{month}}\right)}{\left(S_T * S_C * 3600\left(\frac{s}{hr}\right) * 30\left(\frac{day}{month}\right)\right)} * D \tag{5}$$

$$D = \frac{9\%(BH\ share)}{50\%(Maxlaod)} * \frac{50\%}{15\%} \tag{6}$$

Whereby, S is the subscriber assumed in scenario one and two, S_T is no. of sites, S_C is sector per BS, D_P data plan and D is the distribution. D is based on BH share, maximum load in which 50% of the traffic is carried by 15% of the cells.

$$\eta = P \cdot R_{av} \cdot \frac{G}{BW} \tag{7}$$

Whereby, η is spectral efficiency in bps/Hz/cell, P = 600 Mbps is peak data rate, R_{av} = 64QAM → 2/3 is average modulation and coding scheme, G = 2 is gain by MIMO and CA and BW = 40 MHz.

The required bandwidth for the deployment of LTE-A in either *full* or *progressive deployment* with out-of-band small cells predicted for the study period. The result show that SC1 requires 72.22 MHz while it is 72.21 MHz for SC2. Compared to the existing unutilized bandwidth, i.e. S_{UN} = 89 MHz is sufficient for the study period.

4.3 Modeling

After the required number of eNodeBs, indoor and outdoor small cells are determined from the dimensioning process, the related Capital Expenditure (CapEx) and Operational Expenditure (OpEx) cost driven based on [13, 22, 23, 26] to be used as an input for the *Cost* and *Revenue Modelling*. In [13], CapEx cost for deploying multicarrier BS for LTE-A with existing 4G LTE site and site with no LTE network is assumed. Furthermore, in [23] the total CapEx cost to deploy LTE-A, both for macro cell and small cells, is given. Similarly, in [13, 22], the operational cost related to Operations, Administration and Management (OA&M) given. In [23], OpEx cost for 3 CCs with erosion factor of 5% per year considered for upgrading and new site. For this study average value assumed for new and upgrading existing LTE site [27]. For outdoor and indoor small cells, 20 and 10% of the macro cell CapEx and OpEx cost assumed for the TEA based on respectively [23].

Cost Modeling. Costs can be generally split between CapEx and OpEx and broadly categorized as top-down and bottom-up approaches [20, 26]. In this study, a hybrid approach modelled [27]. Accordingly, the cost model is illustrated as in Eq. (8).

$$C_T = N_M C_M + N_S C_S + C_{Spectrum} + C_{BH} + C_{Core} + OPEX \tag{8}$$

Whereby, C_T is total cost, N_M and N_S are number of macro and small cells, C_M, C_S, $C_{Spectrum}$, C_{BH} and C_{core} are the CapEx cost related to macro, small cell, spectrum, backhaul and core network respectively and *OPEX* cost to run the network.

Revenue Modelling. The number of subscribers, ARPU and market share are the main inputs for calculating the revenues. Together with indirect revenues stemmed from, for example, roaming and advertising [20]. Equation (9) shows the revenue modelling used in the study.

$$R_T = ARPU * S_P * PM_S + I_R \tag{9}$$

Whereby, R_T is the generated revenue, S_P is predicted number of subscribers, PM_S is predicted market share of LTE-A and I_R is indirect revenue from for example roaming, sale of SIM.

4.4 TEA Evaluation

For the TEA evaluation, Cash Flow (CF), Discounted CF (DCF), NPV and PP are used.

CF and DCF. Cash flow is the net amount of cash that is received and disbursed during the study period. Based on cash and revenue model, the CF model is represented as

$$CF = \sum\nolimits_{i=1}^{6} (R_T - C_T)i \tag{10}$$

Whereby, CF is the cash flow, R_T and C_T revenue and cost models and i is the study period.

Whereas, DCF analysis, in Eq. (11), is the process of calculating the present value of an investment's future cash flows in order to arrive at a current fair value estimate for the investment i.e., time value of money [20].

$$DCF = \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \cdots + \frac{CF_i}{(1+r)^i} \tag{11}$$

Whereby, CF is cash flow, r is discount rate and i the study period.

NPV and PP. These parameters are used to evaluate the viability of the investment. NPV calculated, as indicated in Eq. (12) [20]. Whereas, PP calculated as in Eq. (13).

$$NPV = CAPEX - DCF = CAPEX - \sum\nolimits_{i=1}^{6} CF_i/(1 + r)^i \tag{12}$$

$$PP = LA + [\frac{Abs(CCF)}{TCF}] \tag{13}$$

Whereby, LA is last period of negative cumulative CF, CCF = value of cumulative CF at the end of LA, TCF = Total CF i.e. the period after LA.

5 Result and Discussion

5.1 Techno-Economic Evaluation Result

The existing 329 LTE BSs upgraded and new sites added relying on BS requirement, see [27]. In terms of spectrum, no cost for additional band is assumed since the unutilized bandwidth from the existing network is used. Then the output of techno-economic evaluation is performed for the two scenarios, as discussed below, using key TE evaluation parameters like NPV, CF, DCF and PP.

Full LTE-A Deployment with Out-band Small Cells: Figure 7 shows results of cash flow analysis for coverage and capacity for SC1. The output indicates positive NPV for DU, U and SU with payback period in the range of 3 to 5 years in terms of coverage,

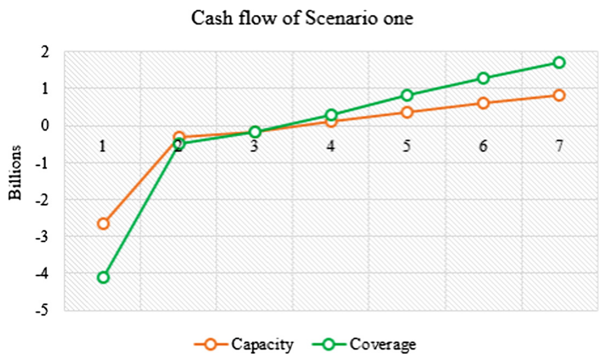


Fig. 7. Cash flow analysis for scenario one

Table 5. Whereas, in the case of capacity the result shows negative NPV with extended PP or more time period than the predefined study period.

Table 5. Summary of TEA for SC1

Deployment	Dimensioning	NPV	Payback period
DU and U	Coverage	1,162.47	4.13
SU		1,975.89	3.19
DU, U and SU		565.52	4.68
DU, U and SU	Capacity	(1,277.80)	-

Progressive LTE-A Deployment with Out-band Small Cells: Techno-economic evolution result for both coverage and capacity point of view shows positive NPV with payback period in the range of 3 to 5 years, Table 6. Cash flow for coverage and capacity analysis illustrated in Fig. 8.

Table 6. Summary of TEA for SC2

Deployment	Dimensioning	NPV	Payback period
DU and U	Coverage	2,290.55	3.29
DU and U	Capacity	644.02	4.70

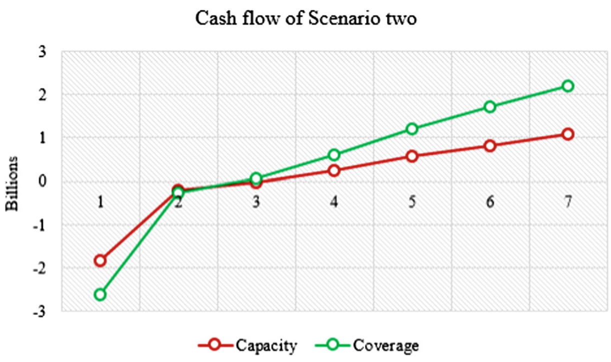


Fig. 8. Cash flow analysis of scenario two

5.2 Discussion

There is always a tradeoff between QoS and cost. In such scenario, efficient planning helps to properly manage resources and save unnecessary wastage. In scenario-planning approach presented in this study, the list of possible LTE-A deployment options are proposed and two of them identified as viable deployment scenarios and evaluated techno-economically. Subsequently, market analysis and dimensioning processes performed by applying the modified TERA framework and developed TEA tool.

In SC1, the NPV and PP indicated that LTE-A is feasible in DU, U and SU areas from coverage perspective with 4.13 and 3.19 PP for DU/U and SU respectively. However, SC1 is capacity limited and the NPV and PP results indicated that it is not feasible in the defined study period. While in SC2, the NPV and PP indicated that LTE-A is feasible in DU and U areas both in coverage and capacity perspective with PP 3.29 and 4.70 respectively.

In general, the TE evaluation result showed that progressive deployment of LTE-A is practical and feasible deployment option of LTE-A for Addis Ababa.

6 Conclusion

The study themed with describing the cellular technology evolution giving focus on LTE-A and the ever-growing end-user demand. The analysis indicates that the current infrastructure cannot cope with demand and requirement of an efficient network. Employing systematic literature reviews, standards, reports, expert opinions and state-of-the-art surveys, a scenario is planned to pinpoint realistic LTE-A deployment options for Addis Ababa. For these PESTLE analysis method is used to study the possible trends and uncertainties for the local context. Positioning on the result macro, small cell and spectrum dimensioning performed to drive needful NE for CapEx and OpEx analysis. In the study, the status and applicability of existing resources, such as spectrum, base station and backhaul for LTE-A are identified. It is then concluded that the backhaul has the capacity to handle LTE-A while LTE base stations can be upgraded and new sites added depending on the scenarios. Concerning spectrum, a six year demand forecasted and the possibility of reusing the unutilized bandwidth verified.

Furthermore, a suitable TEA framework was identified to evaluate the selected deployment scenarios techno-economically. The framework is modified to enclose parameters that will reduce cost and escalate revenue. Based on the modified framework, the results indicated that SC1 is not feasible within the defined study period while SC2 is a viable LTE-A deployment option for Addis Ababa. In general, it is an inevitable that mobile data demand and traffic volumes will continue to increase in the coming years in Ethiopia, specifically in Addis Ababa. Therefore, a technology that can handle this demand is required. It is then important to identify proper planning and scientific analysis method and evaluate technology viability in technical and economic measures.

References

1. Cox, C.: An Introduction to LTE: LTE, LTE-A, SAE, VoLTE and 4G Mobile Communications, 2nd edn. Wiley, Hoboken (2014)
2. Holma, H., Toskala, A.: LTE for UMTS, Evolution to LTE-Advanced, 2nd edn. Wiley, Hoboken (2010)
3. Ethio Telecom: Ethio Telecom company profile, documents and reports 2013–2018
4. Haile, B.B., et al.: Development of the Wi-Fi offloading business concept within the African Market Context. Aalto University, Finland, October 2012
5. 3GPP TS 36.300: Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; Stage 2, Release 15, 3GPP, V15.2.0, June 2018
6. Khasdev, B., Hirwe, A.: Overview of MIMO technology in LTE, LTE-A & LTE-A-Pro. Int. J. Eng. Comput. Sci. **5**(11), 19227–19229 (2016)
7. Kiwoli, L., et al.: Performance analysis of carrier aggregation for various mobile network implementation scenarios based on spectrum allocated. IJWMN **9**(5) (2017)
8. Wannstrom, J.: Carrier aggregation explained. www.3gpp.org/technologies. Accessed 14 Jan 2019
9. 4G Americas: LTE carrier aggregation technology development and deployment world wide, October 2014
10. Sánchez, S.B., Pocovi Gerardino, G.A.: LTE-advanced HetNet investigations under realistic conditions. MSc thesis, Aalborg University, Denmark, May 2014
11. Electronic Design: Carrier aggregation: implications for mobile-device RF front-ends, 19 February 2016
12. Schoemaker, P.J.H.: Scenario planning: a tool for strategic thinking. Sloan Manag. Rev. **36**, 25–40 (1995)
13. Duan, L.: Economic analysis of 4G upgrade timing. IEEE Trans. Mob. Comput. **14**, 975–989 (2015)
14. ElNashar, A.: Design, Deployment and Performance of 4G Networks: A Practical Approach. Wiley, Hoboken (2014)
15. Acharya, J.: Heterogeneous Networks in LTE-Advanced. Wiley, Hoboken (2014)
16. Ek, P.: Deployment of indoor small cells for 4G mobile broad band. MSc thesis, October 2013
17. Buddiket, M., Soni, R.: Small cell technology overview, Alcatel, 13 March 2013

18. Riaz, A.: Spectrum forecasting for future use: methods and techniques. ITU Regional Workshop, Bangkok, Thailand, 3–4 May 2017
19. Chung, W.G.: Calculation of spectral efficiency for estimating spectrum requirements of IMT-advanced in Korean mobile communication environment. ETRI J. **29**, 153–161 (2007)
20. Smura, T.: Techno-economic modelling of wireless network and industry architecture. Ph.D. dissertation, Aalto University, Finland (2012)
21. METIS II: Quantitative techno-economic feasibility assessment. Deliverable D1.2, 15 June 2017
22. Nikolikj, V., Janevski, T.: Applicable cost modeling of LTE-advanced and IEEE 802.11ac based heterogeneous wireless access networks. In: 10th Advanced International Conference on Telecommunications (2014)
23. Oughton, E.J., Frias, Z.: Exploring the cost, coverage and rollout implications of 5G in Britain. MSc thesis, Madrid, Spain (2016)
24. Skinner, T.: 30% of operators investing in LTE-A, VoLTE rollouts grow, 23 July 2015 report. http://telecoms.com. Accessed 14 Jan 2019
25. Stork, C.: Strategic analysis of the telecommunication sector. Research ICT Africa, 9 August 2012
26. Zareen, U.: LTE Advanced: techno economic perspective. Univ. J. Commun. Netw. **2**, 93–100 (2014)
27. Hailegebriel, G.: Techno-economic investigation of LTE-advanced deployment for Addis Ababa, Ethiopia. MSc thesis. Addis Ababa University, AAiT, November 2018
28. Heuveldop, N.: Ericsson mobility report. Ericson, June 2017
29. Ericsson mobility report-interim update, Sweden, February 2018
30. Haile, B.B., Aga, D., Mulu, B.: On the relevance of capacity enhancing 5G technologies for Ethiopia. In: 10th Ethiopian ICT Annual Conference, Addis Ababa, Ethiopia, June 2017

Peak-to-Average Power Ratio Reduction of OFDM Signals Using Selected Mapping

Workineh Gebeye Abera(✉) and Yihenew Wondie Marye(✉)

School of Electrical and Computer Engineering,
Addis Ababa Institute of Technology, Addis Ababa University,
Addis Ababa, Ethiopia
Workinehgebeye@gmail.com, yiheneww@gmail.com
{workineh.gebeye,yihenew.wondie}@aait.edu.et

Abstract. Orthogonal Frequency Division Multiplexing (OFDM) signaling is highly applicable in wireless communications, like 4G cellular networks, as it has better performance in a fading environment. However, OFDM signal has large Peak-to-Average Power Ratio (PAPR) which leads to signal distortion in the electronic components of the OFDM system transmitter. As a result, PAPR reduction techniques are used in OFDM systems. Selected Mapping (SLM) scheme is one method of reducing PAPR of OFDM signals. In this paper, we have demonstrated the PAPR reduction capability of SLM through simulation on MATLAB. In addition, we have shown the computational complexity incurred in using SLM to reduce PAPR. It is shown that the PAPR reduction capability of OFDM increases with increase in the number of candidate OFDM symbols. However, this is achieved at the cost of increased computational complexity, where the complexity also linearly with the number of candidate OFDM symbols.

Keywords: Orthogonal Frequency Division Multiplexing · Peak-to-Average Power Ratio · Selected Mapping

1 Introduction

OFDM is a multicarrier signaling scheme which subdivides an available frequency-selective wideband channel into a number of narrowband subchannels (subcarriers) which are non-frequency-selective. Hence, OFDM signaling has high immunity to frequency selective fading. The subcarriers are overlapping but orthogonal, which avoids the need to separate the carriers by means of guard-bands, and therefore makes OFDM highly spectrally efficient. Furthermore, the arrangement of the subcarriers makes receiver implementation simple as the subcarriers can be perfectly separated. For its good properties, OFDM, in its multiple access version, OFDMA, has been adopted for downlink transmission in long term evolution-advanced (LTE-A) network [1, 2]. In addition, it is used in wireless systems, which use multimedia service applications, like, digital audio broadcasting systems, digital video broadcasting terrestrial TV (DVB-T) systems, asymmetric digital subscriber line (ADSL) systems, IEEE 802.11a/g/n wireless local area network (WLAN), IEEE 802.16 worldwide interoperability for

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 203–212, 2019.
https://doi.org/10.1007/978-3-030-26630-1_17

microwave access (WiMAX) systems, and ultra-wideband (UWB) multiband-OFDM systems [3].

However, the required features of OFDM are obtained at the expense of large envelope variation, which is a major drawback of OFDM and is usually quantified through the Peak-to-Average Power Ratio (PAPR). Such signal envelope variations can be difficult for practical High Power Amplifiers (HPAs) and Digital to Analog Convertors (DACs)/Analog to Digital Converters (ADCs) of the OFDM system to accommodate, resulting in either low power efficiency or signal distortion, including signal clips [4]. Consequently, the signal distortion results in Bit Error Rate (BER) increase and Power Spectral Density (PSD) degradation. These effects can be avoided by making the HPA to work in its linear region with large back-off however, this results in poor power efficiency. Similarly, the DAC/ADC can be designed to accommodate the large dynamic range of the OFDM signal but, this results in a reduced Signal to Noise Ratio (SNR), as the DAC/ADC already has significant amount of quantization noise. Hence, a better solution is to reduce the PAPR of the OFDM signal with some manipulation of the OFDM signal itself [4].

As a result, a number of researches have been done on the development of PAPR reduction schemes for OFDM signals. The first class of schemes reduce PAPR with signal distortion, which include amplitude clipping [5], clipping and filtering [6]; but the schemes have in-band distortion and out-of-band spectral regrowth limitations. Distortionless schemes such as coding [7], Selected Mapping (SLM) [8], Partial Transmit Sequence (PTS) [9], etc. are the second class of PAPR reduction schemes. An overview of the various PAPR reduction schemes can also be found in [4].

In this paper, we will show SLM's, one of the distortionless PAPR reduction schemes, PAPR reduction capability and the associated computational cost incurred in using the scheme. In SLM, for every OFDM symbol intended for transmission, D candidate OFDM symbols are generated (by rotating the phase of every data symbol within an OFDM symbol with D distinct phase rotation vectors $\mathbf{P}^{(\mathrm{d})}$) by the transmitter, among which the one with the least PAPR will be selected for transmission. To recover the original OFDM symbol, the receiver has to know the phase vector $\mathbf{P}^{\bar{\mathrm{d}}}$ (or the index $\bar{\mathrm{d}}$) that gives the least PAPR OFDM symbol. To this end, side information (SI) (with $\log_2 D$ bits) has to be transmitted along with the OFDM symbol, which reduces the information throughput.

Hence, the Author in [8] has proposed a novel technique that helps to recover the transmitted OFDM symbol from the received OFDM symbol itself; i.e. blindly. In the technique, the parameter $\bar{\mathrm{d}}$ is made to be linked with a set of distinct amplitude functions. The Author has shown through simulation that the proposed method gives the same PAPR reduction performance as the original SLM (with SI) method.

However, the paper would have been more comprehensive, had it shown the PAPR reduction performance of SLM as the number of candidate OFDM symbols increase. In addition, the paper doesn't show the computational cost incurred in using SLM to reduce PAPR.

As a result, in our paper we have shown the PAPR reduction performance of the scheme with the number of candidate OFDM symbols. We have also analyzed the computational complexity of the scheme. Furthermore, we have shown the optimum condition for the choice of phase rotation factors so that the best PAPR reduction capability of SLM can be enjoyed.

The remaining part of the paper is organized as follows. Section 2 discusses about the OFDM signal model used in the paper. The Scheme's system model is presented in Sect. 3. The computational complexity of the scheme is analyzed in Sect. 4. Section 5 presents simulation results and discussions. Finally, the concluding remarks are made in Sect. 6.

2 OFDM Signal Model

For a given data vector $\mathbf{S}^{\mu} = [S_0^{\mu} S_1^{\mu} \cdots S_{N-1}^{\mu}]^T$, a sequence of complex numbers drawn from a finite constellation (MPSK or MQAM) in the μ^{th} signaling interval, the baseband OFDM symbol $\left\{s_{n/L}^{\mu}\right\}_{n=0}^{LN-1}$ is an oversampled IFFT output of $\mathbf{S}^{\mu}$. That is [10, 11]

$$s_{n/L}^{\mu} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k^{\mu} e^{j\frac{2\pi}{LN}kn}, n = 0, 1, \cdots, LN - 1$$

$$s_{n/L}^{\mu} = \frac{1}{\sqrt{N}} \left\{ \sum_{k=0}^{N/2-1} S_k^{\mu} e^{j\frac{2\pi}{LN}kn} + \sum_{k=LN-\frac{N}{2}}^{LN-1} S_{k-N(L-1)}^{\mu} e^{j\frac{2\pi}{LN}kn} \right\}$$

$$s_{n/L}^{\mu} = IFFT\left\{ \sqrt{L} S_L^{\mu} \right\} \tag{1}$$

where N is the number of subcarriers, $L \geq 1$ is an integer and it is the oversampling factor, $IFFT\{\cdot\}$ is NL-point oversampled IFFT indexed by n/L, and $\mathbf{S}_L^{\mu} = \left[S_0^{\mu} \cdots S_{\frac{N}{2}-1}^{\mu} 0 \cdots 0 \, S_{\frac{N}{2}}^{\mu} \cdots S_{N-1}^{\mu} \right]^T$ is the L times oversampled equivalent data vector generated by zero padding $\mathbf{S}^{\mu}$ with $N(L-1)$ zeros at its middle [10, 11].

The baseband PAPR is defined as [10, 11]

$$PAPR\left\{s_{n/L}^{\mu}\right\} = \frac{\max_{n \in [0, LN)} \left| s_{n/L}^{\mu} \right|^2}{E\left\{ \left| s_{n/L}^{\mu} \right|^2 \right\}} \tag{2}$$

which is a random variable.

As the passband PAPR is roughly twice (3 dB higher than) the baseband PAPR, it is sufficient to consider only the PAPR of the baseband OFDM signal [10], pp. 22–23. In addition, the cyclic prefix of duration T_g, which is a repetition of part of the OFDM symbol, attached to the OFDM symbol to combat Inter-Symbol Interference (ISI) can be neglected for the purposes of PAPR analysis as the prefix will not produce a peak which is not already present in the OFDM symbol $s^{\mu}_{n/L}$. Theoretically, $\mathrm{PAPR}\left\{\mathrm{s}^{\mu}_{\mathrm{n/L}}\right\}$ approaches $\mathrm{PAPR}\{\mathrm{s}^{\mu}(\mathrm{t})\}$ as L becomes sufficiently large. However, it is has been shown in [11, 12] that when $L \geq 4$ the PAPR of $s^{\mu}_{n/L}$ approximates the PAPR of $s^{\mu}(t)$ and hence $L = 4$ is used in the simulation results of this paper.

3 Selected Mapping System Model

SLM is a distortionless PAPR reduction scheme [8, 13]. In the SLM scheme, the transmitter generates a set of sufficiently different candidate data vectors, all representing the same information as the original data vector, for each data vector intended for transmission by rotating the phase of each data symbol and then selects the candidate data vector with the lowest PAPR for transmission [8, 13].

A set of D markedly different, distinct, pseudorandom but fixed phase rotation vectors [11, 13]

$$\mathbf{P}^{(\mathrm{d})} = [\mathrm{P}^{(\mathrm{d})}_0 \mathrm{P}^{(\mathrm{d})}_1 \cdots \mathrm{P}^{(\mathrm{d})}_{\mathrm{N}-1}]^{\mathrm{T}}, \tag{3}$$

with $\mathrm{P}^{(\mathrm{d})}_{\mathrm{k}} = e^{j\phi^{(d)}_k}, \phi^{(d)}_k \in [0, 2\pi), \mathrm{k} = 0, 1, \cdots, \mathrm{N}-1, \mathrm{d} = 1, 2, \cdots, \mathrm{D}$ must be defined and available both at the transmitter and receiver. The data vector $\mathbf{S}^{\mu}$ is multiplied element-wise with each one of the D phase rotation vectors $\mathbf{P}^{(\mathrm{d})}$, resulting in a set of D different phase rotated data vectors $\mathbf{S}^{(\mu,\mathrm{d})}$ given by [11, 13]

$$\boldsymbol{S}^{(\mu,d)} = \boldsymbol{S}^{\mu} \circ \mathbf{P}^{(\mathrm{d})}. \tag{4}$$

Then, all the D data vectors are transformed into time domain to get D candidate OFDM symbols by using D length-LN IFFTs [11, 13]

$$s^{(\mu,d)}_{n/L} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S^{\mu}_k P^{(d)}_k e^{j\frac{2\pi}{LN}kn},\ n = 0, 1, \cdots, LN-1$$

$$s^{(\mu,d)}_{n/L} = IFFT\left\{\sqrt{L}\boldsymbol{S}^{(\mu,d)}_L\right\} \tag{5}$$

Among the D candidate OFDM symbols, the transmitter selects one, $s^{(\mu,\bar{d})}_{n/L}$, with the lowest PAPR, for transmission where [11, 13]

$$\bar{d} = arg \min_{1 \leq d \leq D} PAPR\left\{s^{(\mu,d)}_{n/L}\right\} \tag{6}$$

Therefore, the SLM-OFDM transmitter is as shown in Fig. 1.

In our work, it is expected that the transmitter and the receiver have the D phase rotation vectors $\mathbf{P}^{(d)}$. But, so as to recover an OFDM symbol, the receiver has to know the phase rotation vector $\mathbf{P}^{(\bar{d})}$ which has been used by the transmitter in generating the least PAPR transmit OFDM symbol. The easiest approach can be to transmit $\bar{d}$ as side information (SI), which requires $\log_2 D$. As SI transmission decreases the information throughput, another scheme to determine $\bar{d}$ is the blind technique [8], where it is determined based only on the received OFDM symbol and the known phase rotation vectors.

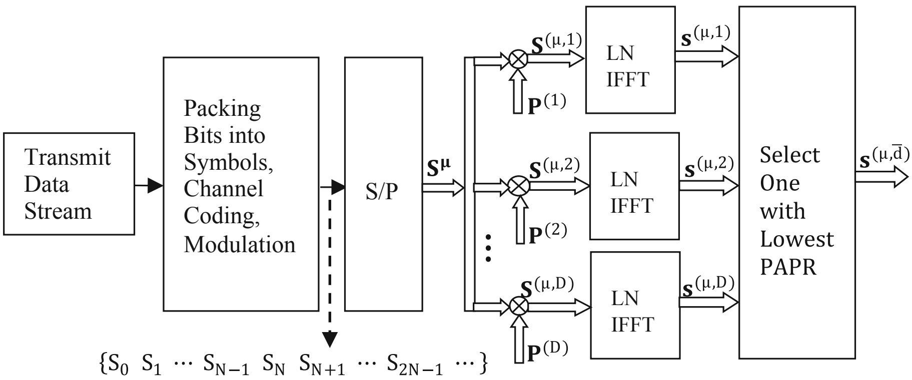


Fig. 1. SLM-OFDM transmitter block diagram.

Then, the original data vector $\mathbf{S}^{\mu}$ is recovered by multiplying the received data vector $\tilde{S}^{\mu}$ element-wise by $\mathbf{P}^{(\bar{d})} = \left[e^{-j\phi_0^{(\bar{d})}} e^{-j\phi_1^{(\bar{d})}} \cdots e^{-j\phi_{N-1}^{(\bar{d})}}\right]$. In the paper, we have assumed that the blind technique is used.

We have shown through simulation in Fig. 3 that for a given D, choosing $P_k^{(d)}$ such that the corresponding $\phi_k^{(d)}$ are uniformly distributed in $[0, 2\pi)$, then the PAPR reduction capability of SLM is the same whether $P_k^{(d)}$ are chosen from set $\{\pm 1\}$ or $\{\pm 1, \pm j\}$.

4 SLM Scheme Computational Complexity Analysis

As it is shown in (4), ND complex multiplications are required to generate $S^{(\mu,d)}, d = 1, 2, \ldots, D$. Then, D length LN IFFTs are required to generate $s_{n/L}^{(\mu,d)}$. Each LN-point IFFT needs $(LN/2)\log_2(LN)$ complex multiplications and $LN\log_2(LN)$ complex additions [14, 15]. Finally, $\left|s_{n/L}^{\mu}\right|^2 = Re\left\{s_{n/L}^{\mu}\right\}^2 + Im\left\{s_{n/L}^{\mu}\right\}^2$ need to be calculated for each n in determining the PAPR, which requires 2DLN real multiplications and DLN additions.

Four real multiplications and two real additions are required for a complex multiplication. On the other hand, a complex addition requires two real additions. Therefore, the total number of real additions A and multiplications M required for the scheme can be summarized as

$$A_{SLM} = DLN(3\log_2(LN) + 1) + 2ND \quad (7)$$

$$M_{SLM} = 2DLN(\log_2(LN) + 1) + 4ND \quad (8)$$

To quantify the computational complexity of SLM, a parameter f, which is the number of addition instructions required for each multiplication operation, is used. Thus, the overall computational complexity of the scheme for $f = 4$ is

$$C_{SLM} = A_{SLM} + fM_{SLM}$$

$$C_{SLM} = [11LN\log_2(LN) + 9N(L + 2)]D \quad (9)$$

The receiver of the scheme uses one FFT, which is the same as the OFDM system without any PAPR reduction scheme, to detect the transmitted OFDM symbol from the received OFDM symbol. Hence, the additional system complexity in using the PAPR reduction scheme is found in the transmitter.

5 Simulation Results

This section presents the simulation results that show the PAPR reduction capability of SLM and the computational complexity of the scheme. The simulations are done using MATLAB.

For every Complementary Cumulative Distribution Function (CCDF) curve, 10^5 OFDM symbols, each containing $N = 256$ data symbols from QPSK modulation constellation, are randomly generated. The oversampling factor used in the simulations is $\mathrm{L} = 4$. The elements of every phase rotation vector $\mathbf{P}^{(\mathrm{d})}, \mathrm{d} = 1, 2, \cdots, \mathrm{D}$, are chosen from the set $\{\pm 1\}$ with equal probability or from the set $\{\pm 1, \pm \mathrm{j}\}$ with equal probability and hence the candidate OFDM symbols are mutually independent.

Figure 2 shows the CCDF of PAPR of SLM-OFDM signal for different number of candidate OFDM symbols **D** along with the CCDF of PAPR of OFDM signal. The elements of every phase rotation vector are chosen from the set $\{\pm 1, \pm j\}$ with equal probability. In fact, the phase rotation vectors are obtained from a randomly generated binary data mapped onto QPSK symbols and hence the number of bits per symbol $\mathrm{m} = 2$ and the number of constellation points $\mathrm{Q} = 2^{\mathrm{m}} = 4$. Thus, the number of elements of the phase rotation factor set is $\mathrm{Q} = 4$.

The plot shows that for $\mathrm{D} = 1$, the CCDF curve overlaps with the CCDF curve of the OFDM signal. For $\mathrm{D} = 1$, the candidate OFDM symbol is the original OFDM symbol itself and hence the plot illustrates that there is no PAPR reduction. Whereas, for $\mathrm{D} = 2$ and above, the PAPR reduces significantly. As an example, for the original OFDM signal the probability that the PAPR is greater than around 10.2 dB is 10^{-2}, i.e.

one in 100 OFDM symbols will have a PAPR greater than 10.2 dB. But, for the SLM-OFDM signal with $D = 2$, to get the same CCDF value, 10^{-2}, the PAPR reduces to around 9 dB. This shows a reduction of around 1.2 dB. For the same CCDF value, the PAPR reduction is even much greater than 1.2 dB as D increases. In summary, the plot illustrates that the PAPR reduction capability of SLM increases with the number of candidate OFDM symbols D. But, this is at the cost of increased computational complexity as shown in Fig. 4.

Figure 3. shows the CCDF of PAPR of SLM-OFDM and OFDM signals. For each D, two CCDF curves are shown where one of the curves is for an SLM-OFDM signal with the elements of every $\mathbf{P}^{(d)}$ chosen from the set $\{\pm 1\}$ with equal probability and the other one is for the same SLM-OFDM signal but the elements of every $\mathbf{P}^{(d)}$ chosen from the set $\{\pm 1, \pm j\}$ with equal probability. The simulation is done to illustrate the PAPR reduction capability of SLM with these two phase rotation factor sets.

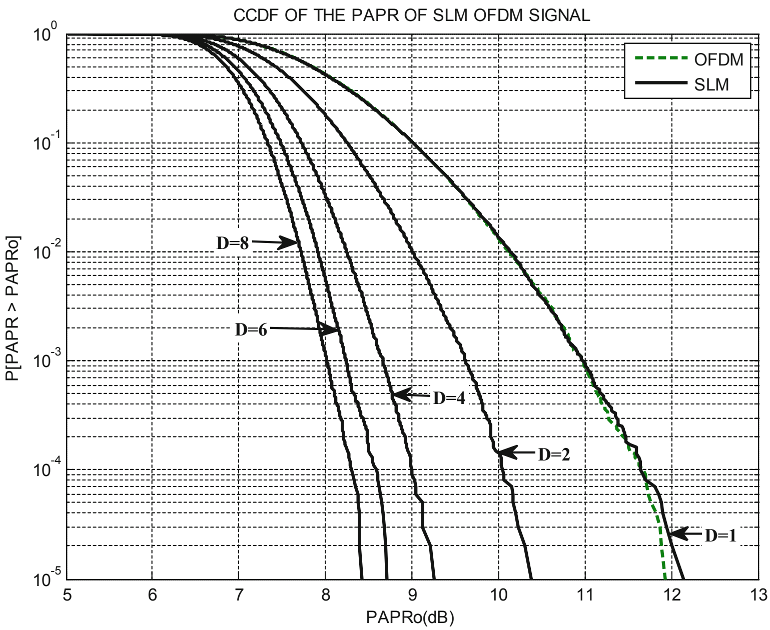


Fig. 2. CCDF of PAPR of SLM and OFDM signals for the elements of $\mathbf{P}^{(d)}$ chosen from the set $\{\pm 1, \pm j\}$.

The phase rotation vectors with elements from the set $\{\pm 1\}$ are actually obtained from a randomly generated binary data mapped onto BPSK symbols, i.e. $m = 1$ and hence $Q = 2^m = 2$.

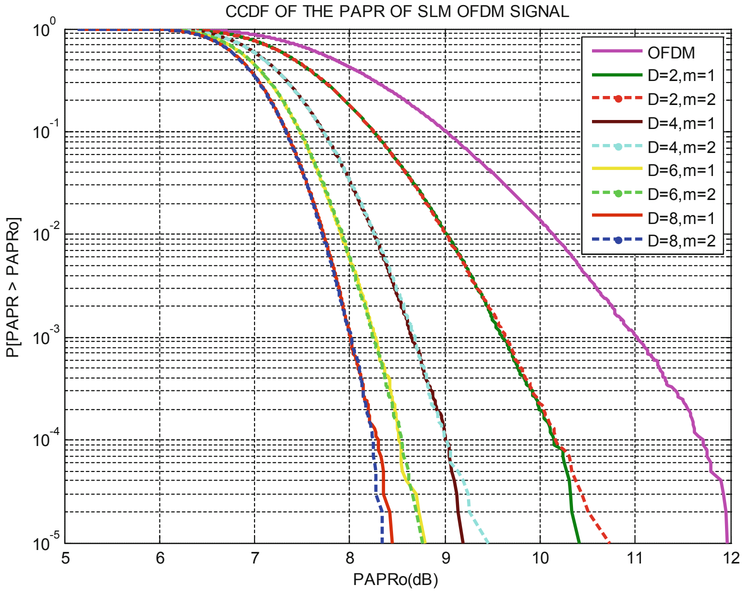


Fig. 3. CCDF of PAPR of SLM and OFDM signals for $P_k^{(d)}$ chosen from $\{\pm1\}$ and $\{\pm1, \pm j\}$.

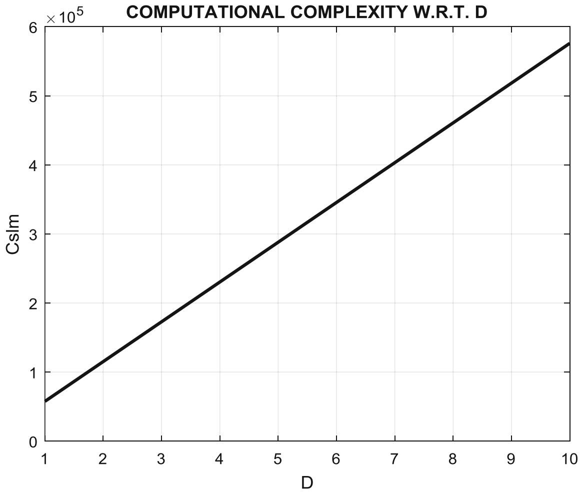


Fig. 4. Computational complexity of SLM with respect to (w.r.t.) D.

The plot illustrates that for a given number of candidate OFDM symbols D, the PAPR reduction capability of SLM is the same whether the elements of the phase rotation vectors are chosen from the set $\{\pm1\}$ or $\{\pm1, \pm j\}$. Hence, irrespective of the phase rotation factor set from which we choose the elements of the phase rotation vectors, the PAPR reduction capability of SLM is the same, if the corresponding phase rotation angles $\phi_k^{(d)}$ are uniformly distributed in $[0, 2\pi)$.

We have shown above that the PAPR reduction capability of SLM increases with increase in D, number of candidate OFDM symbols generated per OFDM symbol for transmission. The plot in Fig. 4 illustrates the computational cost incurred with increase in D. The plot shows that the computational complexity increases linearly with D. As a result, a compromise has to be made between the required PAPR reduction level and the accompanying increase in computational cost.

6 Conclusion

Though OFDM signaling is preferable in a multipath fading environment for its better performance, its large PAPR leads to signal distortion and hence BER degradation in the electronic devices like ADC, DAC, & high power amplifiers. Hence, schemes that can reduce the PAPR of OFDM signal have been introduced.

In this paper, we have shown that the PAPR reduction capability of SLM scheme. In the scheme, for a given OFDM symbol intended for transmission, D length LN candidate OFDM symbols are generated and then the one with the least PAPR is chosen for transmission. The D candidate OFDM symbols are generated using D parallel LN-point IFFTs.

Our simulation results show that the PAPR reduction capability of SLM is the same irrespective of the number of elements of the phase rotation factor set, as far as the phase rotation angles are uniformly distributed in [0, 2π).

The simulation results also show that the SLM scheme can reduce the PAPR of OFDM signal to a required level. The PAPR reduction capability of SLM increases with increase in the number of candidate OFDM symbols D, but at the cost of linear increase in computational complexity with D. Hence, we can make a compromise between the required PAPR reduction level and the associated increase in computational complexity.

References

1. Mahmoud, E., Abdelmohsen, A., Walaa, H.: Cellular LTE-A technologies for the future Internet-of-Things: physical layer features and challenges. IEEE Commun. Surv. Tutorials **19**(4), 2544–2572 (2017)
2. Haitham, S.B.A., Warip, M.N.M., Ong, B.L., Naimah, Y.: An efficient scheduling scheme for heterogeneous services in OFDMA based 5G LTE-Advanced network with carrier aggregation, pp. 275–280. IEEE (2018)
3. Hwang, T., Yang, C., Wu, G., Li, S., Li, Y.: OFDM and its wireless applications: a survey. IEEE Trans. Veh. Technol. **58**(4), 1673–1694 (2009)
4. Jiang, T., Wu, Y.: An overview: peak-to-average power ratio reduction techniques for OFDM signals. IEEE Trans. Broadcast. **54**(2), 257–268 (2008)

5. O'Neil, R., Lopes, L.B.: Envelope variations and spectral splatter in clipped multicarrier signals. In: The 6th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol. 1, pp. 71–75 (1995)
6. Armstrong, J.: Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering. Electron. Lett. **38**(5), 246–247 (2002)
7. Paterson, K.G., Paterson, V.: On the existence and construction of good codes with low peak-to-average power ratios. IEEE Trans. Inf. Theory **46**(6), 1974–1987 (2000)
8. Tong, G.: Peak-to-average power ratio reduction in OFDM: blind selected mapping for PSK inputs. In: IEEE ICASSP, pp. 293–296 (2006)
9. Reisi, N., Attari, M.A., Navid, R.: Embedded SI partial transmit sequence with improved bit error rate. In: IEEE 2nd International Conference on Electrical Engineering, March 2008
10. Jose, T.: Multicarrier Modulation with Low PAR: Applications to DSL and Wireless. Kluwer Academic Publishers, Dordrecht (2002). Norwell
11. Abera, W.G.: Comparative study of the performances of peak-to-average power ratio (PAPR) reduction techniques for orthogonal frequency division multiplexing (OFDM) signals. Master of Science thesis, Addis Ababa University, Faculty of Technology (2010)
12. Ochiai, H., Imai, H.: On the distribution of the peak-to-average power ratio in OFDM signals. IEEE Trans. Commun. **49**, 282–289 (2001)
13. Bäuml, R.W., Fisher, R.F.H., Huber, J.B.: Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping. IEEE Electron. Lett. **32**(22), 2056–2057 (1996)
14. Kang, S.G., Kim, J.G., Joo, E.K.: A novel subblock partition scheme for partial transmit sequence OFDM. IEEE Trans. Broadcast. **45**(3), 333–338 (1999)
15. Cochran, W.T., et al.: What is the fast Fourier transform? IEEE Trans. Audio Electroacoust. **AU-15**(2), 45–55 (1967)

Techno-Economic Analysis of LTE Deployment Scenarios for Emerging City in Africa: A Case of Adama, Ethiopia

Dechasa Negash Fekadu[1](✉) and Beneyam Berehanu Haile[2](✉)

[1] Ethio Telecom, Addis Ababa, Ethiopia
dechasa.negash@ethiotelecom.et
[2] Addis Ababa Institute of Technology, Addis Ababa, Ethiopia
beneyamb.haile@aait.edu.et

Abstract. The exponential growth of mobile broadband demand in emerging cities has been pushing deployment of long-term evolution (LTE) mobile technology. LTE can have various deployment scenarios in terms of various factors including enabled features, bandwidth usage, transmission frequency, requirements and other factors. To plan and deploy comprehensively effective LTE network for a given city, undertaking techno-economic analysis (TEA) is a very important task for mobile operators. Various literature also presents TEA for LTE deployment for different network requirements and deployment environment. Yet, such LTE study is rare for African emerging cities and it has not been done for Ethiopian ones. In this paper, we first formulate potential LTE deployment scenarios for one of the emerging cities in Ethiopia called Adama. Then to thoroughly understand viability of the formulated deployment scenarios, we present TEA assuming 7 years study period and continuation of current monopoly telecom market in the city. For the analysis, we apply modified TERA model that is implemented in MATLAB. The payback period results show that deployment of LTE in the emerging city using 1800 MHz band under high and low capacity demand and using 2100 MHz band under only high capacity demand are economically viable as they have payback periods less than 3.5 years.

Keywords: LTE · Radio access network · Scenario planning · TERA · Techno-economic analysis · Adama · Ethiopia · Africa

1 Introduction

Mobile networks of most operators have been witnessing an unprecedented rise in data traffic globally [1]. This has been happening due to an increase in mobile user demand to access bandwidth-intensive content and increasing penetration of capable mobile devices such as smartphones and tablets. This data demand rise trend has been challenging mobile operators and their networks. Similar data demand growth trend has also been seen in Africa including Ethiopia [2].

To continuously meet the increasing data demand with satisfactory user experience, operators have been expanding their network infrastructure to include contemporary

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 213–226, 2019.
https://doi.org/10.1007/978-3-030-26630-1_18

mobile technology including long term evolution (LTE). LTE represents an important evolutionary step in the development of mobile technology while providing significant improvement in spectral efficiency, cost-per-bit, user data rate and an overall network capacity [3]. This is proved in its penetration rate and as of August 2018, there are around 712 and more than 102 commercial LTE networks globally and in Africa, respectively [4]. Most of the LTE networks in Africa are deployed in capital cities and in the coming years it is expected emerging cities of Africa other than the capitals get LTE services, following the increasing data demand in those cities. For instance, Ethiopia already has LTE in its capital Addis Ababa but not in other emerging cities including Adama, Bahir Dar, Mekelle and Hawassa.

Deployment of mobile technologies in Africa is commonly vendor driven and not supported by localized techno-economic analysis (TEA) that is very important to make viable decisions both technically and economically, very key for developing countries [5, 6, 12]. Yet, it is seldom to find techno-economic study of mobile technologies in African context. We believe related study for Africa is exemplary and important. Therefore, in this paper, we present TEA for potential LTE deployment scenarios for one of Ethiopian emerging cities, Adama using modified TERA model that is implemented in Matlab. The potential LTE deployment scenarios are also formulated using a scenario planning method that is presented in [8]. The TEA is undertaken considering 7 years study period and forecasting Adama's future mobile network environment. The TEA results show that deployment of LTE in Adama using 1800 MHz band under high and low capacity demand and using 2100 MHz band under only high capacity demand are economically viable with payback periods less than 3.5 years.

Remaining parts of the paper are organized as follows. Section 2 presents background on LTE and formulation of deployment scenarios for Adama. Section 3 provides modelling, assumptions and parameter values for the TEA. Then Sect. 4 presents TEA results and discussion which is followed by concluding remarks in Sect. 5.

2 LTE and Deployment Scenarios for Adama

2.1 Background on LTE

To insure high speed demand, 3GPP introduces its new radio access technology called LTE in 2004 to support extensively high throughput and low latency system with improved coverage and capacity performance compared to UMTS 3G systems [8]. LTE developed to satisfy a set of high-level requirements such as reduced cost per bit, simple architecture and open interfaces, flexibility usage of existed and future frequency band and enhanced user experience-more services with lower cost and high speed. LTE's network architecture is introduced to be all - internet Protocol based on simplified network architecture with open interfaces. Its architecture consists of the evolved Node B (eNodeB), evolved packet core (EPC) and user equipment (UE) as shown in Fig. 1. The eNodeB provides the evolved UMTS terrestrial radio access network (E-UTRA) user plane (UP) and control plane (CP) protocol terminations towards the UE. The eNodeB's are interconnected with each other by means of the

X2 - interface. The eNodeB's are connected by means of the S1 - interface to the EPC. The S1 - interface supports a many-to-many relation between eNodeB's and mobility management entity/serving-gateway (MME/SGWs).

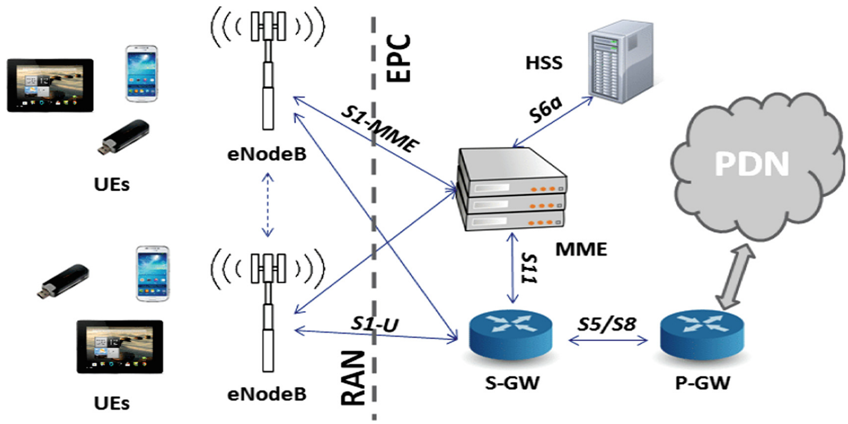


Fig. 1. LTE architecture [5, 7]

LTE is composed of several new technologies when compared to the previous cellular systems (2G/3G) such as multiple-input and multiple-output (MIMO), orthogonal frequency-division multiplexing (OFDM) and single carrier - frequency division multiple access (SC-FDMA). They enable LTE to be able to operate more efficiently with respect to the use of spectrum, and to provide the much higher data rates. LTE offers increased performance attributes, such as high peak data rates, low latency, and greater efficiencies in using the wireless spectrum over other wireless technologies [3].

2.2 LTE Deployment Scenarios for Adama

Adama city is one of emerging city in Ethiopia where potential customers revolve, hosts of many governmental and non-governmental conferences and center of trade next to capital city. In this city; high data rate capable technology such as LTE is vital for the development of the city and enhancement of quality of life in this era. Before proceeding to techno-economic analysis of LTE for Adama, we first need to formulate potential LTE deployment scenarios for the city using scenario planning method presented in [9, 10]. In this method, five important steps depicted in Fig. 2 are undertaken. Following scenario planning method; first the scope, time frame and actors are identified and defined for deployment of LTE in Adama city. Then, factors affecting LTE deployment in Adama city collected and analyzed using political, economic, social and technology (PEST) framework to find uncertain factors and events to deploy LTE in the city. Then, we formulate four potential LTE deployment scenarios using the two most important uncertainties and result is shown in Fig. 3.

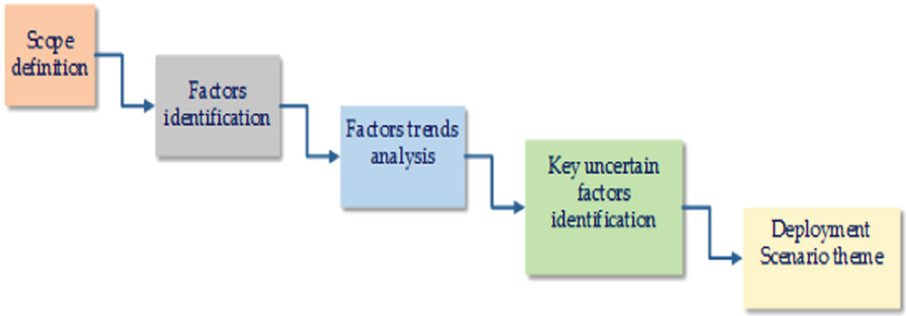


Fig. 2. General scenario planning in five phases [9]

The extrema value of each uncertain event defined as 1800 and 2100 MHz for frequency band, and high and low customers base for demand capacity. In uncertain events extrema value definition; global LTE popular band, handset operating frequency and status of currently existing devices (capability to support LTE) are considered. In addition, in-service spectrum extra bandwidth and its associated cost, trends of existing technology users and ethio telecom marketing plan into account for uncertain extrema definition. Threshold to say high and low demand capacity is defined based on 3G network capacity plan in the city (which is around 168,000 number of users). Then the scenarios defined further based on the other uncertainties and Adama context:

Deployment Scenario 1: Deployment of LTE in 1800 MHz under high demand capacity
Deployment Scenario 2: Deployment of LTE in 2100 MHz under high demand capacity
Deployment Scenario 3: Deployment of LTE in 1800 MHz under low demand capacity
Deployment Scenario 4: Deployment of LTE in 2100 MHz under low demand capacity.

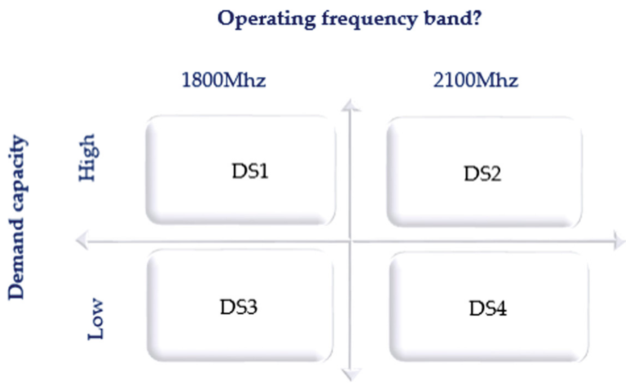


Fig. 3. Formulated LTE deployment scenarios

3 Techno-Economic Modeling and Assumptions

To analyze techno-economic viability of the potential deployment scenarios presented in Subsect. 2.2, we use modified TERA model that is shown in Fig. 4. TERA model is selected for this study as baseline model and widely used techno-economic model in wireless technology and industry [5–7]. We undertake minor modification of TERA model to improve the quality of inputs to suit our study. As shown in Fig. 4, the first part of model investigates marketing analysis which encompasses number of service users, the service user's penetration rate and revenue modeling parameters. The second component is the technology dimensioning which contains capacity and coverage dimensioning parameters. Capacity dimensioning requires input from two sides; from marketing section and the technology characteristics as per formulated deployment scenarios. Whereas, coverage analysis takes inputs of environmental condition and technology parameters to measure pathloss, system gain and loss. In the model, site number attained from both approaches are compared and taken as input to roll out plan to tune number of network elements inline to number of users forecasted within the study period. The third component is the economic part, this block modeled into two capital expenditure (CAPEX) and operational expenditures (OPEX) cost structures. The network dimensioning output, rollout plan and the associated network elements price is determined for CAPEX and OPEX cost modeling. Lastly, based on discount cash flow analysis, economic indicators like net present value (NPV), payback period (PP) and internal rate of return (IRR) are measured for each formulated deployment scenario.

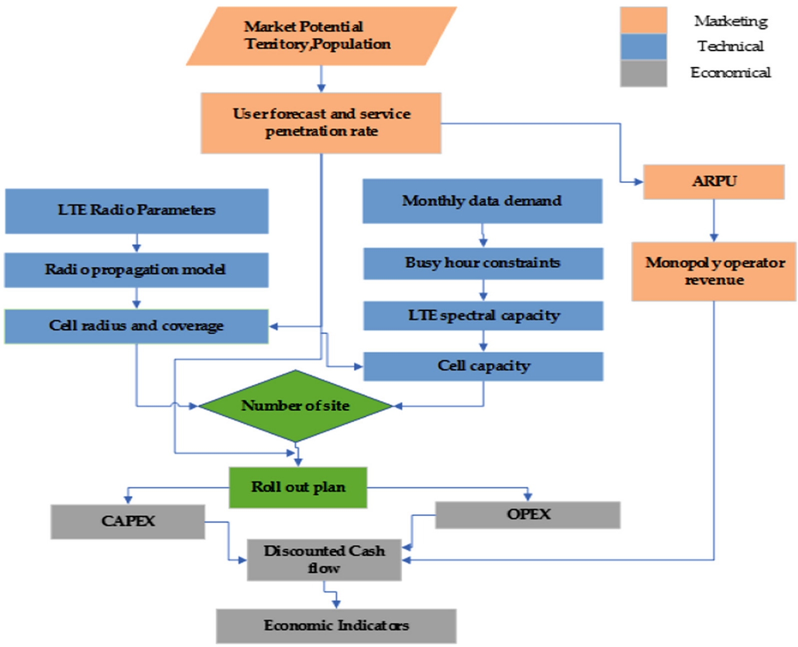


Fig. 4. Techno-economic model

3.1 Market Forecast

Understanding market segmentation, service profile and existing revenue collection is always the starting point for marketing forecasting as noted from different operators experience and literatures [5]. To forecast new technology service users', researchers frequently use the Bass-model. Bass model equation used in this study is stated below [11]. The coefficients determined in this paper in considering in-service technology service users' trends of past five years. Likewise, service penetration rate and operator revenue are projected in considering users forecast, service consumption level (monthly) and charging policy stated by operator. In this paper, market potential expressed with respect to current population of the city.

$$\mathrm{Y(t)} = \mathrm{M}\left[1 - \mathrm{e}^{-\mathrm{t(P+Q)}}\right] / [1 + [\frac{\mathrm{Q}}{\mathrm{P}}]\,\mathrm{e}^{-\mathrm{t(P+Q)}}] \tag{1}$$

Where, M is the market potential, P is innovation coefficient, Q is the imitation coefficient and Y(t) is the number of subscribers at a time t.

3.2 LTE Radio Link Budget Model

The target of the radio access network (RAN) dimensioning is to estimate the required site density and site configurations information. Dimension inputs can be broadly divided into three categories; quality, coverage and capacity-related inputs. Radio link budget has central importance to coverage planning in LTE [5–7]. The link budget calculations estimate the maximum allowed signal attenuation, called path loss, between the mobile and the base station antenna based on the required Signal-to-Interference-Noise Ratio (SINR) level at the receiver and by putting interference ad shadow into consideration. Detail radio link budget parameters assumptions considered in this study are summarized in Annex 1. We selected COST-231 Hata radio propagation model for coverage calculation and it is expressed as follow [8, 9].

$$\begin{aligned}\mathrm{PL(dB)} = 46.3 + 33.9(\mathrm{f}) - 13.82\log(\mathrm{h_{BS}}) + [44.9 - 6.55\log(\mathrm{h_{BS}})]\log(\mathrm{d}) \\ - \mathrm{a}(\mathrm{h_{UE}})(\mathrm{c_m})\end{aligned} \tag{2}$$

Where d is the distance (Km) between the transmitter and receiver antenna, f is the frequency used in MHz, h_{UE} is UE antenna height(m), h_{BS} is BTS antenna height(m), c_m is 0 dB for suburban and 3 dB for urban areas.

3.3 Capacity Dimensioning

To dimension the cell capacity, first it is essential to model data traffic based on monthly user's data consumption level. Accordingly, three user's categories (5, 10 and 15 GB) defined based on service usage ratio of 60, 30 and 10%, respectively. Further, 70/30% downlink/uplink traffic ratio considered for all users' categories. From these, average throughput per subscriber in busy hour is calculated using Eq. (3). Peak throughput per site in both uplink and downlink based on frame structure and coding rate is obtained using Eq. (4) [5, 11, 13]. The maximum number of users per site is calculated in considering users forecast, average throughput per subscribers and peak cell throughput for each formulated deployment scenarios. Hence, total number of sites obtained, in taking maximum number of subscribers from uplink and downlink.

$$\Omega(\text{Kbps}) = \rho * \alpha * \left(8\text{bit} * 10^6/30\,\text{day} * 24 * 60 * 60\right) * \text{ Usage ratio of the services } * \frac{\text{UL}}{\text{DL}}\text{traffic ratio} \tag{3}$$

$$\delta(\text{Mbps}) = \text{data RE/sec} * \text{ bits per RE} * \text{MIMO effect} * \text{ coding rate} * \frac{\text{UL}}{\text{DL}}\text{Traffic ratio} \tag{4}$$

Where Ω (Kbps) is average throughput per subscriber in busy hour (UL and DL), α is traffic ratio of busy hour to whole day traffic, ρ the traffic usage in Month/User. *δ* (Mbps) is the peak throughput per site per modulation, data RE/s is the data in resource element per second, coding rate indicates the volume coding rate of the channel code.

3.4 Cost and Revenue Modeling

The network cost generally splits into investments (CAPEX) and operational costs (OPEX). In techno-economic modelling, usually bottom-up approach is preferable and typically used when calculating investments cost. Whereas, operational costs are often modelled using a combination of the two approaches. CAPEX and OPEX cost modelled using Eqs. (5) and (6) [12, 13]. The cost of each parameter in CAPEX estimation was vendor specific. In this study, already deployed LTE network in Addis Ababa and industry standard per unit cost database considered for CAPEX cost estimation. Similarly, OPEX cost related to a certain project are often more difficult to predict. Therefore, the data for OPEX cost estimation mainly taken from annual OPEX report of currently in-service LTE network in Addis Ababa. In parallel, operator revenue also projected in consideration of number of service users forecasted, monthly data consumption level of each user and stated operator pricing policy (in this paper only service charging and connection fee parameters are considered).

$$CAPEX^{(i)} = \sum M_j^{(i)} C_j^{capex} \left(1 + P_j^{capex}\right)^{i-1} \tag{5}$$

$$\text{OPEX}^{(\text{i})} = \sum \text{N}_\text{j}^{(\text{i})} \text{C}_\text{j}^{\text{opex}} \left(1 + \text{P}_\text{j}^{\text{opex}}\right)^{\text{i}-1} \tag{6}$$

Where, j ∈ (key components of CAPEX), $\text{M}_\text{j}^{(\text{i})}$ is the number of items of type j purchased in year i, $\text{c}_\text{j}^{\text{capex}}$ are the per-unit investment for each asset j in the 1st year, $\text{P}_\text{j}^{\text{capex}}$ its yearly price trends and $\text{CAPEX}^{(\text{i})}$ the investment in year I, j ∈ (network driven and customer driven costs category parameters), $N_j^{(i)}$ is the number of items of type j operated during year i, $C_j^{(opex)}$ is Per unit operating cost for each asset j, P_j^{opex} j its yearly price trends.

3.5 Discounted Cash Flow Analysis

Commonly, if the return on investment is greater than total cost of ownership it is believed that the investment can make money. However, this approach is not enough to determine the feasibility of each formulated LTE deployment scenarios (DSs) as it only shows gross profit. Thus, we considered the economic feasibility indicators like net present value (NPV), internal rate of return (IRR), and payback period (PP) to maximize the decision indicators [5, 6]. All aforementioned economic indicators are calculated based on projected cash flow with in study period. In this paper, 13% discount rate considered for NPV calculation as it is a common value in most of literatures. From NPV rule, If NPV is positive, the formulated LTE DSs are economically feasible.

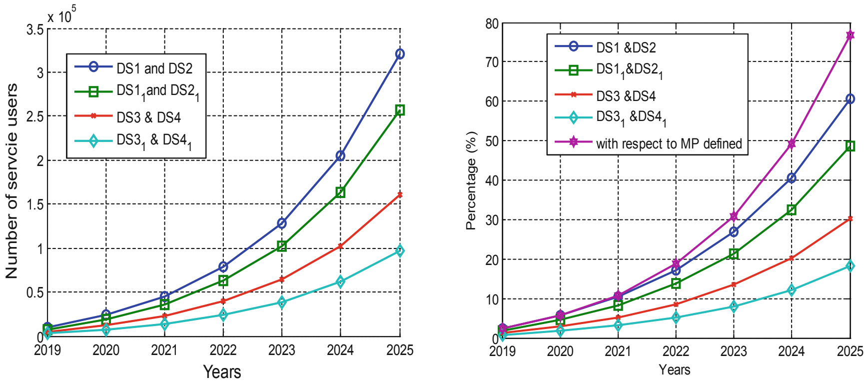
$$NPV = \sum_{t=0}^{T} \frac{CF_t}{(1+r)^t} \tag{7}$$

Where, t – the length of the study period, CFt – cash flow occurring in time period t and r – discount rate (industry average).

The IRR is the discount rate at which the present value of all future cash flow is equal to the initial investment. It is discount rate that makes NPV = 0. Whereas, PP is the number of years in which deployed LTE deployment scenarios takes before the cumulative incomes equal the initial investments.

4 Results and Discussion

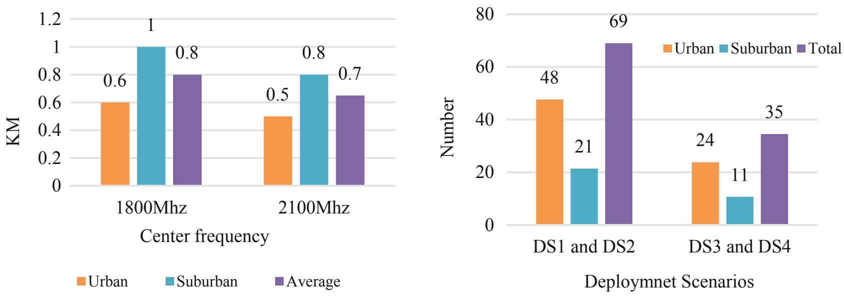
The data service users forecast results for each formulated deployment scenario with in seven years period based on defined market potential depicted in Fig. 5(a). For user forecast, past five years existing technology service users' trends in Adama city considered. Consequently, 2% innovation and 40% imitation coefficients values attained to forecast LTE data service users in Adama city. Users forecast under deployment scenario 1 and 2 (DS1 and 2), and deployment scenario 3 and 4 (DS3 and 4) basis on the market potential (MP) of 100% and 50% respectively. In addition, $DS1_1 \& DS2_1$ and $DS3_1 \& DS4_1$ are specially considered formulated deployment scenarios based on 80% and 30% MP respectively. The service user's penetration rate result with respect to defined MP and projected population of the city depicted in Fig. 5(b). The service users' penetration rate is less than 10% at the starting period of the deployment and after 7 years it is estimated to reach 75% with respect to defined MP for all explicit deployment considered due to the network externality effect.



(a) Service users forecast (b) Service penetration rate

Fig. 5. Service users forecast (a) and penetration rate (b)

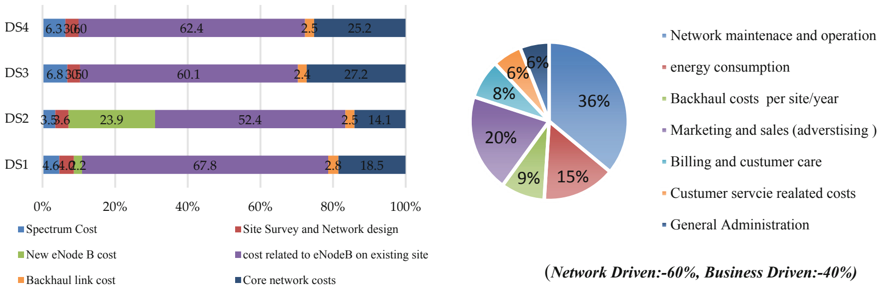
Based on cell edge data rate target definition (i.e. 2 Mbps and 1 Mbps for downlink and uplink respectively), projected SINR at the cell edge are 1.06 and 1.55 dB for both downlink and uplink respectively. Estimated cell radius in both considered environmental model and center frequency via COST-231 propagation model was depicted in Fig. 6(a). Further, to comprise the effect of multipath, average cell radius taken for number of site calculation based on geographical information and environmental model defined. From obtained cell radius and defined area, urban environmental conditions required a high number of evolved Node B's to mitigate the loss due to a multipath effect of transmission system compared to suburban areas.



(a) Cell radius (b) Number of sites

Fig. 6. Cell radius (a) and number of sites via capacity dimensioning (b)

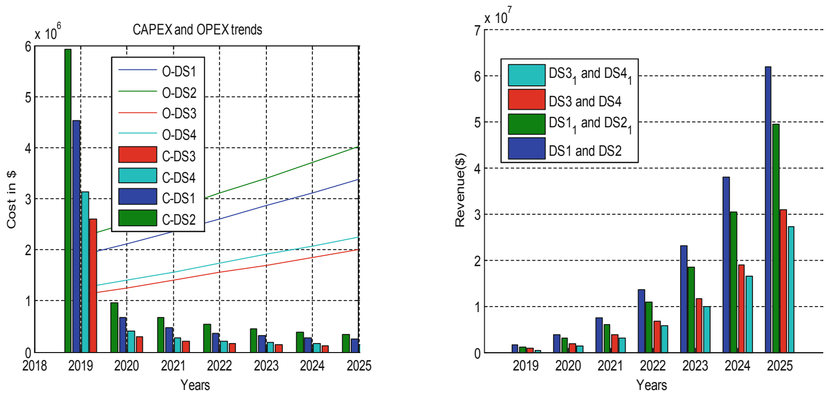
To dimension network capacity, number of users at the end of the study period considered as input. Number of evolved Node B's via traffic model are depicted in Fig. 6(b). In result, average downlink throughput per subscriber is almost twice that of the uplink due to the natures of data service users. Likewise, downlink throughput was greater than uplink due to slower transmission mode in uplink. Further, to minimize the cost of network elements and to use existing resource wisely, existing technology site number and setup considered through overlay network deployment strategy. Also, 70/30 rollout approach applied for chosen number of sites. The initial year investments cost and OPEX cost breakdown percentage with respect to its key component parameters for each deployment scenarios are shown in Fig. 7(a) and (b).



(a) Initial year investment breakdown (b) OPEX breakdown per site

Fig. 7. Initial year investment cost and annual OPEX breakdown

Yearly CAPEX and OPEX cost trends within the study period is calculated in consideration of 5% yearly equipment price trends as shown in Fig. 8(a). In view of ethio telecom pricing policy and monthly users data consumption survey (1 GB for residential and 10 GB for enterprise user's category), operator revenue is projected in a monopoly telecom market as shown in Fig. 8(b). In Fig. 8(a), "O-DS1 means OPEX cost for deployment scenario 1 (DS1), C-DS1 means CAPEX cost for deployment scenario 1(DS1) and etc…".



(a) CAPEX and OPEX cost trends (b) Operator revenue

Fig. 8. CAPEX and OPEX cost trends (a) and operator revenue (b)

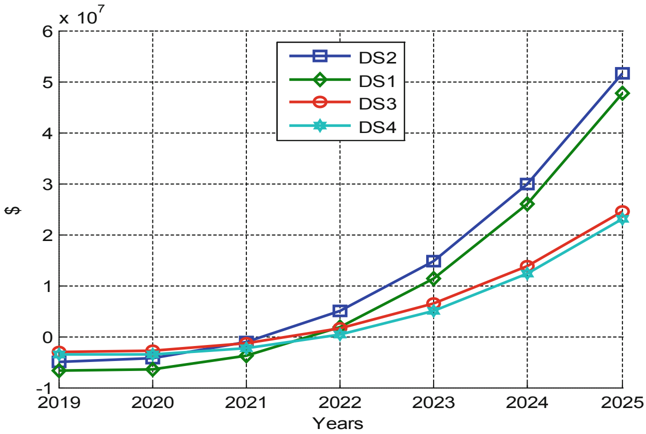


Fig. 9. Net present value (NPV)

As shown in Fig. 9, NPV is negative in the first three years and after the third year of the deployment, it is increases exponentially to high positive value. This exponential incremental arises due to the service user's behavior. Moreover, in all formulated deployment scenarios positive net present value recorded at the end of study period which indicates the economic feasibility of all formulated deployment scenarios. Similarly, as shown in Table 1, the obtained internal rate of return for all considered deployment scenarios are greater than 13% (value defined initially to calculate cash flow). Higher IRR values (greater than discount rate defined) indicates the economic feasibility of formulated deployment scenarios.

Table 1. Economic feasibility indicators value for each DS

	Deployment scenarios			
	DS1	DS2	DS3	DS4
Internal rate of return (IRR) (%)	82.1	68.4	72.7	65.3
Payback period (PP) (in years)	2.79	3.1	3.3	3.72

Moreover, the formulated LTE deployment scenarios are profitable and feasible if the payback period was in between 3 and 4 years (less than 3.5 years) in span 7 years as noted from several literatures. Hence, the payback period recorded for DS1, DS2, DS3 and DS4 are 2.79, 3.1, 3.3 and 3.72 years respectively. That means, the time (year) required to return the investment for deployment of LTE in an emerging city under DS1 is 2.79 years. Whereas, for the rest of LTE deployment scenarios the return on investments (ROI) achieved in 3.1, 3.3 and 3.72 years for DS2, DS3 and DS4 respectively.

5 Conclusion

Significantly increasing demand of innovative data service in emerging cities of developing countries is pushing deployment of state-of-the art mobile technologies including LTE. Yet, to be technically and economically viable, decisions on moving to LTE should be made after techno-economic study which is rare in the context of Africa mainly due to the common vendor driven mobile network deployment approach. Aiming to contribute in addressing this challenge, in this work we have presented techno-economic analysis of LTE for one of the emerging cities in Ethiopia, Adama. The analysis has been performed for formulated potential deployment scenarios (DSs) for the city assuming 7 years study period and continuation of existing monopoly telecom market in the study period. The analysis is made using modified TERA model that is implemented in Matlab.

From the results, market potential and operating frequency have a great impact on network capacity, coverage and number of evolved Node B which in turn influences the rate of return on investments in all formulated LTE deployment scenarios. In most of the explicit deployment, the system is coverage limited. And, for all formulated deployment scenarios, LTE data service users increases exponentially with slow start rate within the first three years of operation period. Likewise, total OPEX cost is greater than twice of CAPEX cost and high share of CAPEX cost taken by evolved Node B related cost due to chosen network deployment strategy (60–67%). Deployment of LTE in 1800 MHz band under high and low demand capacity and in 2100 MHz band under high demand capacity are feasible technically and economically for an emerging city with the payback period of less than 3.5 years within the span of 7 years operation duration. Additional, from specially considered formulated deployment scenarios, the optimistic and coverage favored deployment scenarios (in 1800 MHz band under high and low demand capacity) are the most feasible LTE DSs for an emerging city with payback period range of 3–3.25 years.

Annex 1

Detail radio link parameters assumptions.

Parameters	Variable for both UL and DL
Duplex Mode	FDD
Operating frequency	1800/2100
System Bandwidth (MHz)	20
Number of PRB per assigned BW	100 for 20
MIMO Scheme (MS)	1 × 2 UL, 2 × 2 DL
Cell Edge Rate (Mbps)	1 for UL, 2 for DL
Allocated RB at cell edge (10%)	10 for 20BW

(*continued*)

(*continued*)

Parameters	Variable for both UL and DL
Factor A	0.42 X2 for 2 × 2 MIMO, 0.42 for 1 × 2 MIMO
Factor B	0.85 for 2 × 2MIMO
PRB bandwidth (Mbps)	180 kHz
Rate per PRB (Mbit/s)	Cell edge/Allocated PRB
Required Spectral efficiency (bits/s/Hz)	Rate per PRB/PRB
SINR (LINEAR)	Factor B * (2^ (SE/Factor A)-1)
SINR (dB)	10 * log SINR(Linear)
eNodeB - UE	
Number of PRBs (10%)	10
Downlink data rate (Mbps)	1 for UL, 2 for DL
eNodeB TX power (dBm)	46
eNodeB antenna gain (dBi)	~18
eNodeB antenna cable loss (dB)	1–2
EIRP (dB)	=gains − losses
UE receiver characteristics	
UE noise figure (dB)	~7 dB
Thermal noise	B * T * PRB * BW
Receiver noise floor (dBm)	NF + B * T * PRB * BW
Required SINR (dB)	10 * log SINR(Factor B * (2^ (SE/Factor A) − 1))
Receiver sensitivity	Required SINR+ Receiver Noise
Control channel overhead (dB)	5%–25% (1 dB–4 dB overhead)
Rx antenna gain	0 for handset
Body loss	3–5 dB
Environmental characteristics	
Indoor Penetration Loss (dB)	12–16 dB
Cell Edge Coverage Probability (%)	90–95%
Shadowing Margin (dB)	9.25–10.5
Interference margin (dB)	5–8
Maximum Allowed propagation loss (MAPL) (dB)	=gains − losses

References

1. Cisco: Global mobile data traffic forecast update 2016–2021 (2017). https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html. Accessed 10 Oct 2018
2. Haile, B.B., Bulti, D.A., Zerihun, B.M.: On the relevance of capacity enhancing 5G technologies for Ethiopia. In: ICT Annual Conference (EICTC-2017), Addis Ababa, Ethiopia. https://www.researchgate.net/publication/320730399_On_the_Relevance_of_Capacity_Enhancing_5G_Technologies_for_Ethiopia. Accesses 10 Oct 2018

3. Knoll, T.: LTE network design from a techno-economic perspective. LTE-business-model, pp. 1–68 (2012)
4. Global Mobile Suppliers Association (GSA): Evolution from LTE to 5G: global market status, January 2019. https://gsacom.com/paper/evolution-lte-5g-2/. Accessed 02 Feb 2019
5. Fekadu, D.N.: Techno-economic analysis of LTE deployment scenarios for emerging city: a case of Adama, Ethiopia. Master's thesis. AAiT, Addis Ababa, Ethiopia 2018 (2014)
6. Smura, T.: Techno-economic modelling of wireless network and industry architectures. Doctoral dissertations. Aalto University School Electrical Engineering, Helsinki (2012)
7. Frias, Z., Pérez, J.: Techno-economic analysis of femtocell deployment in long-term evolution networks. EURASIP J. Wirel. Commun. Netw. **2012**, 1–15 (2012)
8. GPP Website: 3GPP technologies. http://www.3gpp.org/techneologies. Accessed 23 Feb 2018
9. Schoemaker, P.: Scenario planning: a tool for strategic thinking. Sloan Manag. Rev. **36**, 36–40 (1995)
10. Haile, B.B.: Development of the Wi-Fi offloading business concept within the African market context, Addis Ababa, Ethiopia (2012)
11. Sokele, M.: Analytical method for forecasting of telecommunications service life-cycle quantitative factors. Doctoral dissertations. University of Zagreb, Croatia (2009)
12. Vidal, J., Pérez, J., Agustín, A., Marina, N., Høst-Madsen, A.: Technoeconomic evaluation of cooperative relaying transmission techniques in OFDM cellular networks. EURASIP J. Adv. Signal Process. **60** (2012)
13. Arif, M., Yameen, M., Matin, A.: Femtocell suburban deployment in LTE networks. Int. J. Inf. Electron. Eng. **3**(2), 208 (2013)

SAMoC: Self-aware Access Monitoring and Controlling Framework for Android

Nanda Kumar Thanigaivelan, Ethiopia Nigussie(✉), Seppo Virtanen, and Jouni Isoaho

Department of Future Technologies, University of Turku, Turku, Finland
{nakuth,ethnig,seppo.virtanen,jisoaho}@utu.fi

Abstract. We present the concept and specification of a self-aware access monitoring and controlling framework for Android. The lack of users' security awareness, Android operating system's security limitations and the widespread use of Android for personal and organizational activities emphasizes the importance of improving its security. The proposed SAMoC framework strengthens the security of Android through learning based operation refinements. The introduction of a monitoring agent on each resource and a self-aware layer in the Android software stack allows to detect unexpected events and perform fine-grained access control over the resources. The self-aware layer can automatically refine the policies or provide suggestions depending on its configuration. The SAMoC framework enforces specific restrictions on an individual user basis and allows the users to play a role in the system operation optimization due to its transparent nature.

Keywords: Self-aware security · Mobile security · Access control

1 Introduction

Mobile devices are becoming the dominant devices for communication and online service consumption among individual users and organizations. The exponential growth in the number and diversity of the available applications is the main reason for widespread popularity of mobile devices. The increase in popularity of mobile devices along with users' limited security awareness, and operating system vulnerabilities make these devices easy target for security attacks. An increasing number of organizations is allowing employees to access the enterprise's resources through mobile devices and even encouraging to use their own devices (Bring Your own Devices (BYoD)). In addition to enterprise apps, a number of other apps from various channels may also be installed in the same device, exposing the enterprise to numerous security threats and attack vectors.

This work focuses on access control for Android based mobile devices to ensure their security. Android is an open-source operating system developed based on the Linux kernel [1]. The Android operating system deploys mechanisms such as application sandboxing and the Android permission model. The

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 227–238, 2019.
https://doi.org/10.1007/978-3-030-26630-1_19

sandboxing provides application isolation and containment using Linux access control and process protection mechanisms. The permission model restricts an application's capabilities by regulating sensitive API calls that access protected resources. Permissions to access certain resources are granted during first access request. Users tend to approve all requested permissions since the users may not understand fully the consequences. Though runtime permission revocation is introduced in Android 6.0. it still fails to allow fine grained permission control. Third-party libraries (for example, advertisement and analytics) cannot be prevented from abusing the granted permissions of their host apps because of the unavailability of separation mechanisms. In other words, Android lacks runtime configurable access control for preventing an application from accessing any open interfaces of another application and abusing its granted permissions.

The future mobile devices will be part of the internet-of-things (IoT) and they can operate as sensing, actuating, and intermediate gateway devices in IoT. Intercommunication among mobile devices is highly likely in future IoT applications. These activities will enhance the roles and usage of mobile devices which makes them an attractive target for attacks since the impact of attacks can be higher. Therefore, it is necessary to come up with a system which is able to reconfigure and adapt its operation to ever evolving threat. To counter such circumstances, the system has to be self-aware.

In this work, we present the concept and specification of a self-aware access monitoring and controlling framework for Android. The main working principle of the framework is to monitor, to learn and then to refine existing policies or defining new ones at runtime. A self-aware layer is introduced in the Android software stack with the sole intention of learning from the monitored information and implementing appropriate controlling measures. The contributions of this work are as follows:

- Introduction and adaptation of self-awareness in Android Mandatory Access Control (MAC): to detect and/or prevent unexpected events including the current and future intentional or unintentional malicious activities and resource abuses.
- Self-optimization of operations: through continuous monitoring and learning, the framework defines new access policies or refines existing policies which in turn optimizes the operation and resource usage.
- Multiuser environment support: to allow enforcing different restriction policies and customizing them accordingly for each user account in a single device.

The rest of the paper is structured as follows. Android and its security limitations are presented in Sect. 2. The key features of the proposed self-aware access monitoring and controlling (SAMoC) framework are discussed in Sect. 3. The concept of self-awareness in general and its importance in Android security context are explained in Sect. 4. The detailed specification of the SAMoC framework is presented in Sect. 5. Finally, discussion and conclusion are presented in Sects. 6 and 7, respectively.

2 Android and Its Security Limitations

Android is a Linux-based open source mobile operating system, developed by Open Handset Alliance led by Google. The Linux kernel is customized for Android to provide features such as Binder, asynchronous shared memory, process memory allocator and power management. The kernel layer acts as a hardware abstraction layer and provides various services such as networking, file system and device drivers. The middleware layer consists of Android framework, native and runtime libraries including Android Runtime (ART). It is responsible for handling application life cycle management and other system services along with the access restrictions on application resource accessibility. The application layer hosts the core/system applications and other applications written by Android developers. In addition to Java, developers can use C/C++ through the Java Native Interface (JNI). The Android application architecture is unique in such a way that it is designed to provide application compatibility, portability and security.

2.1 Android Security Model

Since Android is Linux-based operating system, it inherits user resource isolation and process isolation from Linux. Android extends the user resource isolation feature by extending to application level. Under this modification, each application is assigned with unique ID (UID) upon installation and executes it as a dedicated process using the same UID. Also, dedicated data directory for its resources is given and restricts to access other files unless explicitly permitted. The system resources such as daemons and system applications are owned either by system or root user and executed under pre-defined UID. Thus, it achieves isolation in terms of both process level as well as file level.

Due to the isolation, application cannot access any resources other than their own resources. In order to provide access to the resources such as internet connectivity, data and hardware features, Android introduced access rights in the form of permissions. Application required to access those features need to obtain appropriate permission and the developer must declare the required permission in the AndroidManifest. Android offers 130 permissions and also allow developers to define their own permissions to enforce access restrictions on their application's critical application resources [7]. Android framework enforces access verification during runtime to ensure applications has appropriate access rights. There are few permissions which are not monitored by the Android framework since they are mapped to the low level operating system control.

2.2 Limitations

Isolation of applications and their resource accessibility is contained through application sandboxing. Android also caters permission to access other resources in order to capitalize their features for enhancing user experiences. Due to its popularity and continuous increase in user base, it has attracted the attention of research community to identify the vulnerabilities and possible solutions [2–5,8–10]. Also, it has gained the attention of adversaries, who aim to exploit the vulnerabilities for their own benefits. To enhance the security through the enforcement of MAC at kernel layer, SE Linux is introduced into the Android platform since version 4.3 (as SEAndroid). The implementation of SE Linux is still evolving in Android from permissive mode in version 4.3 to full enforcement mode in version 5.0. SEAndroid provides huge advantage in protecting the device, for example, it has the capability to prevent application installation based on its signatures [6]. However, it is not possible to define the certificates for all available applications in the policy. Extending SE Linux to cover the Android's middleware is challenging mainly because of implementations complexity as well as requiring an invasive and costly set of changes [6]. Due to these constraints the middleware is loosely protected by Android's default security model. SE Linux also requires specialized skills for implementing and enforcing policies.

From version 6.0 onwards, several enhancements are introduced and one example is runtime permission revocation. It allows users to change (grant or revoke) the status of the applications' permissions as needed. This is one important step in improving security but it fails to provide fine granularity for controlling the resources. In this work, we are presenting a framework which provide fine granular access control, and more controllable system features.

3 Key Features of SAMoC Framework

A self-aware system is capable of knowing its environment, the on-going activities and implementing appropriate corrective measures at run-time in order to achieve and maintain the required performance. By introducing a self-aware concept in mobile MAC, we are aiming to achieve the following main goals:

- Self-awareness: to reduce user intervention in security policy enforcement and maintenance by creating self-aware components. These components can access logs, learn, able to predict unexpected behaviors/events, and refine/define new policies accordingly.
- Highly refined policies: to attain higher granularity in policy refinement, access restriction enforcement based on the resource features rather than the enforcement on application level or resource level will be introduced. For example, in the proposed framework, instead of blocking entire internet access resource, it is possible to allow certain protocol on particular port(s) and deny the rest of the protocol for an application.
- BYOD and parental solutions: The system will provide multi-user support and it is tightly coupled with device users. This allows enforcement of different sets of control policies and customization of each policy set for individual user account in a device.
- Improved system resilience: The integration of five subsystems into the framework (application installation and monitor, device resource controller, communication controller, self-aware subsystem, and SAMoC user interface)

enhances the system's resiliency by allowing to run atleast one subsystem continuously. The user can decided to enable/disable the certain subsystem(s)/policies and keep the rest active all the time.
– Easy to use: since there is no predefined structure/syntax for writing most of the policies, the personal users can easily adopt to the system.
– Transparency: Allowing device users to access the system logs and the enforced policies enables the users to understand behind-the-scene operations either fully or partially depending on their knowledge.

4 Adaptation of Self-awareness in Security Context

A number of researches have been conducted and many solutions have been proposed to thwart the threats in the Android but most of them are concentrated on either providing resource access restriction or prevention of identified threats and few works target both cases [5,6]. The existing research on mobile security fail to recognize the assimilation of mobile devices in IoT and the resulting change in the mobile usage landscape. Mobile device are part of IoT and play an important role in IoT as a sensing, actuating, and/or intermediate gateway. The interoperability of mobile in IoT is facilitated due to the introduction of 6lowPAN based wireless sensor networks. There is also on-going effort to connect IP networks through bluetooth low energy [11,12]. One potential example of this integration is the possibility of gathering information through crowd sensing by using the sensors available in the mobile devices along with other sensor networks [13,14]. Intercommunication between the mobile devices is highly likely in future IoT applications. In these circumstances, it is difficult to know the security threats beforehand. In order to protect the device resources, we need to develop a system which is capable of self-reconfiguration and self-adaptation. To realize this objective, we have to make the system self-aware.

Any system which has the capability to learn and adapt itself through monitoring its own operations and environment is termed as self-aware system. Implementation of self-aware systems is not an easy task but they can offer numerous benefits upon proper employment.

4.1 Self-awareness in Security Context

The advantages of self-awareness especially detection and prevention of malicious activities through monitoring and learning are critical to the security of the systems since threats or threat sources are dynamic and evolving. To capitalize on these benefits, we have adapted the self-awareness into the security context. The customized self-aware agent is shown the Fig. 1.

The self-aware agent comprises of five components and the tasks of each component are described as follows.

Self-configure: the purpose of self-configuration is to update the settings of the self-aware agent components by considering the previous and current circumstances to ensure the required objectives are fulfilled with minimal overheads.

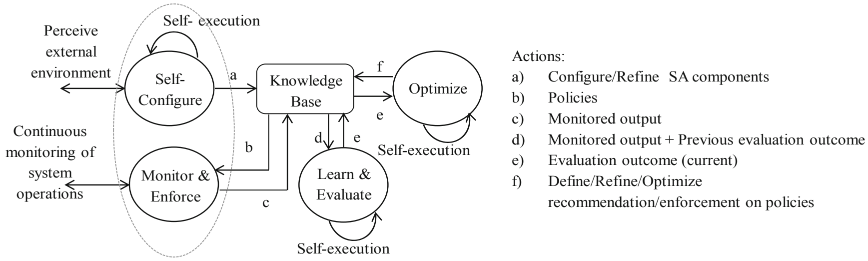


Fig. 1. Customized self-aware agent for security context

The self-configure component is also responsible for guiding the rest of self-aware agent components execution for efficient power and device usability. It will oversee the internal and external factors such as peek device usage, active and idle duration, and then define/refine execution period for itself, *Learn and Evaluate* and *Optimize* components.

Monitor and Enforce: it engages in continuous observation and enforcement of the policies on the systems activities. It ensures that the applications behavior and system resources accessibility conform with the provided policies. It is also responsible for transferring the observation to the *Knowledge Base (KB)*. The information observed by this component will be used as a basis for the *Learn and Evaluate* component.

Learn and Evaluate: this component has two operating phases. In the learning phase, information gathered by *Monitor and Enforce* component are analyzed in order to derive the common system behavior patterns. On successive execution of this phase, it will generate a new pattern if there is a change in the systems application behavior, otherwise the same common pattern will be generated. In the evaluate phase, the identified pattern is compared against the common patterns. Upon on detection of deviation, it will construct appropriate changes and submit the recommended changes to the *KB*.

Optimize: this component examines the recommendations provided by the *Learn and Evaluate* component and change the policies accordingly or provide them as suggestions to the user depending on the chosen configuration. It is also responsible for restoring the policies to the previous states if the applied changes fails to fullfill the requirement.

Knowledge Base: is used as a storage space for all activities. It contains the policies, system activities as logs, derived patterns, and policy recommendations, along with the setting configuration of the self-aware agent. As can be seen from Fig. 1, all the inputs and outputs for the other four components of self-aware agent are from and to the *KB*.

5 SAMoC Framework Specification

The primary objective of the proposed framework is to achieve higher degree of security by inhibiting malicious or unintended activities through highly refined policy enforcement, self-learning, and self-reconfiguration. The system which is developed by incorporating the SAMoC framework will reduce the requirement for human involvement in making more appropriate decisions which in turn improve the security of the mobile device as well as system usability. The framework engages in continuous monitoring of the resources and communication channels to ensure that all apps are working appropriately according to the defined policies. It also performs periodic assessment of monitored information as this is necessary for policies and processes refinement. Since the framework will be ingrained into the Android platform, it will start functioning upon device boot. The framework operates along with Android's default security implementation.

5.1 Assumptions

As in other system developments, we have defined certain assumptions to ensure proper functioning of the system. The users and administrators who are configuring the polices are fully trusted and they required to posses knowledge on system functionality principles. We assume that no application is granted root permission and the device will not be rooted in any circumstances as this may result in applications to abuse system resources intentionally or unintentionally. The Android operating system and the components that we are introducing are fully trusted.

5.2 SAMoC Mobile

The mobile part consists of the following components: handlers, self-aware components and synchronization components. The architecture of the SAMoC mobile is shown in Fig. 2.

Handlers: The handlers consist of two major components: policy handlers and event-log handlers. They play a crucial role in the framework since all the communication towards logs and policies are directed by them. Policy handlers are responsible for providing appropriate policy to the monitoring and enforcing component to act upon the resources. Event-log handlers perform log maintenance by allowing other SAMoC mobile components to read or write the logs as and when they needed.

Self-aware Component: Self-aware component is the core of the framework. It comprises of self-configuration, monitoring and enforcing components, evaluation and optimization components. The self-configuration component is responsible for guiding the rest of self-aware components by updating their execution settings. Monitoring and enforcing enhances the normal Android execution flow by placing appropriate hooks in the Android components which are responsible for handling applications and their device services accessing capabilities. The evaluation component is responsible for making policy recommendations by accessing the logs, identifying the correlation and establishing relationship among the correlated information. The optimization component will refine the existing policies, define new policies or recommend changes to the user based on the recommendations provided by the evaluation component.

Synchronization Components: Synchronization component is responsible for handling request access to the policies by the SAMoC UI. In addition, it will also handle compression, decompression and purging old logs from the repository.

Repository is a storage component which contains the policies and logs generated by the framework components. It will also acts as knowledge base for self-aware components by storing and retrieving the settings required for their self-execution, behavioural patterns and policy recommendations.

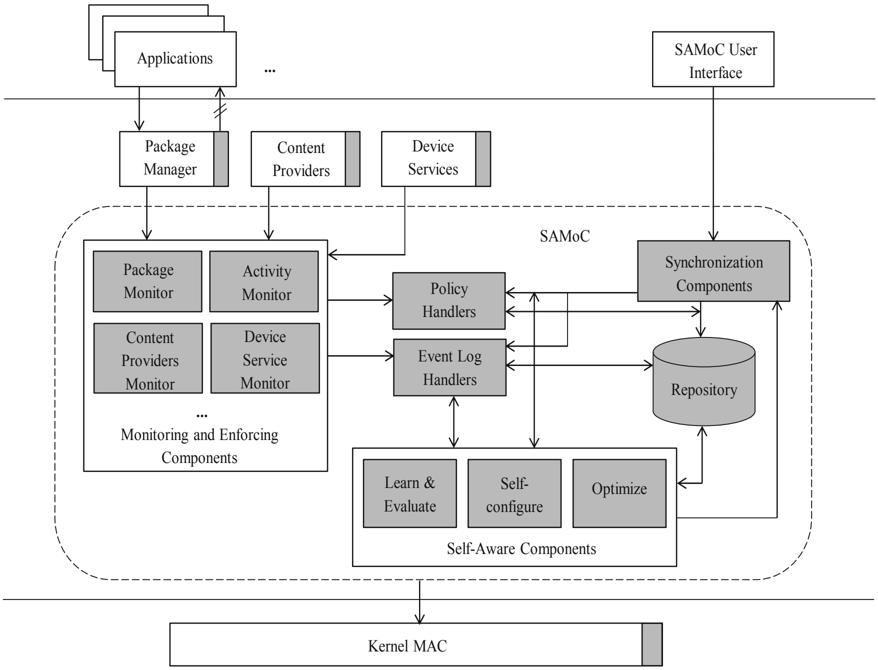


Fig. 2. Mobile software stack with SAMoC framework. The fully shaded blocks are the proposed new modules of the framework and the partially shaded blocks consist of framework module extensions in Android.

5.3 SAMoC Mobile Subsystems

One of the goals of the framework is to guarantee the existence of least degree of security at any time in order to maintain system resiliency. In order to achieve this goal, SAMoC mobile is classified into five subsystems: application installation and monitor, device resource controller, communication controller, self-aware subsystem and SAMoC user interface. The subsystems relies on different components and in certain cases, they can share the same components for the fulfilment their operations.

Application installation and monitor subsystem is responsible for handling the applications related activities including installation restriction and inter-communication between the applications.

Device resource controller functions are responsible to oversee the resource accessibility by the applications and enforce appropriate policies to restrict the access.

Communication controller will regulate the applications communication with external networks. It is in charge for allowing or refusing the outside connections based on the protocols and networks.

Self-aware subsystem observes the external environment and configure its own operations accordingly. It will also access logs, learn, create and test the established patterns, construct and apply/revoke suggestions to improve the framework operations.

SAMoC user interface is the interface application for SAMoC mobile which provides appropriate user interfaces for different subsystems. The user is able to control the subsystems or access logs through the interface.

5.4 SAMoC Mobile Software Architecture

In order to reduce development complexity as well as future enhancement and maintenance, we have adopted layered architecture for the implementation. The layered architecture allows to have complete control over the SAMoC mobile components individually and it is easier to add additional components in future. The architecture also helps in restricting the visibility of the implementations. The SAMoC mobile software architecture is shown in Fig. 3.

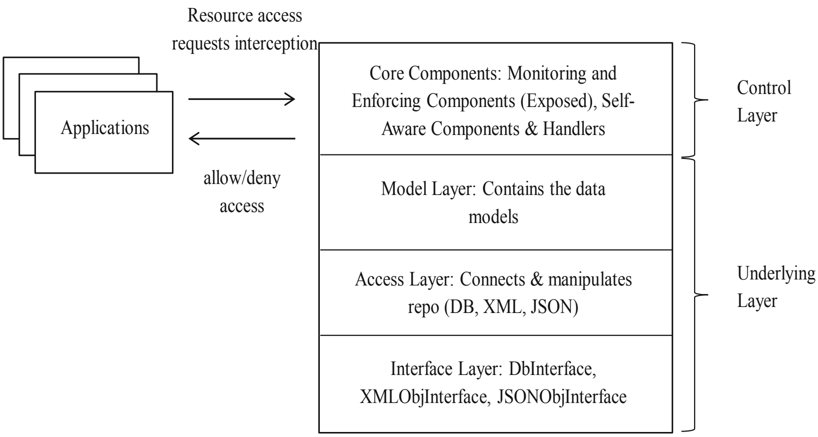


Fig. 3. SAMoC mobile software architecture

The entire architecture is grouped into two layers: control layer and underlying layer. The control layer contains all the implementation of the SAMoC mobile components (handlers, self-aware component and synchronization). The visibility of the handlers and self-aware component except the monitoring and enforcing tasks are hidden. The monitoring and enforcing components are visible and hooked into the Android's system service implementations while synchronization component provides interface to SAMoC application (SAMoC user interface). Using SAMoC user interface, users are able to perform administration tasks, such as writing policies and reviewing logs. It will be installed by default as system application in Android operating system.

The underlying layer comprises of three sublayers (Model, Access and Interface Layers) which contains the basic classes required to manipulate the repository. The model layer represents the data models of the database tables and XML structures. The access layer will manipulate the repository objects whose references will be provided by the interface layer. The Control layer components use the models to store/retrieve the information by calling the Access layer methods which in turn get the appropriate reference on the information objects from the Interface layer.

5.5 SAMoC Mobile Policy

SAMoC mobile policies are highly refined which restrict access to resources using its very own features. The efficiency of the framework depends on the policies available in the repository which will be defined by the users. The defined policies play a crucial role in determining the system efficiency during initial phase of its operations. There are certain control policies which will be delivered as settings to control the self-aware agent behaviors such as automatic enforcement of learned changes or provide them as recommendation to the user.

There are no pre-defined profiles/profile names in the SAMoC framework but pre-defined policies are available. SAMoC framework has default policies which will be enforced upon device boot for example denial of non-market application installation and denial of adb install command. Policies are created and enforced on individual account basis.

SAMoC framework does not force users to learn/develop special skills before using the system since most of the policies do not have any syntax or pre-defined structure. One of the goals of SAMoC is to make the system easy to use even for the user who has limited knowledge. To realize this, most of the policies will configured through the SAMoC user interface application. It will provide appropriate user interface in the form of settings for policy writing.

6 Challenges

At the current development stage, self-awareness including self-configuration of the self-aware agent is adopted. The self-configuration allows to configure and adapt the learn and evaluate processes, and to optimize the components of the self-aware agent according to the external environment and device usage. The learn, evaluate and optimize processes can not run continuously due to the following reasons: (1) for optimal learning, they require a sufficient number of log entries, (2) initiating the learning process results in resource wastage if there is no considerable number of new log entries recorded in-between the successive learning sessions, and (3) these processes should not consume the resources while the device is busy with other activities. These three processes run on certain time frames and the self-configuration subsystem is responsible for deciding the ideal time frame for the execution. Care has to be taken in determining the successive time frames. If the interval between successive time frames is too large, the device may fail to detect unexpected events or new type of attacks because the self-aware agent has not learned the new patterns. If it is too small, it leads to unnecessary resource consumption. As discussed, the self-configuration component is in a key role in configuring the initial settings of self-aware agent components which will have impact on identification of new threat patterns and optimization of the controlling process. To devise appropriate logic for implementation of these tasks requires careful analysis.

In the framework, two options for optimization are given. The first option is the self-aware agent taking the optimization action by itself without requiring the user's approval. The other one is the self-aware agent providing suggestions for optimization to the user. The decision to accept or deny the recommendations will be made by the user. In this case, the enterprise or individual users, who may hesitate to adopt the SAMoC framework can choose the second option. These options grant flexibility in controlling the optimization component only. The rest of the self-aware agent components will operate on their own but their processes can be reviewed through monitored logs. It is not wise to control the entire self-aware agent operation since the threats evolve dynamically. Therefore, the system which handles the threats needs to be continuously improved, especially on its own. In addition, to detect the unforeseen events or unknown threats, the system requires agility, and to be agile, it has to operate on its own.

7 Summary

The concept and specification of a self-aware access monitoring and controlling framework (SAMoC) for Android were presented. The SAMoC framework introduces the self-aware agent into the Android operating system to harden the security of the devices. The agent performs information gathering, learning from gathered information, evaluation of the available policies in comparison with the learning outcome and optimization or suggestion of the necessary improvements. The integration of a self-aware agent enables the SAMoC Framework to detect unexpected events, enforce highly refined restrictions over the resource accessibility and change the statuses of the enforcement (revoke and grant rights) during runtime. Provisions are provided to control the framework's subsystem operations, such as disabling an individual subsystem and implementing/suggesting the policies' refinements. The framework also allows to apply different sets of control policies in a single device in a multi-user environment and has provisions to engage users in the security operations and enhancements. All these features of the framework make it a prominent choice for Android security.

References

1. The Android source code. https://source.android.com/source/index.html
2. Arzt, S., et al.: Flowdroid: precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps. ACM SIGPLAN Not. **49**(6), 259–269 (2014)
3. Heuser, S., Nadkarni, A., Enck, W., Sadeghi, A.-R.: ASM: a programmable interface for extending Android security. In: Proceedings of 23rd USENIX Security Symposium (SEC 2014). USENIX (2014)
4. Wang, X., Sun, K., Wang, Y., Jing, J.: DeepDroid: dynamically enforcing enterprise policy on Android devices. In: Proceedings of 22nd Annual Network and Distributed System Security Symposium (NDSS 2015). The Internet Society (2015)
5. Bugiel, S., Davi, L., Dmitrienko, A., Fischer, T., Sadeghi, A.-R.: Xmandroid: a new Android evolution to mitigate privilege escalation attacks. Technical report TR-2011-04. Technische Universität Darmstadt (2011)
6. Smalley, S., Craig, R.: Security enhanced (SE) Android: bringing flexible MAC to Android. In: Proceedings of 20nd Annual Network and Distributed System Security Symposium (NDSS 2013), vol. 310, pp. 20–38. NDSS (2013)
7. Android Developers-Mainfest.permission. http://developer.android.com/reference/android/Manifest.permission.html
8. Enck, W., et al.: TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. ACM Trans. Comput. Syst. (TOCS) **32**(2) (2014)
9. Fawaz, K., Shin, K.G.: Location privacy protection for smartphone users. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 239–250. ACM (2014)
10. Wei, F., Roy, S., Ou, X.: Amandroid: a precise and general inter-component data flow analysis framework for security vetting of Android apps. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 1329–1341. ACM (2014)
11. Isomaki, M., Nieminen, J., Gomez, C., Shelby, Z., Savolainen, T., Patil, B.: IPv6 over BLUETOOTH (R) Low Energy (2015)
12. Wang, H., Xi, M., Liu, J., Chen, C.: Transmitting IPv6 packets over Bluetooth low energy based on BlueZ. In: Proceedings of 15th International Conference on Advanced Communication Technology (ICACT), pp. 72–77. IEEE (2013)
13. Skorin-Kapov, L., Pripuzic, K., Marjanovic, M., Antonic, A., Zarko, I.P.: Energy efficient and quality-driven continuous sensor management for mobile IoT applications. In: Proceedings of 2014 International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), pp. 397–406. IEEE (2014)
14. Angelopoulos, C.M., Evangelatos, O., Nikoletseas, S., Raptis, T.P., Rolim, J.D.P., Veroutis, K.: A user-enabled testbed architecture with mobile crowdsensing support for smart, green buildings. In: The Proceedings of 2015 IEEE International Conference on Communications (ICC), pp. 573–578. IEEE (2015)

Internet of Things Technology for Agriculture in Ethiopia: A Review

Tesfa Tegegne(✉), Hailu Beshada Balcha(✉), and Mebiratu Beyene(✉)

Bahir Dar Institute of Technology, ICT4D Research Center, Bahir Dar University, Bahir Dar, Ethiopia
tesfat@gmail.com, beshadahailu@yahoo.com, mabiratubeyene@yahoo.com

Abstract. The IoT (Internet of Things) provides an opportunity for technology to transform many sectors, especially agricultural sectors. In agriculture sector, IoT technologies appear in different application areas like precision agriculture or smart farming, irrigation monitoring, environmental monitoring, cattle animal monitoring and in all other fields of agriculture. The different technologies of IoT, such as radio frequency identification (RFID), wireless sensor network (WSN), sensors and global positioning system (GPS), had been widely used in different agricultural industry process. The advancement IoT technology has been widely applied in agricultural production, process, sales and circulation which brought a number of a great benefit for maintain and monitor agricultural and animal product process easier and automatically, improving the efficiency and speed of operations in their process management. This review paper aimed to review and introduce application of IoT in agriculture, particularly, to encourage the adoption of IoT technologies for Ethiopian agricultural sectors. Indeed, this reviewed paper the key technologies of IoT in agriculture, the general framework model of IoT, IoT based system deployment architecture and potential challenges of applying IoT based agriculture were discussed.

Keywords: IoT · Agriculture · Production · RFID · WSN · Sensors · GPS

1 Introduction

Agriculture plays a significant role for economic and social development in most undeveloped countries like Ethiopia. “Agriculture is the backbone of the Ethiopian economy. This particular sector determines the growth of all other sectors and consequently, the whole national economy” [1]. Ministry of Finance and Economic Development (MoFED), revealed that agriculture contributes greatly in terms of export, employment, and subsistence to the Ethiopian economy. It contributes 42.7% of GDP, 85% of employment (the rural population of Ethiopia), 90% of earnings from export, and 70% of raw material requirements for large and medium industries that are agricultural processing [1, 2]. Agricultural extension is an educational service which brings

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 239–249, 2019.
https://doi.org/10.1007/978-3-030-26630-1_20

information and new technologies to farming communities to enable them improve their production, incomes and standards of living. With the problem that extension agents face in facilitating direct contact with farmer clients and with researchers due to the physical distances involved and lack of transportation needed for their mobility. It is obvious that, while production and productivity targets are generally achievable, the country needs to adopt more cost-effective, innovative and modern approaches to agricultural knowledge management, and to reform and modernize its agricultural extension system. Advancement in information and communication technologies (ICT) has facilitated doing business and economic development in the world [3, 4]. Use of ICTs can be a major intervention for more efficient agriculture.

The widespread of Internet in the last two decades brought countless benefits to citizens and organizations around the world [5]. Information of adequate quality is a necessary condition for improvement of all areas of agriculture. The application of ICTs offers excellent possibilities, for strengthening research extension systems and beyond the urban focus. Recently, the IoT is promising to bring the same benefits to everyday objects, giving us a way to extend our perception and our ability to modify the environment around us. The experience in Ethiopia rapid development of ICT, which facilitates the flow of data and information, has tremendously enhanced the knowledge management practice in agriculture. In this context, agricultural industrial and environmental fields are ideal candidates for the deployment of IoT solutions because they occur in wide areas that need to be continuously monitored and controlled [6]. At the same time similarly, “IoT opens new opportunities beyond ground floor automation when the collected data are used to feed machine learning algorithms to provide predictions, easing decision planning and decision making for owners, managers, and policy makers” [6].

IoT can be used at different levels in the agricultural production and processing in Ethiopia like precision agriculture or smart farming, irrigation monitoring, environmental monitoring, cattle animal monitoring and in all other fields of agriculture. It can help to evaluate field variables such as soil state, atmospheric conditions and biomass of plants or animals. It can also be used to assess and control variables such as temperature, humidity, vibration, and shock during the product transport [6]. It can be used to monitor and predict the product state and its demand on shelves or inside refrigerators. In addition, it can provide information to the final user/consumer about the origin and properties of the product. The IoT applied to the agricultural industry can contribute to create an informed, connected, developed and adaptable rural community.

The objective of this paper is asses the state of the art for IoT based Ethiopian agricultural system and identify potential application areas of IoT in Ethiopian agricultural system to improve productivity and economic, social and environmental sustainability.

2 IoT in Agriculture

With the continuous development of times and information technology (IT), networking technology provide important input for industry development. The easiest thing to understand IoT is based on Internet, mobile communication, and other communication networks, through intelligent sensors, radio frequency identification, infrared sensors, global positioning systems, laser scanner s, remote sensing, etc. [7]. Agricultural IoT technology is used to combine variety of sensor organizations to form the sensor network, through which we can collect farmland sensor information (field planting, horticulture, livestock and poultry, aquatic product s' site information breeding, agricultural logistics, and other areas) and real- time analysis of the results transmitted to farmers to enable them to make the most profitable decisions.

Most researchers done IoT based agricultural system on different specification and parameters. Currently, IoT applications in agriculture are mainly in the following mature technologies:

1. Agricultural sensor technology. Agricultural products have been covered by many categories of sensors such as soil sensors, water sensors, meteorological sensors, heavy metal detection sensors, biosensors, gas sensors, and so on [7].
2. Radio transmission technology. ZigBee wireless sensor networks can achieve self-organizing wireless data transmission, which has been widely applied in large-scale farming.
3. RFID (Radio-frequency identification) technology. It is widely used in animal identification, which can be performed on animals to achieve intelligent recognizing, positioning, tracking, monitoring, traceability, and management.
4. Technical quality safety of agricultural products. In the agricultural "production circulation sales" industrial chain, recording and monitoring the chain can realize the whole process of regulation [7].
5. Intelligent irrigation technology. Relying on satellite positioning network and "shallow wells underground cables + field + automatic irrigation system pipe" technology, it can collect irrigation water, electricity, irrigation, and time data to achieve automation of farmland irrigation and through a comprehensive analysis of information technology software to guide irrigation.
6. Precision seeding and spraying techniques. Relying on technology combined with GPS navigation technology, variable rate fertilization, and seeding technology, it can achieve uniform implementation of the planting, spraying, and improving the utilization of seeds, pesticides, and so on.

The Table 1 below presents the studies conducted on IOT based agricultural system in different agricultural parameters and potential application areas.

Table 1. Studies conducted on IoT based agricultural system

Authors	Parameters	IoT Technology	Protocol	Application areas
Abd El-kader et al. [8]	Potato crop	Agricultural sensor technology	APTEEN protocol	Plant monitoring (precision farming)
Mohanraj et al. [3]	Plant growth and Irrigation Planner	TI CC3200 Launchpad interconnected sensors and KM-Knowledge base	Not specified	Field Monitoring and Automation
Nikolidakis et al. [9]	Irrigation	Agricultural sensor technology	Equalized Cluster Head Election Routing Protocol	Irrigation control
Li et al. [10]	Apple orchards	Agricultural sensor and GPRS technology	ZigBee	Air monitoring
Chen et al. [11]	Soil temperature and moisture	Agricultural sensor and GPRS technology	ZigBee	Soil monitoring
Yue et al. [12]	pH, oxygen density and turbidity	Wireless sensor network technology	IEEE 802.15.4	Water monitoring
Fourati et al. [13]	Humidity, solar radiation, temperature, and rain	Wireless sensor network technology, GPS and GIS	Zigbee, GSM and RS-232	Plant monitoring
Ehsan et al. [14]	Wildlife tracking	Delay-tolerant networks (DTNs)	Not specified	Animal monitoring
Jain et al. [15]	Behaviour and migration	Wireless sensor network technology and GPS	Zigbee	Animal monitoring
Khriji et al. [16]	Water need for plant	Wireless sensor network technology and GPS	IEEE 802.15.4	Irrigation control
Pahuja et al. [17]	Greenhouse climate	Wireless sensor network technology	ZigBee based on IEEE 802.15.4	Fertilizer and pesticide control
Luvisi et al. [18]	Vineyards	RFID	Not specified	Production monitoring
Vellidis et al. [19]	Cotton	Wireless sensor network technology and Radio Frequency Identification	RFID tags	Production monitoring
Papetti et al. [20]	Cheese	Radio Frequency Identification	Not specified	Sales monitoring and traceability
Wang et al. [21]	Meat	Radio Frequency Identification and GIS system	Not specified	Sales monitoring and traceability
Abad et al. [22]	Fish	Radio Frequency Identification sensors	Not specified	Circulation monitoring
Amador et al. [23]	Pineapples	Radio Frequency Identification	RFID tags	Circulation monitoring
Kim [24]	Crop diseases	Cloud-based and Wireless sensor network technology	Backhaul IPv4	Smart farming

3 Potential Applications of IoT in Agriculture

Internet of Things can be incorporated in different environmental applications in order to help human's day to day activities. To produce dense and real-time maps of air and water pollution, noise level, temperature and harmful radiation among others IoT technology is the important and emerging technology [6]. Now a days, agricultural technology and precision farming have arisen as new scientific fields that use data intensive approaches to drive agricultural productivity while minimizing its environmental impact [25]. IoT has been used in several application areas; one of the dominant domains is agriculture. Zhao et al. [26] stated that IoT is dominant used in greenhouse technology to monitor production environment and control system; temperature, humidity, and soil signals are collected in real-time in the production process using wireless sensor networks through machine-to-machine support system. The application of IoT in Agriculture is versatile. In the subsequent section we present the major areas that IoT can be used in the domain.

(a) **Smart Farming**

Farming is one of the potential application areas of IoT. The adoption of IoT for of smart technology includes using sensors, robots, systems of control, and autonomous vehicles [27, 28]. IoT farming also offers farmers with appropriate pesticides, climate information, soil information (acidity, moisture, nitrogen level etc.), fertilizer information, crop disease management, and weather forecast. Precision farming [28] is one of the most famous applications of IoT in agriculture sector. For example, Crop Metrics is a precision agriculture organization working on the management of precision irrigation to monitor the soil moisture and to increase water use efficiency in order to maximize yields [28].

(b) **Green Housing**

Light, temperature and humidity determine productivity and plant growth [29]. Continuous monitoring of these environmental variables provides valuable information to the farmer to know the effect of each factor and how to increase crop productivity [29]. Adjusting the greenhouse climate enables the farmer to improve productivity. Several studies have been conducted on IoT application to monitor greenhouse environment using wireless sensor technologies.

(c) **Livestock Management/Monitoring**

There are enormous number of publications on animal monitoring using ICT. IoT applications can be used to collect data regarding the location, well-being, and health of their cattle For example, some of the applications of IoT utilized in livestock management monitoring are the study of animal behavior monitoring (ZebraNet in Kenya), the behavior analysis of sheep within vineyards based on IoT technology [30], and the animal posture behavior [31]. Williams et al. [32] studied the grazing behavior of cows using GPS. In their study, they classified the behavior into three states: grazing, walking and resting. Another IoT intervention area is estrus detection [32, 33]. Missed inseminations causes additional feed costs for fertile cows, artificial insemination costs, and reduce production of calves and milk yield. Kim et al., used bio-sensor that measures the body temperature of cows in real time that allows accurate detection of estrus. The biosensor is inserted orally into the cow; and transmit biometric information is forward to data gathering box and the data is transmitted to the server using cellular or WiFi technology for data analysis. The server notifies the status of estrus detection (whether the cow is in estrus cycle or not) to the manager or veterinarian through a mobile application. Several related studies have been conducted to detect estrus [34, 35].

Generally, IoT can be applied in every fields of Agriculture. It has immense potential to increase productivity and in general it is a means to foster socio-economic development.

4 Architecture of IoT in Agriculture

Based on the review of different research work on smart agriculture and its analysis on IoT's critical technologies in smart agriculture. In the general framework of IoT based agriculture, we shall ensure the size of IoT, mobility sensor communication and its data security. The application of IoT is still relatively very less in Ethiopia, and there is no large-scale, systematic development trend, so a scalable and open architecture of IoT should be established in the agricultural industrial chain technology as well as application and formation system in ideal aspects in order to break through the barriers of application in large-scale and promote the IoT industry's transition from the start-up period to the growth period. The overall general architecture of IoT includes data acquisition layer, information exchange layer and application layer. The overall technical architecture of IoT is shown in Fig. 1.

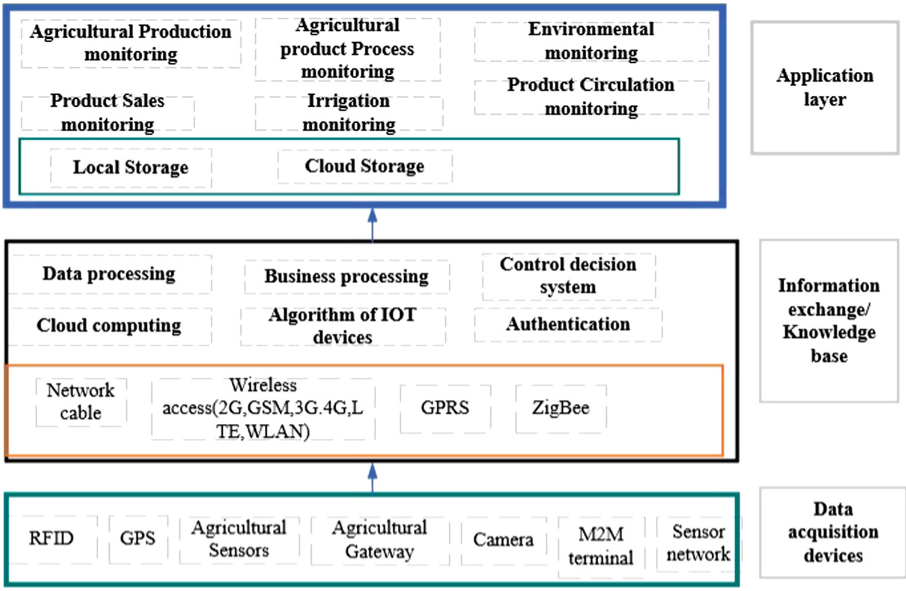


Fig. 1. The general architecture of IoT

4.1 IoT Based Monitoring System in Agriculture

In the process of agricultural production, the most important thing is the real time data collection in terms of data acquisition sensors like temperature, moisture, carbon dioxide content, and soil temperature and soil moisture content and etc. Monitoring System in agricultural based on IoT aims at the target of making information collection towards several indexes in agricultural production, process, sales, environmental monitoring, irrigation and in product circulation and carrying out systematic monitoring towards the plantation area, crop pattern, crop growing, the breaking out and development of agricultural damages, crop output and so on. The monitoring system is composed by the wireless sensor monitoring network and distant monitoring information system. The Fig. 2 describe a typical IoT deployed for agricultural applications.

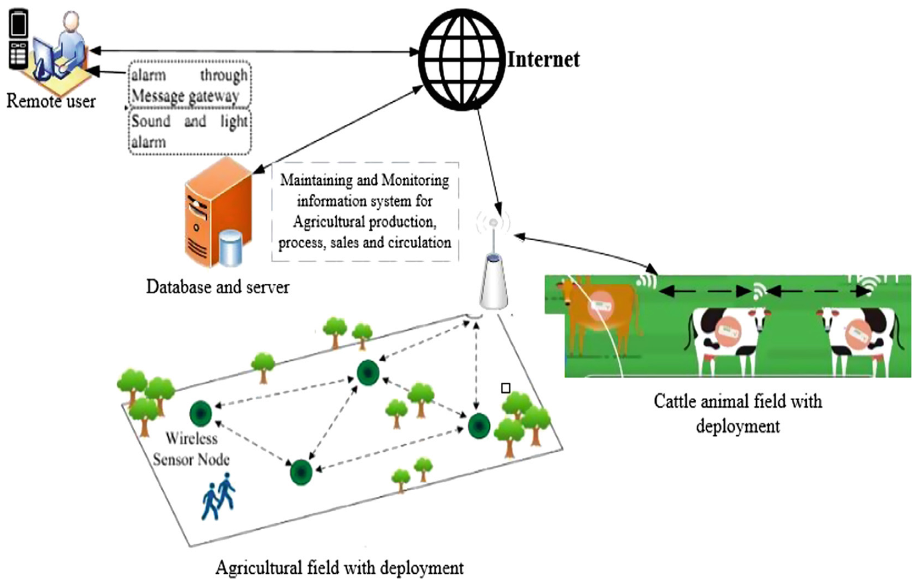


Fig. 2. IOT deployed for agricultural applications

5 Challenges of Using IoT in Agriculture

As we have discussed in section two, IoT is the network of physical objects, devices, vehicles, buildings and other items which are embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data. Implementing this concept in agricultural field is not an easy task. These is may be for many reasons including the complex nature of the different components of the ecosystem of IoT. The challenges of IoT in agriculture limit to meet the needs of the future along with efficient use of resources in Ethiopia. There are considerations to adopt IoT technologies in agriculture of Ethiopia. In this regards the popularity of any technology is depends on the ease of its adoption and implementation [36]. According to [37] studied there are some challenges to implement the application of IoT required high cost than expectation that the technology must available at low cost with a large number of objects. Many other challenges are also faced in order to adopt IoT application for developing countries, in these regards for Ethiopia because of different factors. Challenges of IoT like scalability, self-organizing, data volumes, data interpretation, interoperability, automatic discovery, software complexity, fault tolerance, power supply, wireless communication, security and privacy are discussed [37–39] studies. For instance: complexity of software and hardware are some of the challenges associated with the IoT in agriculture for developing countries like Ethiopia. Minimal resources and software infrastructures are needed in smart agriculture. The study by [40] examined that the appropriate programming tools and low-power capabilities are mandatory, since the frequent battery replacement of the stations. This is not easy to deploy in large scale open field. In the study power harvesting was mentioned as a solution to some extent, however, the power consumption has still to be within the power budget of small power harvesting modules like solar panels, wind turbines etc. According to [40, 41] studied the large number of interconnected devices produces an incredibly large amount of data, which will soon be beyond the resource capacities of small-scale server infrastructures to handle. The other is challenges of security, the IoT technology is an interconnected smart device with more devices together. This make the negative impact on smart agriculture because of it provide the decentralized energy points for creating malware. In this regards it become the main challenge of IoT in agricultural sector. Security and privacy is the major issue for the prevention of widely adopt IoT in agriculture [42]. The three requirements of security are authentication, confidentiality and access control. The architecture of security has four layers. Based on these layers the IoT requirements are to insure security in the whole system of smart agriculture. According to [40] the most security issue include information acquisition security and physical security of hardware in perception layer of security architecture in IoT technology. Another important issue is the requirement of technical skill, in smart agricultural sector capacity building is the essential task. For farming and create an impact on the overall performance in field of agriculture technical skill is required. In order to enable IoT applications in agricultural field like in Precision Farming, Livestock Monitoring, Agricultural Drones and others, IoT technician or skilled individuals on different smart devices are required for farmers to maximize yields using minimal resources such as water, fertilizer and seeds.

6 Conclusion and Future Direction

The introduction of IoT technology is featured to be useful for advancing the agricultural process and animal monitoring process by introducing new dimensions. Ethiopia is an agrarian country where 80% of the population are living in rural area and their income is from Agriculture. However, farming in Ethiopia is very traditional and technology intervention is very minimal. Therefore, IoT can support to modernize Ethiopian agriculture System. There is huge amount of farmland; and the environment is very conducive for much of the crop to grow and cultivate. Besides, Ethiopia ranked first in livestock in Africa and fifth in the World [43]. But, the cattle management, disease monitoring, diary, beef, meat production, etc. trends is very traditional and not reliable. Thus, IoT has a huge potential to boost production in livestock sector.

In agriculture sector, IoT technologies appear in different application areas like precision agriculture or smart farming, irrigation monitoring, environmental monitoring, livestock monitoring and in all other fields of agriculture. Through review on the key technologies of IoT, and the applications in monitoring agricultural products during production, processing, circulation, sales of the whole supply chain and traceability, it showed that consumers can acquire agricultural products information for risk management and traceability, and producers can get high quality and yield products with minimum effort in agriculture based on IoT technology. Nowadays, RFID, WSN, sensors, GPS technology and LPWAN (low power wide area network) protocols such as LoRa, SigFox and NBIOT had been widely used in smart farming and greenhouse management and environmental monitoring, irrigation controlling, animal monitoring cold chain monitoring and tracking, and traceability, etc. This technology made the monitoring of agricultural product processing easier, automatic, effective and real-time, which greatly improved agricultural industrialization. Although there exist challenges in IoT applications, it has many benefits and it has some potential in increasing productivity and alleviate poverty. Some of the challenges IoT faces are: scalability, self-organizing, data volumes, data interpretation, interoperability, automatic discovery, software complexity, fault tolerance, power supply, wireless communication, security and privacy. At present, the demand for IoT is increasing rapidly, and its application in agricultural product processing is a new attempt for Ethiopia. The future works mainly focus on the following aspects:

1. Agricultural industrialization is one of the targets for Ethiopian Government and developing IoT-based framework model is the future work for each application in agricultural sectors.
2. Besides, adopting and developing full-fledged IoT based system which support agricultural industrialization in Ethiopia is also another future work.

References

1. Gebre-Selassie, A., Bekele, T.: A review of Ethiopian agriculture: roles, policy and small-scale farming systems. Ethiopia and DR Congo (2010)
2. Ethiopia Institute of Agricultural Research, Agricultural Mechanization Research. http://www.eiar.gov.et/index.php/agricultural-mechanization-research. Accessed 20 Feb 2019
3. Mohanraj, I., Ashokumar, K., Naren, J.: Field monitoring and automation using IOT in agriculture domain. Procedia Comput. Sci. **93**, 931–939 (2016)
4. Harivallabha, V. M., Anithamahalakshmi, I., Senthamizh Selvi, S.: A model to improve the reliability of ICT in agriculture. In: 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (2016)
5. World Bank. ICT in Agriculture (Updated Edition): Connecting Smallholders to Knowledge, Networks, and Institutions. The World Bank (2017)
6. Talavera, J.M., et al.: Review of IoT applications in agro-industrial and environmental fields. Comput. Electron. Agric. **142**, 283–297 (2017)

7. Huang, B., Yao, Y.: Proceedings of the 5th International Conference on Electrical Engineering and Automatic Control. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-48768-6
8. Abd El-kader, S.M., Mohammad El-Basioni, B.M.: Precision farming solution in Egypt using the wireless sensor network technology. Egypt. Inform. J **14**(3), 221–233 (2013)
9. Nikolidakis, S.A., Kandris, D., Vergados, D.D., Douligeris, C.: Energy efficient automated control of irrigation in agriculture by using wireless sensor networks. Comput. Electron. Agric **113**, 154–163 (2015)
10. Li, D., Liu, Y., Chen, Y.: Computer and Computing Technologies in Agriculture IV: 4th IFIP TC 12 Conference, CCTA 2010, Nanchang, China, October 22–25, 2010, Selected Papers. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-18333-1
11. Chen, K., Zhang, H., Wu, T., Hu, J., Zhai, C., Wang, D.: Design of monitoring system for multilayer soil temperature and moisture based on WSN. In: 2014 International Conference on Wireless Communication and Sensor Network, pp. 425–430 (2014)
12. Yue, R., Ying, T.: A novel water quality monitoring system based on solar power supply & wireless sensor network. Procedia Environ. Sci. **12**, 265–272 (2012)
13. Fourati, M.A., Chebbi, W., Kamoun, A.: Development of a web-based weather station for irrigation scheduling. In: 2014 Third IEEE International Colloquium in Information Science and Technology (CIST) (2014)
14. Ehsan, S., et al.: Design and analysis of delay-tolerant sensor networks for monitoring and tracking free-roaming animals. IEEE Trans. Wirel. Commun. **11**(3), 1220–1227 (2012)
15. Jain, V.R., Bagree, R., Kumar, A., Ranjan, P.: wildCENSE: GPS based animal tracking system. In: 2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (2008)
16. Khriji, S., Kanoun, O., El Houssaini, D., Viehweger, C., Jmal, M.W., Abid, M.: Precision irrigation based on wireless sensor network. IET Sci. Meas. Technol. **8**(3), 98–106 (2014)
17. Pahuja, R., Verma, H.K., Uddin, M.: A wireless sensor network for greenhouse climate control. IEEE Pervasive Comput. **12**(2), 49–58 (2013)
18. Luvisi, A., Triolo, E., Rinaldelli, E., Bandinelli, R., Pagano, M., Gini, B.: Radiofrequency applications in grapevine: from vineyard to web. Comput. Electron. Agric. **70**(1), 256–259 (2010)
19. Vellidis, G., Tucker, M., Perry, C., Kvien, C., Bednarz, C.: A real-time wireless smart sensor array for scheduling irrigation. Comput. Electron. Agric. **61**(1), 44–50 (2008)
20. Papetti, P., Costa, C., Antonucci, F., Figorilli, S., Solaini, S., Menesatti, P.: A RFID web-based infotracing system for the artisanal Italian cheese quality traceability. Food Control **27**(1), 234–241 (2012)
21. Wang, Z., Ye, T., Xiong, A.: Research of food traceability technology based on the internet of things name service. In: 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 100–106 (2016)
22. Abad, E., et al.: RFID smart tag for traceability and cold chain monitoring of foods: demonstration in an intercontinental fresh fish logistic chain. J. Food Eng. **93**(4), 394–399 (2009)
23. Amador, C., Emond, J.-P., do Nascimento Nunes, M.C.: Application of RFID technologies in the temperature mapping of the pineapple supply chain. Sens. Instrum. Food Qual. Saf. **3**(1), 26–33 (2009)
24. Kim, S., Lee, M., Shin, C.: IoT-based strawberry disease prediction system for smart farming. Sensors **18**(11), 4051 (2018)

25. Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: a review. Sensors **18**(8), 2674 (2018)
26. Zhao, J.-C., Zhang, J.-F., Feng, Y., Guo, J.-X.: The study and application of the IOT technology in agriculture. In: 2010 3rd International Conference on Computer Science and Information Technology (2010)
27. Stokes, P.: IoT applications in agriculture: the potential of smart farming on the current stage. Medium, 02 October 2018. https://medium.com/datadriveninvestor/iot-applications-in-agriculture-the-potential-of-smart-farmingon-the-current-stage-275066f946d8. Accessed 05 Feb 2019
28. IoT Applications in Agriculture—IoT For All. IoT For All, 03 January 2018. https://www.iotforall.com/iot-applications-in-agriculture/. Accessed 05 Feb 2019
29. Akkaş, M.A., Sokullu, R.: An IoT-based greenhouse monitoring system with Micaz motes. Procedia Comput. Sci. **113**, 603–608 (2017)
30. Nobrega, L., Tavares, A., Cardoso, A., Goncalves, P.: Animal monitoring based on IoT technologies. In: 2018 IoT Vertical and Topical Summit on Agriculture - Tuscany (IOT Tuscany) (2018)
31. Umstätter, C., Waterhouse, A., Holland, J.P.: An automated sensor-based method of simple behavioural classification of sheep in extensive systems. Comput. Electron. Agric. **64**(1), 19–26 (2008)
32. Williams, M.L., Mac Parthaláin, N., Brewer, P., James, W.P.J., Rose, M.T.: A novel behavioral model of the pasture-based dairy cow from GPS data using data mining and machine learning techniques. J. Dairy Sci. **99**(3), 2063–2075 (2016)
33. Kim, H., Oh, S., Ahn, S., Choi, B.: Real-time temperature monitoring to enhance estrus detection in cattle utilizing ingestible bio-sensors: method & case studies. J. Korean Inst. Inf. Technol. **15**(11), 65–75 (2017)
34. In, K., Lee, J., Xu, Z., Park, D., Chung, Y., Chang, H.-H.: Real-time vocalization aquisition and oestrus detection of Korean native cows. J. Korean Inst. Inf. Technol. **13**(12), 123 (2015)
35. Zubair, M., Hartmann, K., Loffeld, O.: Design and implementation of a wireless geophone sensor node for target detection and classification. In: Automatic Target Recognition XXII (2012)
36. Kaur, K.: The agriculture internet of things: a review of the concepts and implications of implementation. Int. J. Recent Trends Eng. Res. (IJRTER) **2**(04), 2–4 (2016)
37. Zeinab, K.A.M., Elmustafa, S.A.A.: Internet of things applications, challenges and related future technologies. World Sci. News **2**(67), 126–148 (2017)
38. Davies, R.: The Internet of Things opportunities and challenges. European Parliamentary Research Service (2015)
39. Mattern, F., Floerkemeier, C.: From the internet of computers to the internet of things. In: Sachs, K., Petrov, I., Guerrero, P. (eds.) From Active Data Management to Event-Based Systems and More. LNCS, vol. 6462, pp. 242–259. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17226-7_15
40. Tzounis, A., Katsoulas, N., Bartzanas, T., Kittas, C.: Internet of Things in agriculture, recent advances and future challenges. Biosyst. Eng. **164**, 31–48 (2017)
41. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
42. Kumar, S.A., Vealey, T., Srivastava, H.: Security in internet of things: challenges, solutions and future directions. In: 2016 49th Hawaii International Conference on System Sciences (HICSS) (2016)
43. Cook, R.: World Cattle Inventory: Ranking of countries (FAO). Drovers. https://www.drovers.com/article/world-cattle-inventory-ranking-countries-fao. Accessed 19 Feb 2019

Improvement of LEACH Protocol for Wireless Sensor Networks

Atallo Kassaw Takele[1(✉)], Towfik Jemal Ali[2], and Kebebew Ababu Yetayih[2]

[1] Assosa University, Assosa, Ethiopia
atallo2116@gmail.com
[2] Jimma University, Jimma, Ethiopia
towfikjemal@gmail.com, kebebewss2@yahoo.com

Abstract. Wireless sensor networks are aggregates of large number of small, cheap, self-configured and low power sensor nodes. These nodes will sense, and gather various types of information from the environment or system and transmit the data to the base station for monitoring a system or an environment. However, they are characterized by limited resources (like battery, CPU and memory storage) because, they are small in size. Among those battery is a crucial issue, since they are deployed in remote locations. Most battery is consumed when a node sends data and when it routes where to send the data. Hence, energy efficient routing protocols have received significant research attention. This paper focuses on modifying LEACH protocol, which is a well-known energy efficient protocol for minimizing energy and overheads. On the first-round clusters will be formed as in the original LEACH protocol, then nodes will be scheduled for serving as a cluster head on each round turn by turn sequentially. The scheduled node will automatically be cluster head until two-third of its total energy is consumed. After two-third of the total energy has elapsed, re-clustering will be held in order to remove low powered nodes and dead nodes from the schedule. If the scheduled node dies before it serves as a cluster head, the successor scheduled node will serve. The proposed work has been evaluated comparatively using Castalia 3.3 simulator and it out performs the existing LEACH protocol.

Keywords: WSN · LEACH · Cluster-head selection · WSN routing · Scheduling

1 Introduction

In recent advances with micro electro-mechanical systems technology, wireless communications, and digital electronics, the design and development of low-cost, low-power, multi-functional sensor nodes that are small in size and communicate untethered in short distances have become feasible [1]. The ever-increasing capabilities of these tiny sensor nodes, which include sensing, data processing, and communicating, enable the realization of wireless sensor networks (WSNs) based on the collaborative effort of a large number of sensor nodes. Wireless sensor networks (WSN) are a collection of various nodes which consist of sensors, analog to digital converter (ADC),

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 250–259, 2019.
https://doi.org/10.1007/978-3-030-26630-1_21

micro-controller, memory, radio electronics and antenna, battery and dc-to-dc converter [2]. They are distributed embedded system consisting of large collection of short range collectively work together to monitor a system or an environment. These nodes will sense, and gather information from the environment or system and transmit the data to Base Station.

They are commonly used for military reconnaissance, vehicular movements, volcanic earthquake timing, weather forecasting, environment monitoring like temperature, humidity, pressure, motion, health care applications, etc.

As presented above wireless sensor networks are widely applicable in several areas in our lives, but they have limited resource such as battery, processing capacity, bandwidth and memory [8, 9]. Among those energy is an important resource in order to maximize the lifetime of the network. The replacement of batteries in the nodes is infeasible as most of WSN nodes are deployed normally in hostile and remote locations. Therefore, the node's battery energy drain should be considered while designing a routing protocol since routing process between nodes consumes a lot of energy which determines the network's lifetime [3, 4].

More work has been done to design energy efficient routing protocol for wireless sensor networks such as, PEGASIS [11], TEEN [6], LEACH [5], etc. But, still they have to be improved so as to make them more energy optimized and customized in specific application scenarios. This paper focus on the improvement of LEACH protocol. LEACH is a cluster based energy efficient routing protocol in which cluster head is elected and cluster are formed in the first phase and actual data communication is performed in the second phase in each round [5]. In this case selecting cluster heads and forming clusters in each round consumes time and energy. To minimize this overhead this paper applies a scheduling technique. On the first-round clusters will be formed as in the original LEACH protocol, then nodes will be scheduled for serving as a cluster head on each round turn by turn sequentially until two-third of the total energy will be elapsed.

The rest of the paper is organized as follows. Section 2 states about related works. Section 3 describe about the proposed work. Then Sect. 4 talks the simulation and evaluation of the proposed work and the last section concludes the proposed work.

2 Related Work

Low Energy Adaptive Clustering Hierarchy (LEACH) protocol uses hierarchical topology with which the data is aggregated and the same is sent to the BS. It is self-adaptive and self-organized. The nodes are divided into groups or clusters with uniform size, and a CH is elected for every cluster. The nodes collect physical phenomena related data and send the same to their respective CHs using TDMA schedule. If a node continues to remain as a CH, its energy is drained and the node dies sooner. In order to avoid such a situation, the CH role is rotated after each round of operation. A node forwards the data to its CH and the CH gathers the data from each node supported by it and the CH computes the aggregate of the collected data and the same is forwarded to the BS [5].

Much improvement has been done to overcome the drawbacks of LEACH protocol. Some works are done to make it multi-hop by considering the drawbacks of distance from the base station to the cluster-heads and others are done on single-hop as it is by considering other parameters like energy, way of cluster-head election, etc. Some of them are summarized as follows;

Authors in [8], proposed another improvement on LEACH by introducing an energy efficient Cluster Head (CH) selection considering the energy and distance as parameters. The probability function is modified with the energy and distance metrics to choose energy efficient CH for data transmission. In addition to probability function score function is introduced which is calculated from distance between the base station and the sensor nodes and residual energy. This paper tries to consider the two important parameters (distance and residual energy) but sensor nodes will have more overhead in order to calculate and exchange distance and residual energy for each round.

Authors in [9], tries to minimize advertisement packets of LEACH protocol to form the cluster in each round. But in this approach sensor nodes will send or receive the control packets when it changes the state. This approach considers the energy thresh-hold of the cluster-head to inform the cluster-head, it will no longer be used, initial formation of the cluster to elect the cluster-head, joining of sensor nodes through handshaking and sensor node leaving the cluster. But, if much number of cluster-head have frequent thresh-hold messages there will be frequent formation of cluster hence it will consume extra resource.

Authors in [10], proposed IL-LEACH to suppress the correlated data transmission inside the cluster. This approach assumes that we know location of the sensor nodes. The solution focuses to reduce the data transmissions based on grouping of sensor nodes into a virtual correlated cluster (VCC). The grouping is done based on calculation of a threshold sensing coverage (TSC) and Euclidean distance. Then, on the basis of residual energy of each sensor node, one of the nodes is selected to transmit data from the group. In LEACH the nodes send data based on TDMA schedule made by the cluster head. In this method IL-LEACH all the nodes that are selected to be in a same VCC (virtual correlated cluster) are given same TDMA so that only one of them sends their data to the cluster head.

Gambhir and Parul in [7], proposed OE-LEACH that aims to minimize the energy west during idle listening state. There are many situations where the sensor nodes do not have data to send regularly as they may be event driven. In that case, data are available only when they sense the event. So, the sensor nodes need not to listen the channel at all times. Energy is wasted more on idle listening than that dissipated in transmitting and receiving. Slots are not allocated to the nodes which have no data to send and free slots are converted into useful slots. This will reduce the idle listening problem. This method will decrease the waiting time for nodes as now sensor nodes can get more than one slot in a frame which means throughput of the network increases. But it does not consider the residual energy while selecting cluster heads.

Authors in [4], proposed an improved multipath leach protocol which uses at most one intermediate cluster head. LEACH sends data from the Cluster-Head to the sink directly. Thus, the necessary energy for transmitting data is high. Even more, if the sink is too far the transmission is impossible and the data is lost. The aim of this protocol is to extend the lifetime of the network and send more data compared to the original protocol. It uses only one intermediate cluster head hence it is not scalable for large networks.

In this we have reviewed related works that have been done by different scholars on enhancing LEACH protocol, to conserve energy consumption of wireless sensor networks. Some of them considers parameters like residual energy and distance, minimization of advertisement packets, scheduling densely deployed nodes by turning each node on and off alternatively and optimal cluster head selection. Others made an enhancement on shifting from single hop to multiple hop approach. But, none of them consider minimizing overhead by eliminating unnecessary advertisement messages. The proposed work focuses on improving the battery consumption and delays while nodes select cluster heads and form clusters on each round by scheduling cluster heads for avoiding bulk re-clustering advertisement messages.

3 Proposed Approach

In this paper, we have proposed a modified LEACH routing algorithm in order to minimize energy consumption. The new approach tries to minimize the over-head of selecting cluster head and cluster formation on each round. On the first round, clusters will be formed and nodes will be scheduled to become a cluster head for the next rounds. The first cluster head is elected randomly like the original LEACH and non-cluster heads join the cluster head in order to form the cluster. After the cluster is fully formed, the cluster head schedules cluster heads for the next rounds.

Once the first cluster head schedules the next round cluster heads, the scheduled node will automatically be cluster head until two-third of its total energy is consumed. After two-third of the total energy is elapsed, re-clustering will be held in order to remove low powered nodes and remove dead nodes from the schedule. Then, re-clustering will continue after all members serve as cluster head at least once. Re-clustering process helps to regroup alive nodes and to remove dead nodes from the cluster member cache. Before, re-clustering residual energy of individual nodes will be checked, to decide a cluster-head within at least one next group of rounds. Hence, the cluster-head scheduling removes the over-head of cluster-head advertisement message to get members, join request message to the cluster-head and computation for electing cluster-heads. If the scheduled node dies before it serves as a cluster head, the successor scheduled node will serve.

According to the working process of LEACH, there are several communication patterns for nodes. With those patterns, the energy consumption of nodes can be classified into two parts according to their roles, as demonstrated in Table 1 [12]. In this model broadcast CH message, receive join message from non-CH and send acknowledgement message to non-CH will be eliminated after forming cluster within each group of rounds at the cluster-head. Similarly, receive CH advertisement message and send join request message to CH will also be eliminated at the member nodes. Energy consumption of the proposed model after each group of rounds is demonstrated in Table 2.

Table 1. Energy consumption of nodes in leach algorithm [14].

CH energy consumption	Broadcast CH message
	Receive join message from non CH
	Send acknowledgement message to non-CH (TDMA schedule)
	Receive data from non CH
	Send aggregated data to BS
Non CH energy consumption	Receive CH advertisement message
	Send join request message to CH and receive TDMA schedule
	Send data to CH

Table 2. Energy consumption of nodes in modified leach algorithm after each group of round.

CH energy consumption	Send TDMA schedule to member nodes
	Receive data from member nodes
	Send aggregated data to BS
Non CH energy consumption	Receive TDMA schedule from the CH
	Send data to CH

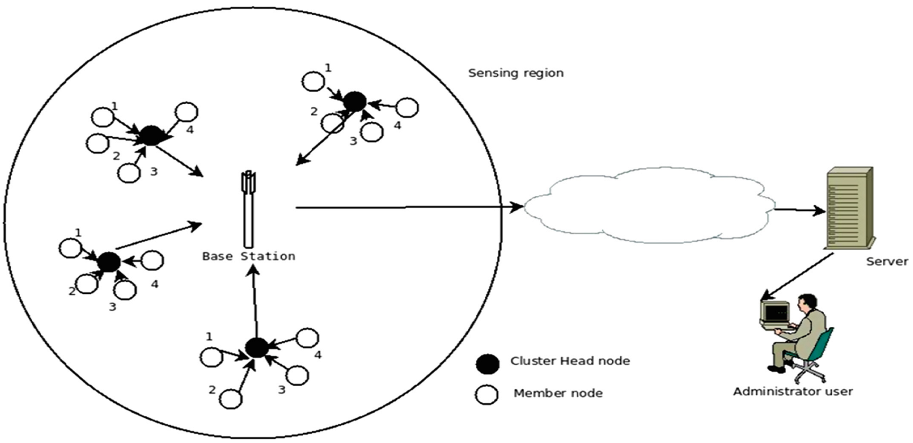


Fig. 1. Topological Architecture of the proposed work

The topology of the proposed work includes member nodes, cluster-head nodes, sink node, server node and an administrator user as shown in the Fig. 1. The sensed data flows from member node to the administrator user. Member nodes send the sensed information to the cluster-head. The cluster-head will aggregate the data that comes from its members, then it will pass to the sink node. Since the sink node is assumed to have enough resource it encapsulates the aggregated data and will send to the main server. The administrator will process the sensed data that comes from the sensor nodes to make a certain useful decision.

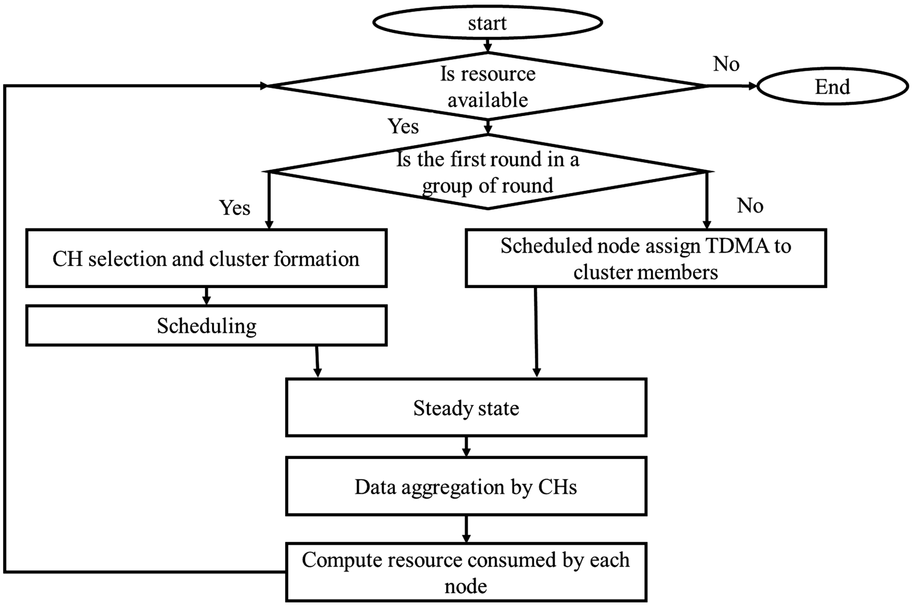


Fig. 2. Flow chart for the whole process in modified model

Figure 2, shows the flow chart of the proposed algorithm. First it checks if enough resource (energy) is available or not. If enough energy is available, it will again check whether it is the first round in a group of round. If it is the first round in a group of round, new cluster will be formed and cluster head scheduling will be held. Group of rounds refers a collection of rounds that all members of the cluster serve as a cluster head at most once and the time from the starting until two third of the total energy consumed is considered as on group of round.

4 Simulation and Evaluation

To evaluate the performance of the proposed model, we have conducted simulation experiment using a simulation tool. Simulator tools are effective tools to evaluate algorithms and protocols at design, development and implementation stages. A number of these tools are available each with different features, characteristics, models and architectures for performance testing in WSNs. Some of these are NS2, OMNeT++ and TOSSIM [13, 14].

Among those we have selected OMNeT++ simulator with Castalia framework for the following reasons. First, Castalia supports networks of low-power embedded devices such as WSNs. Second, it can be utilized to test the distributed algorithms and protocols in realistic radio models and wireless channel. Third, Castalia embraces additional features such as: several popular routing protocols and MAC protocols, a model for temporal variation of path loss, and RSSI calculation, which can provide more convincing and accurate simulation results [13].

4.1 Simulation Result and Discussion

We assume that initial energy for all nodes is identical. All nodes are active as long as they have enough energy to communicate and every node can reach to the sink. We also assume that one third of residual energy is enough for running at least one round. Table 3 presents simulation parameters used in the simulation. In this simulation 100 sensor nodes has been deployed in the surface of 100 × 100 m^2 for about 100 s. As stated in [5], the optimal number of cluster heads in LEACH protocol is 5% that why we make CH probability 0.05. We have executed the simulation for 10 times and the base station is deployed statically at (1,1). For the first scenario, we made 1 configuration and multiple configurations has been done by varying simulation time and number of nodes for the second and third scenarios respectively.

Table 3. Simulation parameter list

No.	Parameter list	Specified
1	Simulation time	100 s
2	Topology size	100 × 100 m^2
3	Number of nodes	100
4	CH probability	0.05
5	Number of trials	10
6	Position of base station	(1,1)
7	Number configuration	1
8	Slot length	0.2 s

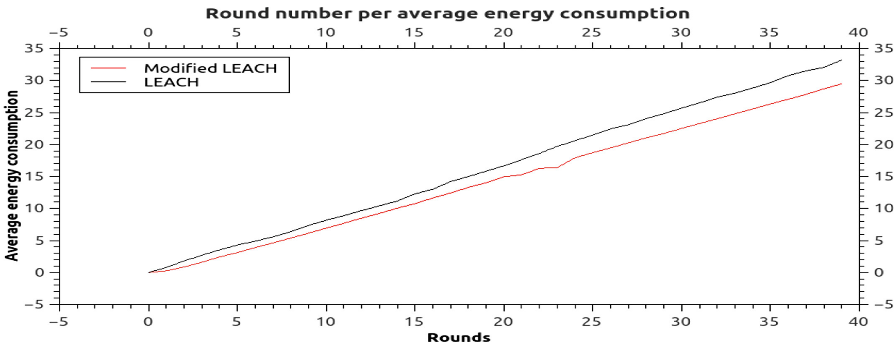


Fig. 3. Energy consumption of the modified and existing protocols on each round

Average consumed energy of the proposed model and the existing model per each round is depicted in Fig. 3. The figure clearly shows that energy consumed per each round in the modified protocol is lower than that of the existing protocol. As we go from the first round to the end the gap between the new and the original protocol increases, which enhances the network life time.

Figure 4, depicts that there is an inclination in average consumed energy with an increase in simulation time in both of the existing and modified protocol. However, the gap between the two graphs increases as the simulation time increases. Hence, as the simulation time increases the average energy consumption of the modified protocol decreases. This will enhance the lifetime of the network.

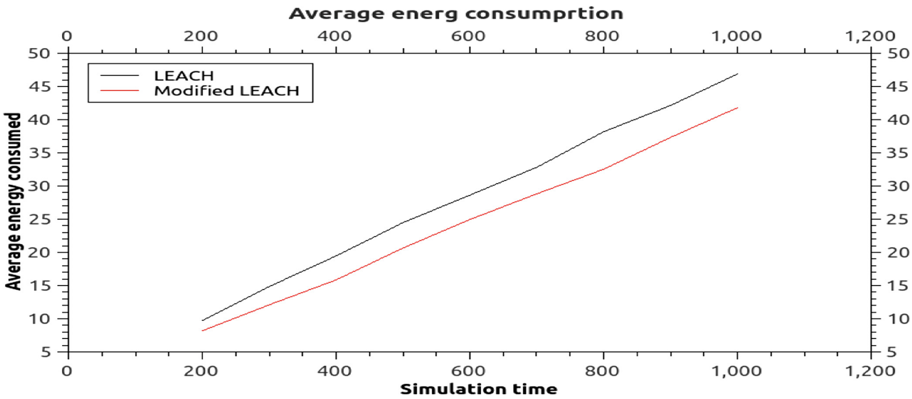


Fig. 4. Simulation result by varying simulation time

Figure 5, shows number of packets received per node to the base station. As the graph shows the proposed model has better performance than the existing model.

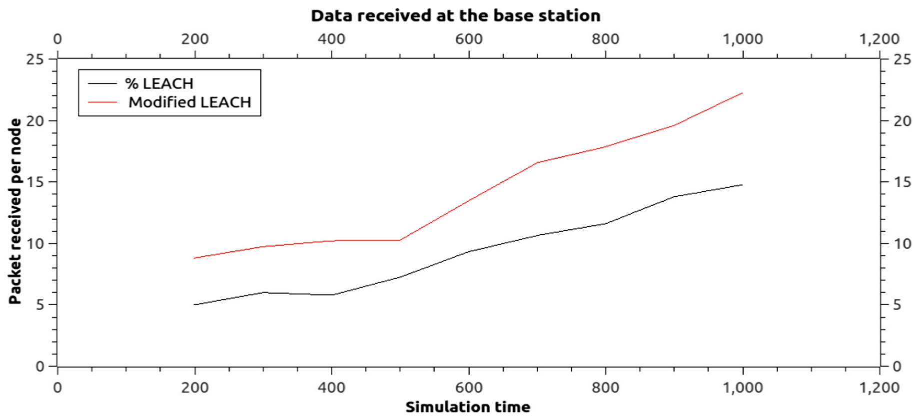


Fig. 5. Packets received per node with varying simulation time

Figure 6, shows the average consumed energy of both the existing and modified protocol as the number of nodes increases. As clearly shown from the graph the average consumed energy of the modified protocol is proportionally lower than that of the existing protocol. Therefore, the proposed protocol is energy efficient within different network sizes.

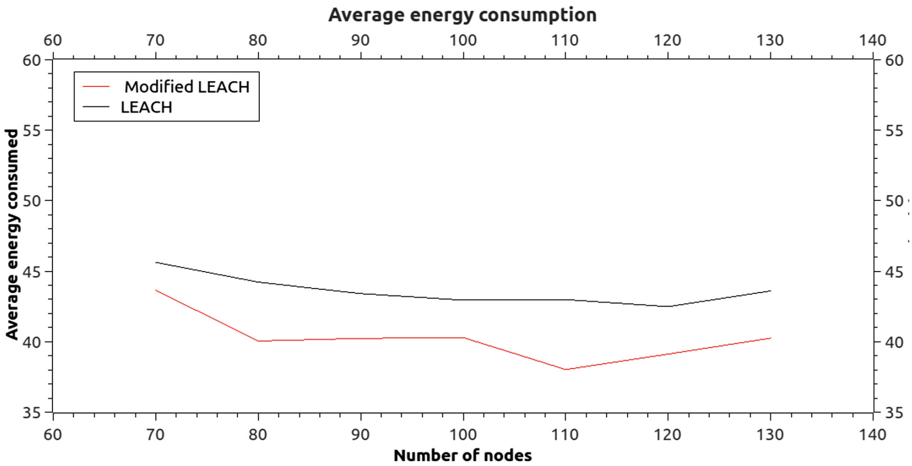


Fig. 6. Simulation result by varying number of nodes

5 Conclusion

Wireless sensor networks are formed from small self-powered devices, consisting of one or more sensors, a microprocessor and one or several radio transceivers. The sensors include an analog-to-digital converter, and the microprocessor can be a microcontroller or a low power processor. The sensor node uses an energy source, usually with very limited capacity. Wireless sensor networks are self-configuring networks since, they are mostly deployed randomly without careful planning.

In wireless sensor networks, energy is an important resource in order to maximize the life time of nodes; they are also characterized by limited CPU and low memory storage. Replacing the battery of the sensor nodes is also difficult since they are mostly deployed in difficult areas such as in forests for sensing fire, in deserts for sensing earthquake, etc. Hence, we have to consider battery consumption while we design protocols and applications to the sensor devices.

This paper puts forward an enhancement on LEACH routing protocol in order to minimize energy consumption and unnecessary over-heads. Modification is applied on the selection of cluster-heads, by scheduling members to be a cluster-head for the next rounds sequentially.

Finally, we have implemented and evaluated the modified protocol using OMNeT++ simulator with Castalia framework. Simulation result shows that the proposed model out performs the existing protocol in terms of energy consumption and the number of packets received by the base station.

References

1. Akyildiz, I.F., Vuran, M.C.: Wireless Sensor Networks. Wiley, Chichester (2010)
2. Kodali, R.K., Kiran A., V.S., Bhandari, S., Boppana, L.: Energy efficient m-level LEACH protocol. IEEE (2015)
3. Sony, C.T., Sangeetha, C.P., Suriyakala, C.D.: Multi-hop LEACH protocol with modified cluster head selection and TDMA schedule for wireless sensor networks. IEEE (2015)
4. Bennani, M., Taj, M., Kbir, M.A.: ICH-LEACH: an enhanced LEACH protocol for wireless sensor network. IEEE (2016)
5. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Published in the Proceedings of the Hawaii International Conference on System Sciences, Maui, Hawaii, 4–7 January 2000 (2000)
6. Manjeshwar, A., Agrawal, D.: TEEN: a protocol for enhanced efficiency in wireless sensor networks. In: Proceedings of the 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, San Francisco (2001)
7. Gambhir, S., Parul: OE-LEACH: an optimized energy efficient LEACH algorithm for WSNs. IEEE (2016)
8. Amirthalingam, K., Anuratha, D.: Improved LEACH: a modified LEACH for wireless sensor network. IEEE (2016)
9. Bharti, A., Devi, C., Bhatia, D.V.: Enhanced energy efficient LEACH (EEE - LEACH) algorithm using MIMO for wireless sensor network. IEEE (2015)
10. Kandpal, R., Singh, R.: Improving lifetime of wireless sensor networks by mitigating correlated data using LEACH protocol. IEEE (2016)
11. Lindsey, S., Raghavendra, C.: PEGASIS: power efficient gathering in sensor information systems. In: Proceedings of the IEEE Aerospace Conference, Big Sky, Montana (2002)
12. Luo, F., Jiang, C., Zhang, H., Wang, X., Zhang, L., Ren, Y.: Node energy consumption analysis in wireless sensor networks. IEEE (2014)
13. Chhimwal, P., Rai, D.S., Rawat, D.: Comparison between different wireless sensor simulation tools. IOSR-JECE **5**, 54–60 (2013)
14. Korkalainen, M., Sallinen, M., Kärkkäinen, N., Tukeva, P.: Survey of wireless sensor networks simulation tools for demanding applications. IEEE (2009)

Towards a Design Research Model for Climate Services: Experiences from a Development Project in Mozambique

Jaakko Helminen[1(✉)], António José Leão[2], Ville Myllynpää[1], Silvia Gaiani[1], Mikko Apiola[1], and Erkki Sutinen[1]

[1] University of Turku, Turku, Finland
{jaakko.a.helminen,ville.myllynpaa,mvapio,erkki.sutinen}@utu.fi, silviagaiani@yahoo.it
[2] Eduardo Mondlane University, Maputo, Mozambique
ajleao62@gmail.com

Abstract. As one part of our research efforts and exploration in Africa, we have visited the Chokwe Irrigation System area in Mozambique in order to develop hybrid grass-root mobile climate services for and with local farmers to support their cultivation practices. The lessons learned along the explorations inspired us to elaborate an appropriate design research framework for the particular action design research process, which would facilitate us to develop holistic grass-root mobile climate services for local farmers in a developing country. The main outcome of the study was the updated design research framework for the task.

Keywords: Climate Change · Mobile technology · Design Science Research

1 Introduction

With the ongoing climate change and changes in climate variability we encounter quite a variety of climate impacts affecting everyday life, especially in a developing countries. In terms of the design science framework introduced by Hevner et al. [1] as a milestone of the topic we may ask whether we can apply it, when we in a participatory way with local partners co-create climate services appropriate for end-users. Following the design thinking paradigm [2, 3] we started our work with its first stage – empathy within the inspiration space. We acted based on the recommendations of Brown [2], who suggests that the intention during an empathy stage in a project is to experience a problem or opportunity that sets researchers in motion.

We started our project by exploring local climate services and background information in Chókwè District, Mozambique. We had joined a pre-negotiated program, where we had to adjust our tasks to the overall objectives of the program. Still, this arrangement left us with the required freedom to explore pertinent problems as well as opportunities to build an appropriate design research framework.

In the following we briefly describe the situation at the time of our field trip to Chókwè, Mozambique, during August - December 2011. Then we discuss how we processed parts of the interview material and how recent international agreements

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 260–269, 2019.
https://doi.org/10.1007/978-3-030-26630-1_22

[4–8], including agenda for humanity [4], and UN Framework Convention on Climate Change [8], which steered our efforts. Thereafter we show relevant results of the interviews as well as the induced inspirations to tune the Design Science Research framework of Hevner et al. [1] so that we can apply it to develop appropriate hybrid [9] grass-root mobile climate services for local farmers in Chókwè District. In the conclusions we share these inspirations with the readers.

2 Project Consortium

In the context of "Programme of cooperation in Science, Technology and Innovation between Finland and Mozambique" (STIFIMO), we established a joint project "Development of innovative community climate services for agriculture in Chókwè District" (DICCLISEAG) with the Mozambique National Meteorology Institute (INAM) as our local host and together with the District Services of Economic Activities of the Chokwe District Administration (SDAE/CDA), South Zone Center of the Mozambican Institute of Agricultural Research (CZS/IIAM) as well as Instituto Superior Politéchnico de Gaza (ISPG), (Polytechnic Institute of Gaza in English).

3 Trends, Evolutions, and Challenges of the Chókwè Irrigation System and Its Surroundings

The authors of [10–13] discussed the history and present stage of the Chókwè Irrigation System (CIS) and its vicinity. The evolution of events led to a situation, where the government of Mozambique favored large-scale rice cultivation in CIS at the expense of land ownership by local smallholder farmers [11, 12]. In addition, the local historical and social dynamics contributed to gender imbalance towards women. Conflicts between them and the CIS administrator, A Hidráulica do Chokwe, E.P. (HICEP) [10] led to a further weakening of the status of local smallholder farmers.

Chókwè District is a semi-arid region in the flat Limpopo drainage delta area. More upstream there are large elevation differences and also some reservoirs of limited capacity along the route of the Limpopo river. The volume of the reservoirs is not adequate to damp downstream effects of heavy rain episodes in the upper reaches of Limpopo drainage. Hence Chókwè District as part of Limpopo delta area is prone to relatively frequent recurring flash floods, which in the recent past have caused severe disasters, also in CIS.

As climate, its change and the changes in its variability are pertinent factors in the semi-arid region of Chókwè District, it became clear to us that local people, typically smallholder farmers, needed appropriately tailored climate services for actions to improve their baseline and livelihood in a sustainable manner.

In our research, we drew inspiration from the increasing recognition of the social dynamics in climate issues. The World Humanitarian Summit (WHS) 2016 released the Agenda for Humanity (AH) [4] with its five core responsibilities. In his WHS Chief's Summary the then UN Secretary-General Ban Ki-Moon urged the nations to stand up for the humanity and to commit themselves to action [5]. In the context of core responsibility four “Change people’s lives – From delivering aid to ending need” its roundtable five discussion under the theme “Natural Disasters and Climate Change: Managing risks and crises differently” [6] led to five core commitments related to climate issues as follows: implement the related strategies, reinforce national and local management, invest in climate services, build community resilience, and ensure regional and global humanitarian assistance.

4 Data and Methods

4.1 Research Question

Based on the community’s crucial needs for tailored climate services to improve their livelihood in a sustainable manner, we sought to gain baseline information about current needs, strengths, infrastructure, ICT usage patterns, and climate information sources. We set the research question as follows:

What are the current baselines for crucial information needs, current strengths, infrastructure, skillsets, climate information, and mobile usage patterns in Chókwè District?

In our reflection of the results, we sought to find pertinent additional local issues to environment, IS research and knowledge base columns of the design science research framework of Hevner et al. [1] to tune it for the development of holistic grass-root mobile climate services for local smallholder farmers of Chókwè District.

4.2 Questionnaire and Processing Responses

For the general interviews in Chókwè District we used a questionnaire as our data collection protocol. Group interviews were conducted, thus one response in the questionnaire represented one group. After the interviews had been conducted we scrutinized the responses and took into account the parts we considered to be relevant for the task of our research question. The questionnaire consisted of a general part and specific sections covering agriculture, climate and ICT items. The objective of the interview was to find out both strengths and gaps in baseline items pertinent to the hybrid approach of the project.

Based on recommendations of our partners, we initiated the project by making group interviews and conducting surveys in five communities (acronyms in parentheses); Lionde (**L**), Chókwè (**S**) and Chilembene (**B**), all within the CIS area. The remaining communities Macino (**M**) and Chekelane (**A**) were outside of CIS. In the large farm premises of two foreign agribusiness companies Mozfoods Industrias Alimentares (MIA), inside CIS, and EmWest (EW), close to the northwestern end of CIS, we surveyed and discussed their automatic weather stations. In the results, one answer represents one community.

5 Results

5.1 Quality of Life

In regard to quality of life the communities responded on a scale of "excellent", "good", "fair" and "bad". One community considered their life to be "Good" (B), two communities "Fair" (S and M) and two communities "Bad" (L and A). Thus, there were two modes (Mo), "Fair" and "Bad".

5.2 Prominent Main Needs

In regard to the main needs of the communities, the following responses were given. All five communities mentioned the need for farm equipments, including ploughs, tractors and water pumps. Four communities mentioned the need for seeds and electricity (L, S, M and A). Two communities mentioned also the need for water and cows (L and A). Individual communities mentioned access to credit, transport, hospital and channel renovation as their main needs. As a secondary need, three communities mentioned local schools.

5.3 Sources of Help

When asked if any organizations were currently addressing their prominent main needs, only two communities had received help in meeting their needs. Community (L) had got an offer from the agribusiness company Mozfoods Industrias Alimentares (MIA) to provide improved maize seeds, but as a part of that deal they would need to sell the yield to them at their price. In addition, community (B) had received two excavators, which were donated by HICEP, but there is still need for additional excavators and other machinery.

5.4 Evaluation of Infrastructure

In regard to evaluation of infrastructure, on a scale of "Good", "Acceptable" and "Not Good" interviewees gave the following responses. In regards of both *irrigation systems,* as well as *marketing services,* all five communities ranked them as "Not good". Four communities (L, S, M and A) ranked *roads, means of transportation*, and *housing* as "Not good", three communities (L, S, and B) ranked *schools* as "Not good". One community (A) ranked *health centers* as "Not good", while three communities (L, S and B) ranked *health centers* as "Acceptable". Two communities ranked (M and A) *schools* as "Acceptable". *Roads* (B), *housing* (B), *availability of electricity* (B) and *means of transportation* (M) where all ranked as acceptable by one (n = 1) community. Only one community (M) ranked one part of infrastructure as "good", which was the *health centers*. Thus, the communities had a relatively negative perception about most infrastructure items in the questionnaire.

5.5 Expectations for the Future

In regard to expectations for the future, communities' responses where mainly related to the *ability to grow various plants and keep livestock*. One community (L) wants to *cultivate maize and vegetables*, while one community (S) wanted to *cultivate rice* and *larger quantities of vegetables,* such as tomatoes, cabbages, beans and onions. Community B wanted to *develop agricultural practices* by cultivating larger areas and keeping livestock, such as cows, pigs and goats. Another community (M) wanted to develop agricultural practices, cultivate potatoes and tomatoes, keep cows and be able to sell various agricultural products including meat. Community A wanted to cultivate peanuts, small beans, cassava, potatoes and pineapples. In additional all communities reported that they would like to have *better machines, knowledge* and *orange plants*. Also, many of the interviewees in community A affirmed that they would like to start cashew nut production and to get a better and bigger storage facility for processing cashew nuts.

5.6 Malnutrition Situation

When asked if their families had run out of money to buy food in the past 30 days, the following answers where received. Representatives from two communities (L and S) replied that it had happened often (=more than 5 days), whereas representatives from three other communities (B, M and A) replied that it had happened only sometimes (=between 1 to 5 days).

5.7 Cultivation Practices

In regard to cultivation responses of the communities brought up following problems. Two communities (L and S) mentioned that the *irrigation system was not functioning well* and there were *not enough available tractors and other machines for seeding, harvesting and transpor*t. S community also pointed out that *uneven regulation of water channels* could cause issues. Other community (B) pointed out the *lack of fertilizers, poor water quality and the lack of credit* as well as *marketing efforts*, while another community (M) faced *issues with land preparation, plant diseases and problems with wild animals*. One community (A) also complained that *plant diseases were out of control* and there was a *lack of saws to cut trees*.

5.8 Production Constraints

In terms of production constrains, the following responses we given by the communities. Three communities (L, S and A) mentioned *lack of rain*, two of them (L and S) mentioned also *lack of credit*. Two communities (S and M) mentioned *lack of transportation* and *poor market conditions* as well as *very low prices*. Community B referred to their answer for the previous question.

5.9 Suggestions for Changes in Community

In regard to *suggestions for changes in the community*, this was presented as an open-ended question. The responses given by the informants were, in most cases, related to *availability of agriculture inputs*, *equipment* and *construction of various facilities*. Two communities (L and S) proposed to have better *availability of tractors*, *fertilizers*, *seeds* and *finances*. One community (B) suggested to have an *agro-center*, while another (M) proposed to have a co*oling house with needed transportation equipment*. One community (A) proposed to have *a borehole based irrigation system*.

5.10 Past and Present Climate Hazards

When asked how severe the impacts of climate hazards to their crops are, all communities uniformly answered that they are severe. In regard to local responses to climate hazards, all communities again uniformly assessed them to be fair.

5.11 Suggestions to Improve Local Responses

When asked how to improve local responses to climate hazards, communities' responses varied quite much. Two communities (L and S) suggested to *update knowledge on hazards and their occurrence*. One community (B) suggested to *establish their own weather station*, whereas another community (M) proposed that *each community should have its own leader*. Community A anticipated that once their own irrigation system would be established, problems could become more manageable. Another option would be to move farms to more humid areas.

5.12 Access to Climate and Weather Information

In terms of access to seasonal meteorological and agro-meteorological forecasts, three communities (L, S and A) had no idea where to get that information, whereas two communities (B and M) knew that Radio Mozambique broadcasts those information.

This same ignorance applied also to weather and climate services, since all the other communities were unaware of their existence, expect community S, which mentioned radio and TV.

5.13 ICT Ownership

In regard to ownership of ICT technologies, most commonly owned technologies were mobile phones (63 out of 80 responses), radios (9 responses) and TV (7 responses). Both computer and video-TV got one response.

5.14 Most Often Used ICT

Most used technologies in the communities were mobile phones, in addition community S used TV and community S used radio. Community M did not respond to this question.

5.15 Willingness to Contribute to the ICT Development in Your Village

In regard to being interested to contribute to developing/implementing some ICT uses in their village, the following answers were obtained. Community L did not know and community S did not answer. Community M just said "yes", without further elaboration, while community M said that they could help spread the information. Community A said that a local weather station is needed, as well as other equipment's and education.

6 Discussion

In terms of our research question we focus on items pertinent to the tuning of Hevner design research framework [1]. According to the general questions (Sects. 5.1–5.9) the administrative practices of the government steered the small-scale farmers to limit their efforts and co-operation among families, neighbors and, outside of the CIS, also among communities. In these circumstances the smallholder farmers remained poor and underdeveloped, as observed by [14]. The organizations, which addressed these needs to some extent, were MIA and HICEP (Sect. 5.3). For its conditional help in terms of seeds provision MIA kept the right to buy the total crop yield by the price it settled. The farmers of B pointed out their local needs for CIS channel maintenance. The needs for facilities were considerably bigger than the help received from HICEP. Mr. Mugabe of SDAE/CDA said that the whole CIS was in a demand of renovation. The lack of systematic help to smallholder farmers was in line with the preferences of the government to favor large-scale foreign investments and actions. Hence the administrative practices should be included in the environment column of the design research framework.

Another important point inspired by almost all shown responses was the need to establish increasing willingness for open, hybrid, constructive and trans-disciplinary cooperation across the whole society. This aspect should be brought into the environment column of a design model. The surveys and replies in regard to Sects. 5.7–5.9 on farming indicated that the way farming practices had been organized by the big farms (MIA and EW) could give pertinent ideas for the local farmer communities on how to organize their facilities and activities. The answers in Sect. 5.11 spread widely. Regarding Sects. 5.10 and 5.11 one pertinent reaction among farmers was the need to improve their own knowledge of floods and draughts and their occurrence. According to Sect. 5.12 the local smallholder farmers did not know anything about weather and climate services. This indicated lack of knowledge in agricultural practices and in basic climate concepts and led us to replace capacities by capacity building in the environment column.

With the surveyed widespread use of mobile phones it was encouraging to recognize that in communication we could rely on modern devices and networks. In addition (Sect. 5.15) many interviewees indicated their interest to develop and implement ICT uses in their village, although they lacked capacities and know-how.

Many responses indicated that the baseline of livelihood and the abilities to meet impacts of climate hazards were inadequate. To improve the general awareness of the communities we suggest baseline to be included in the knowledge base column. In the context of administrative practices we highly suggest that legislation as well as historical, political and social dynamics are actively kept into consideration.

All in all we could say that the interviewees were aware of the impacts of climate variability in their everyday life in farming issues. However, the awareness of basic climate service concepts and communication skills were inadequate if not non-existent. Hence the lack of knowledge about climate and climate services (Sects. 5.10, 5.11, 5.12) proposes us to emphasize the role of climate knowledge [15] in the context of the baseline in the knowledge base column.

As many of the past flash floods have had severe impacts on the CIS area (Sect. 5.10) proper dimension values e.g. for additional reservoirs upstream from the Limpopo delta area would give a good base for actions to prevent these impacts to occur. Hence also the need of proper dimension values should be emphasized in the context of the baseline in the knowledge base column.

The interviews and assessment of social behavior, like motivation and trust, must be included in the knowledge base column in the design models. With the ongoing climate change and changes in climate variability we used for expert prototype interviews the twelve question set of Critical Systems Heuristics (CSH) [16] and found this approach promising. We suggest heuristics to be part of the knowledge base column.

As we have suggested open, holistic, constructive and trans-disciplinary cooperation across the whole society, we propose synthesis to be added to the justify/evaluate the slot of the IS research column.

With the addition of willingness for an open, hybrid and constructive cooperation, the list of revised and added suggestions to the environment column of the design research framework is as follows:

- capacity building,
- administrative practices as well as
- cooperation.

and to the foundations of the knowledge base column

- baseline as well as
- heuristics.

7 Conclusions

ISPG together with us submitted a proposal to STIFIMO for the project to prepare and discuss the project at the final DICCLISEAG assembly. Due to lack of open, hybrid and constructive cooperation and misunderstandings between local institutes and ministries the proposal was turned down. We were left with our first lesson. Finally, we come to our lesson learned, which can be put in a nutshell as depicted in Fig. 1.

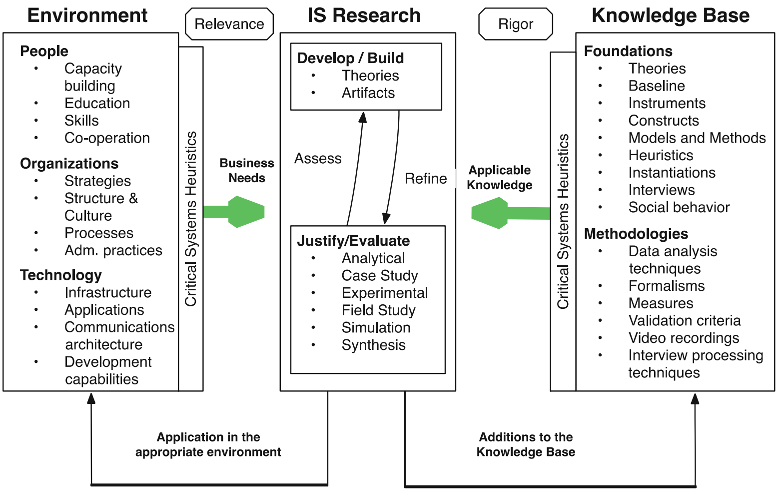


Fig. 1. Elaborated design research framework, based on [1, 16].

Both the environment and knowledge base have in front of them the filter "Critical Systems Heuristics" (CSH) of Ulrich and Reynolds [16]. In essence CSH refers to its twelve question set, through which the suggestions for business needs and applicable knowledge are scrutinized before they enter the IS research column.

Regarding the people in the environment column we have replaced capacities with capacity building not to omit the importance of capacities but to emphasize the needs to replace their lack with capacity building. We have also added co-operation to the essential qualities we expect people to have. As for organizations we found administrative practices to be scrutinized and also followed up.

To the foundations of the knowledge base we added heuristics with the need to take into account increasingly hectic changes in the whole system of systems. We renamed the concept frameworks by baseline in order to widen the scope in system considerations, like climate impacts.

With these updates our first lesson turned out to be quite valuable so that we are now ready to proceed with considerations of our Systems Action Design Research (SADR) in our future projects.

References

1. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
2. Brown, T.: Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. HarperCollins, New York (2009)
3. Both, T.: Design Project Scoping Guide. Hasso Plattner Institute of Design at Stanford, Stanford University, Palo Alto, California, USA (2016). Open Access: https://dschool.stanford.edu/resources/design-project-guide-1

4. Agenda for humanity. Agenda for Humanity Annex to the Report of the Secretary-General for the World Humanitarian Summit (A/70/709, 2 February 2016). United Nations (UN) Office for the Coordination of Humanitarian Affairs, New York, USA (2016). Open Access: https://www.agendaforhumanity.org/agendaforhumanity
5. World Humanitarian Summit. Chair's Summary by the United Nations Secretary-General: Standing up for Humanity: Committing to Action," in World Humanitarian Summit (WHS), Chair's Summary, Istanbul, Turkey (2016a). Open Access: https://www.unocha.org/sites/dms/Documents/Chairs Summary.pdf
6. World Humanitarian Summit. Natural Disasters and Climate Change: Managing Risks & Crises Differently. In: WHS, Key Documents – High-Level Leaders' Round Tables – Core Commitments (2016b). Open Access: http://whsturkey.org/the-summit/key-documents-for-whs
7. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development A/RES/70/1. Resolution adopted by the General Assembly on 25 September 2015 (A/RES/70/1), UN, New York, USA (2015). Open Access: https://sustainabledevelopment.un.org/content/documents/21252030 Agenda for Sustainable Development web.pdf
8. United Nations Framework Convention on Climate Change. Report of the Conference of the Parties on its twenty-first session, held in Paris from 30 November to 13 December 2015. FCCC/CP/2015/10/Add.1, UN, New York, USA (2016). Open Access: https://unfccc.int/process/conferences/pastconferences/paris-climate-change-conference-november-2015/paris-agreement
9. Heeks, R.: Information systems and developing countries: failure, success, and local improvisations. Inf. Soc. **18**(2), 101–112 (2002). https://doi.org/10.1080/01972240290075039
10. Pellizoli, R.: 'Green revolution' for whom? Women's access to and use of land in the Mozambique Chókwè irrigation scheme. Rev. Afr. Polit. Econ. **37**(124), 213–220 (2010)
11. Veldwish, G.J., Beekman, W., Bolding, A.: Smallholder irrigators, water rights and investments in agriculture: three cases from rural Mozambique. Water Altern. **6**(1), 125–141 (2013)
12. van der Struijk, L.F.: Smallholders' strategies in response to contract farming programs in Chokwe, Mozambique – farmers' reactions to the process of relocation of land and water access in the Chokwe Irrigation System. Master thesis, Water Resources Management in International Land and Water Management, Wageningen University, The Netherlands (2013)
13. Veldwisch, G.J.: Contract farming and the reorganization of agricultural production within the Chókwè Irrigation System, Mozambique. J. Peasant Stud. (2015). https://doi.org/10.1080/03066150.2014.991722
14. Kojo, A., Chichava, S.: South-south cooperation, agribusiness, and African agricultural development: Brazil and China in Ghana and Mozambique. World Dev. **81**, 13–23 (2016). https://doi.org/10.1016/j.worlddev.2015.11.021
15. Meinke, H., Nelson, R., Kokic, P., Stone, R., Selvaraju, R., Baethgen, W.: Actionable climate knowledge – from analysis to synthesis. Clim. Res. **33**, 101–110 (2006). www.int-res.com/articles/cr_oa/c033p101.pdf
16. Ulrich, W., Reynolds, M.: Critical systems heuristics. In: Reynolds, M., Holwell, S. (eds.) Systems Approaches to Managing Change: A Practical Guide, pp. 243–292. Springer, London (2010). https://doi.org/10.1007/978-1-84882-809-4_6. Link to the chapter http://oro.open.ac.uk/21299/1/systems-approaches

An Efficient Security Implementation with Power Cane for Visually Challenged

J. Saravanan[1], A. Clement Raj[1(✉)], Raushan Kumar Singh[2], and Ararso Taye[1]

[1] Bule Hora University, Bule Hora, Ethiopia
saravanan.j2k@gmail.com, clementlecturer@gmail.com, arstay19@gmail.com
[2] Spectrum Solutions, Pondicherry, India
spectrumsolutions.md@gmail.com

Abstract. Eyes considered as the most precious organ of the human body, being blind is considered one of the biggest physical disaster for the human. As per the survey, 39 Million people are blind throughout the world. These blind people go through several challenges throughout their life 24 × 7. Some of the major critical issues they face are their movement through the lanes especially in dense traffic areas, they are highly vulnerable against robbery, kidnapping, and molestation due to their impairment, and they can't find they are moving through dark or lightened lanes. These are some of the major issues they face which can be mitigated with the efficient use of technology and electronic components. We can enable the walking sticks of the blind person to act as the Stun Gun which produces 20 KV in emergency condition for their protection using Electric Arc Generator and power source. Using Ultrasonic Sensor provides the estimation of the obstacle and vibrates the stick if comes nearby any obstacle below the threshold limit. The major electronic components we use are Ultrasonic Sensor (HC-SR04), Arduino Nano, Electric Arc Generator, LDR, Dry cell, and Vibration Motor. We can design an efficient blend of technology and logic to short out the issues of the blind community which will be our gift to humanity.

Keywords: HC-SR04 · Electric Arc Generator · LDR · Arduino Nano · Vibration motor

1 Introduction

Blind means the disability of vision which reciprocates to the impairment in society too. Visually impaired go through several social and physical traumas. Being impaired and helpless they are not able to defend themselves when finding in a critical scenario like in heavy traffic zone or in case of criminal incidences [2]. There have been many types of research for helping blinds with their proper navigation using several sensors [3, 4, 6]. Especially to find their path and get alert about their navigation path. There

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 270–277, 2019.
https://doi.org/10.1007/978-3-030-26630-1_23

has not been any attempt to enable the physically impaired to find their path in any condition and protect against criminal activities combine. Our proposed system deals with the blinds to help themselves in case of an emergency. To fulfill the requirements we proceed with the brain of our system which is Atmega328 P PU microcontroller based Arduino Nano. It is a small and lightweight yet effective solution which is idle for this scenario due to its tiny size to fit on walking sticks. It is programmed with Embedded C code to perform the task. Then we proceed with our second requirement which has two applications. First to create a sensor based navigation system to alert the blind for any obstacle and second to produce high voltage arc to give a shock to criminals in case of emergency. To perform the first task we use the Ultrasonic Sensor (HC-SR04) which acts with the same fashion as the Bats act using ultrasonic frequency to detect the obstacles. For the second we deal with high voltage arc generator which produces huge power to shock the culprit. For the third we use LDR to make him understand whether he is moving through light or darkened lanes.

1.1 Microcontroller (Brain of the System) Selection

We use Arduino Nano to be used as the logical think-tank to perform our tasks for blind security and navigation system. It has the capability of steady design, smaller form factor and yet powerful analytical ability based on incorporated programming.

1.2 The First, Second and Third Initiative

The first initiative is to solve the issue of obstacle detection and avoidance for blind people. We use ultrasonic sensors for that. The ultrasonic sensor identifies the distance and provides the logic to the microcontroller which activates vibration motor. The vibration motor vibrates the walking stick so the blind person gets alert about the obstacle (Fig. 1).

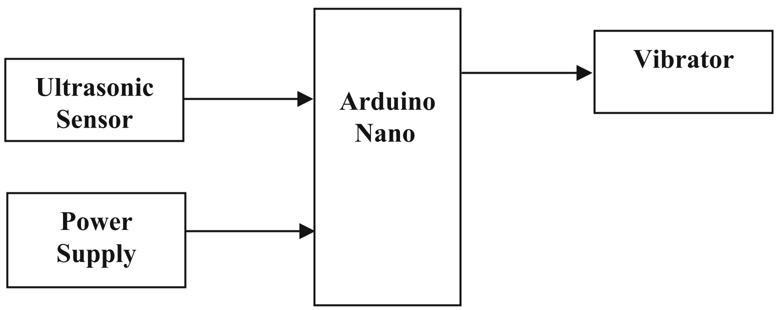


Fig. 1. The obstacle identification and alert scenario

The second initiation is for high voltage current production to provide sufficient shock to the criminals in case of emergency. For that, we use electrical Arc device which can generate a voltage as heavy as 20 kV. This shock does not kill the victim but makes them paralyzed for hours. The device can be operated and energized by just 1.5 V battery itself. It is a lethal weapon if the high voltage battery is used. We have a panic button in the stick itself which activates this feature for the cane. It's so powerful that it can be used for a group of people by producing power from a small battery (Fig. 2).

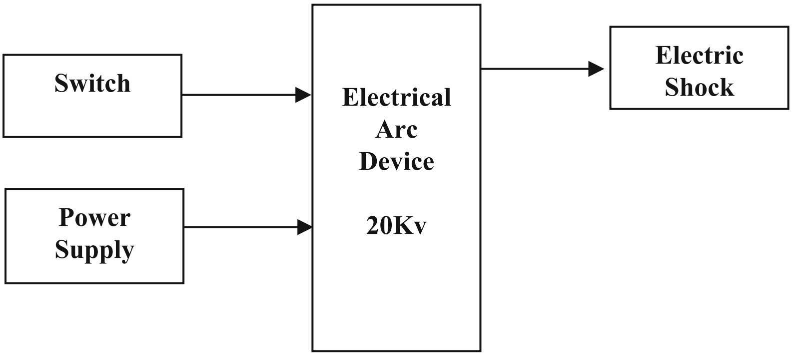


Fig. 2. The electric shock scenario

The system is highly useful and the first of its kind to provide the safety to the blind individual.

1.3 Advantages of the Proposed System

1. Real-Time implementable system
2. Practically Usable
3. Light Weight
4. Compact Size
5. Low Power Consumption
6. Minimum Hardware Possible.

1.4 Objective of the Project

- To make a practically usable lightweight system
- To make a blind stich this actually can be used by a blind person.
- To use a smart-talking system without increasing the hardware and system weight
- To make the system in minimum possible cost.

2 Hardware Components Specifications and Usage

2.1 Ultrasonic Sensor (HC-SR04)

The sensor deals with the process of ultrasonic frequency by transmitting the signal from the transmitter and after reflection from any reflecting body, it gets received by the receiver unit (Fig. 3).

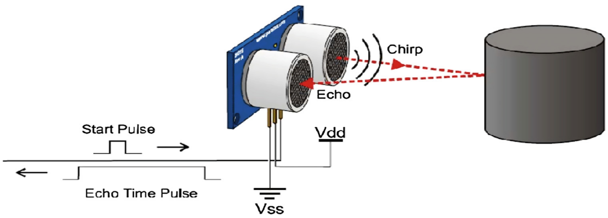


Fig. 3. The working principle of the ultrasonic sensor

2.2 Arduino Nano

It is the think-tank of our system which deals with the C Programming language to provide logic. All the sensor logic is processed and provided to output motor device to vibrate the stick (Fig. 4).

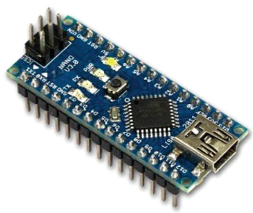


Fig. 4. Arduino Nano

2.3 Electric Arc Device

This device is a powerful shock provider even from a Drycell. The shock is so powerful that it can paralyze a person for hours. This is not deadly but acts as the stun gun in case of emergency (Fig. 5).

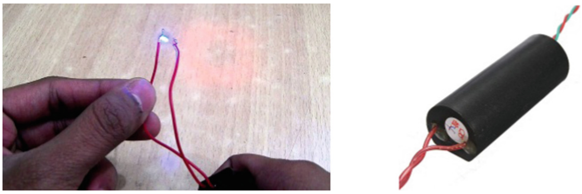


Fig. 5. Electric arc working

The combined utilization of the above components can be used to create a powerful supporting and defending cane system for blinds. All necessary information about the combined approach is mentioned below in the form of overall block diagram.

2.4 LDR

It stands for Light Dependent Resistor or Photoresistor, which is passive electronic component, basically a resistor which has a resistance that varies depending of the light intensity. The resistance is very high in darkness, almost high as 1 MΩ but when there is light that falls on LDR, the resistance is falling down to a few KΩ (Fig. 6).

Fig. 6. LDR

2.5 Vibrator Motor

The vibration motor is a special type of motor which produces heavy vibration during obstacle detection. Whenever the vibration is produced in accordance with an obstacle then the user can understand he/she is nearing the obstacle and gets alerted (Fig. 7).

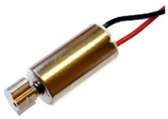


Fig. 7. Vibration motor

3 Design Diagrams and Programming

3.1 Microcontroller Program

```
const int trigPin = 2;
const int echoPin = 4;
#define LEDPin 13 // Onboard LED
void setup()
{
 Serial.println("$bc11");
 Serial.begin(9600);
 pinMode(LEDPin, OUTPUT);
 Serial.println("$bc6");
 Serial.println("Blind navigation system ");
 Serial.println("*blind navigation system ");
 delay(2000);
 Serial.println("$ta90");
 Serial.println("$ts40");
 delay(1000);
}
void loop()
{
Serial.println("$bc12");
 long duration, inches, cm;
 pinMode(trigPin, OUTPUT);
 digitalWrite(trigPin, LOW);
 delayMicroseconds(2);
 digitalWrite(trigPin, HIGH);
 delayMicroseconds(10);
 digitalWrite(trigPin, LOW);
 pinMode(echoPin, INPUT);
 duration = pulseIn(echoPin, HIGH);
 inches = microsecondsToInches(duration);
 cm = microsecondsToCentimeters(duration);
 Serial.println("$ts60");
 Serial.print(inches);
 Serial.print(" Inches ");
 Serial.print("      ");
 Serial.print(cm);
 Serial.print(" C.M");
 Serial.println();
 delay(500);
  if (inches <= 2)
  {
to indicate "out of range" */
 Serial.println("$bc7");
 Serial.println("$ts60");
 Serial.println("*Obstacle is there ");
 delay(1000);
 Serial.println("Obstacle ");
 delay(300);
 Serial.println("$ts40");
 digitalWrite(LEDPin, HIGH);
 delay(300);
 }
 else {
  turn LED OFF to indicate successful reading. */
 Serial.println(inches);
 digitalWrite(LEDPin, LOW);
 }
}
long microsecondsToInches(long microseconds)
{
 return microseconds / 74 / 2;
}

long microsecondsToCentimeters(long microseconds)
{
 return microseconds / 29 / 2;
}
```

3.2 Overall Block Diagram

See Fig. 8.

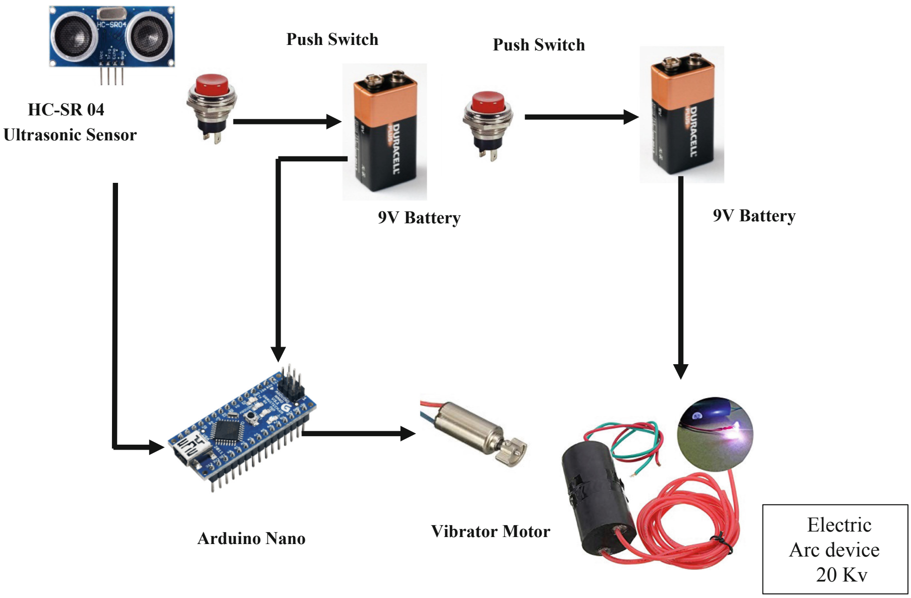


Fig. 8. Overall block diagram

4 Conclusion

Thus we developed a system for visually impaired individuals by a perfect blend of ultrasonic, microcontroller and power generation technology. The efficient utilization of these resources can bring ultimate change in the degrading condition of the blind. It can act as the support system and a helping hand in the case of emergency for physically challenged individuals.

Acknowledgments. This work is the efficient utilization of the perfect technological balance of J. Saravanan, A. Clement Raj, Raushan Kumar and Ararso Taye. The idea of paper and technical implementation is carried out mutually by the authors and brings the perfect implementable concept for social welfare.

References

1. Bernabei, D., Ganovelli, F., Di Benedetto, M., Dellepiane, M., Scopigno, R.: A low-cost time-critical obstacle avoidance system for the visually impaired. In: International Conference on Indoor Positioning and Indoor Navigation (IPIN) (2011)
2. Bach y Rita, P., Kaczmarek, K.A., Tyler, M.E., Garcia-Lara, J.: Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. J. Rehabil. Res. Dev. **35** (4), 427–430 (1998)
3. Vlaminck, M., Jovanovic, L., Van Hess, P., Goossens, B., Wilfried, P., Aleksa dra, P.: Obstacle detection for pedestrians with a visual impairment based on 3d imaging. In: 2013 International conference on 3D imaging (IC3D), pp. 1–7. IEEE (2013)
4. Sainarayanan, G., Nagarajan, R., Yaacob, S.: Fuzzy image processing scheme for autonomous navigation of human blind. Appl. Softw. Comput. **7**(1), 257–264 (2007)
5. Hersh, M.A., Johnson, M.A.: Assistive Technology for Visually Impaired and Blind People, 1st edn. Springer, London (2008). https://doi.org/10.1007/978-1-84628-867-8
6. Hoang, V.N., Nguyen, T.H., Le, T.L., Tran, T.T.H., Vuong, T.P., Vuillerme, N.: Obstacle detection and warning for visually impaired people based on electrode matrix and mobile Kinect. In: 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS), pp. 54–59, September 2015
7. Holz, D., Holzer, S., Rusu, R.B., Behnke, S.: Real-time plane segmentation using RGB-D cameras. In: Röfer, T., Mayer, N.M., Savage, J., Saranlı, U. (eds.) RoboCup 2011. LNCS (LNAI), vol. 7416, pp. 306–317. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32060-6_26
8. Abdullah, S., Noor, N.M., Ghazali, M.Z.: Mobility recognition system for the visually impaired. In: ISTT, Malaysia, pp. 362–367 (2014)
9. Takizawa, H., Yamaguchi, S., Aoyagi, M., Ezaki, N., Mizuno, S.: Kinect cane: object recognition for the visually impaired, Poland, pp. 473–478 (2013)
10. Agarwal, N., Iyer, A., Naidu, S., Rodrigues, S.: Electronic guidance system for the visually impaired—a framework, Mumbai, pp. 1–5 (2015)

Improving Quality of Service of Border Gateway Protocol Multiprotocol Label Switching Virtual Private Network of EthioTelecom Service Level Agreements

Asrat Mulatu Beyene[1(✉)] and Shimelis Asrat Argaw[2]

[1] Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
asrat.mulatu@aastu.edu.et
[2] St. Mary's University, Addis Ababa, Ethiopia
shimelisas@gmail.com

Abstract. The primary goals of Quality of Service (QoS) are managed bandwidth, controlled jitter, minimized latency, and improved packet loss characteristics to provide satisfactory services for users. Shaping network optimization is crucial for the service provider, too. To implement QoS mechanisms optimizing the current network physical and logical architectures is among the best practices. In this paper, an attempt has been made to investigate the end-to-end QoS parameters of multiprotocol border gateway protocol multiprotocol label switching virtual private network (MP-BGP MPLS VPN) EthioTelecom service level agreement (SLA) customers. That is done using differentiated service (DiffServ) model to manage end-to-end traffic delay, jitter, and packet losses. The traffics are classified and marked depending on their priorities. The proposed network architecture has used weighted fair queueing (WFQ) for congestion management and weighted random early detection (WRED) for congestion avoidance. The Huawei's Enterprise Network Simulation Platform (eNSP) and Wireshark are used to design, demonstrate and evaluate the network architectures. When the results of the existing network are compared with the proposed network architecture its delay, jitter, packet loss and traffic utilization have shown improvements.

Keywords: Quality of Service · Virtual private network · Multiprotocol label switching · Multiprotocol border gateway protocol · Service level agreement · Differentiated service model

1 Introduction

Every day new telecommunication technologies are being developed. Enterprises use these new technologies to upgrade their network services and reduce cost.

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 278–288, 2019.
https://doi.org/10.1007/978-3-030-26630-1_24

Now a day, different kinds of traffic such as voice, video, and data are sent over the same network infrastructure. When transferring different traffic types within the same network infrastructure QoS is one the challenges [1]. MP-BGP MPLS VPN is one alternative solution to private wide area networks (WAN) to assure end-to-end QoS.

In existing networks there is significant problem of meeting consumers QoS demands. This is discussed in [1–3]. Moreover, a preliminary investigation based on data collected from EthioTelecom MP-BGP VPN customers shows there were many verified complaints of end-to-end QoS problems. The primary reasons for this are EthioTelecom uses best effort policy to guarantee QoS demands, first in first out (FIFO) to manage congestion, tail drop to avoid congestion. This can be improved in many different ways. Here, we have implemented the DiffServ QoS model, WFQ for congestion management, WRED for congestion avoidance on EthioTelecom MP-BGP MPLS VPN customers network to better meet QoS expectation based on the SLA. To do that, in this research work, a simplified network architecture is built as shown in Fig. 1. It covers the main steps in designing QoS of MP-BGP MPLS VPN networks. The New Generation Network (NGN) network architecture was chosen according to the requirements of designing networks with service provision and end-to-end QoS implementation.

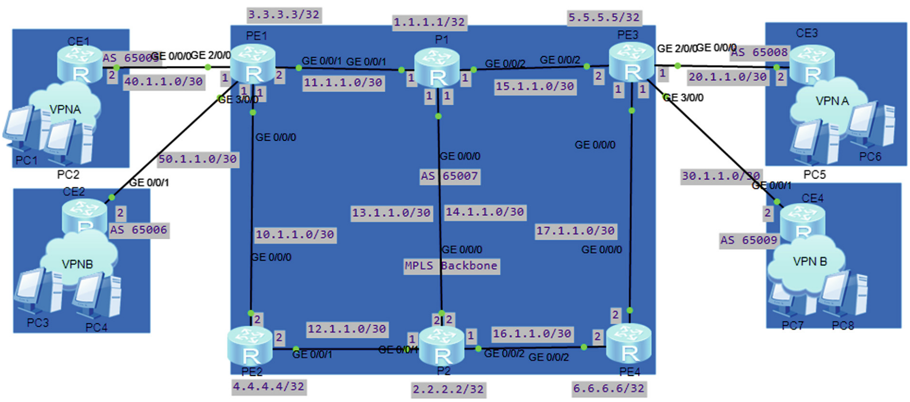


Fig. 1. Simplified MP-BGP MPLS VPN network architecture with end-to-end QoS.

In the proposed network architecture solution there are two provider (P) routers, four provider edge (PE) routers and four customer edge (CE) routers. The P routers are the backbone routers. They provide MPLS label forwarding and maintain public network routing information. PE routers are directly connected with CE routers. The functions of PE routers are maintaining and processing VPN route information, forwarding VPN, running MP-BGP and MPLS. They also do label popping and imposition. CE routers are customer edge routers where the customer's routers or personal computers (PCs) are directly connected.

VPN A and VPN B routers are traffic generators. The two VPN pairs, VPN As and VPN Bs, traffic was evaluated. Both are MPLS based on resource reservation protocol traffic engineering (RSVP-TE) for signaling and tunneling. They use intermediate system to intermediate system (IS-IS) for interior gateway-protocol (IGP) interconnection. Both VPNs use the same networking devices. The P routers in the core network are realized as core routers and route reflectors (RR). These devices are logically divided into two logical systems. These systems are act like separate routers. They have full functional capabilities of two separate hardware devices. The connections between the two logical systems are made with peering the interfaces. The links with the other devices in the networks are recognized with general Gigabit Ethernet interfaces.

The access and aggregation networks are made with secure service routers. They are working in multiprotocol label popping and positioning mode instead of the default packet flow due to the MPLS architecture. The devices are working with Gigabit Ethernet interfaces. The links with the core networks are through Gigabit Ethernet and the links with the end devices are also with full Gigabit Ethernet. The access and aggregation routers apply the QoS to the traffic from the end devices. The two VPN routers are traffic generators. To test the QoS applied in the traffic flow Wireshark and eNSP are used. The two VPNs provide random traffic generation, fixed or non-fixed packet size, and simultaneous generation of multiple traffic flows.

1.1 MP-MP BGP MPLS VPN

In the BGP MPLS VPN, BGP is used to transfer VPN private network route information on the carrier backbone network and MPLS to forward VPN service steams. Depending on the working principles of BGP MPLS VPN there are three aspects: route information advertisement, label distribution and packet forwarding [2,3]. Route information advertisement is used for the exchange of information from the local CE to the ingress PE, from ingress PE to egress PE and egress PE to local CE. Label distribution distributes private network and public network labels. VPN packet forwarding is used for encapsulation, outer packet forwarding on a public network and inner label instructing inner sites of packets [1–3].

1.2 Quality of Service (QoS)

QoS is the mechanism of networks to provide different services to different traffic types [4]. Service providers offer their network service with varying quality levels. To do that they define SLAs. An SLA provides the details of all QoS parameters. It defines the parameters such as end-to-end delay, end-to-end jitter, and packet loss. QoS is not the functionality of a single device and it is an end-to-end mechanism. It provides the intelligence to network devices to treat the different application's traffic as they are defined in the SLA. QoS combines different technologies together such as classifying, marking, scheduling, queuing, allocating and prioritizing bandwidth that are commonly used to provide a scalable end-to-end service [5]. QoS is used to manage the main network performance elements like bandwidth, delay, jitter and packet loss [4,6].

Bandwidth. The amount of data that can be transmitted over the link is bandwidth [4,9]. On the network, IP Packets travel through the best route. The maximum bandwidth of the route is equal to the smallest value of bandwidth on the route. The available bandwidth is the path bandwidth divided by several traffic flows [5,10]. Due to the low bandwidth users experience delay, jitter and packet loss in the communication.

Delay. End-to-end delay is the total time that a packet takes from source to destination [6,7]. End-to-end delay is the sum of processing, queuing, serialization, and propagation delays.

Jitter. Variation in delay is jitter. Packets for the same destination may not arrive at the same rate. Jitter can occur due to different traffic loads on different timings. For voice and video, it is necessary to receive the packets in the same sequence to achieve good quality [10].

Packet Loss. Packet loss occurs due to the low buffer space [8,9]. When the buffer space of interfaces are full packets are dropped. In queue scheduling packet loss occur when the queue is full. Packet loss creates extended delays and jitter. Packet loss can be controlled by applying techniques like tail drop, random early detection, weighted random early detection and traffic shaping and policing [9].

Generally, QoS doesn't depend only on bandwidth, delay, packet loss and delay [10]. It also depends on end-user perception of telecommunication services such as trends, advertising, tariffs and costs which are interrelated with customers' expectation of QoS. Figure 2 shows how end-user perception reaches the QoS satisfaction level.

QoS can be divided into two viewpoints [11,12]. Customer viewpoints and Service provider viewpoints. Customer viewpoints include QoS requirements and perception whereas service provider viewpoints include QoS offered and QoS achieved as shown in Fig. 3.

Generally, if network performance was well optimized, service provider viewpoint reaches the highest level. Moreover, if service provider affords quality services to its customer, customer viewpoint escalates which increases customer quality of experiences.

2 Designed QoS of MP-BGP MPLS VPN

The designed QoS-based network architecture provides different levels of service quality based on end-to-end QoS targets and International Telecommunication Union (ITU) threshold quality requirements for different VPNs. Managing

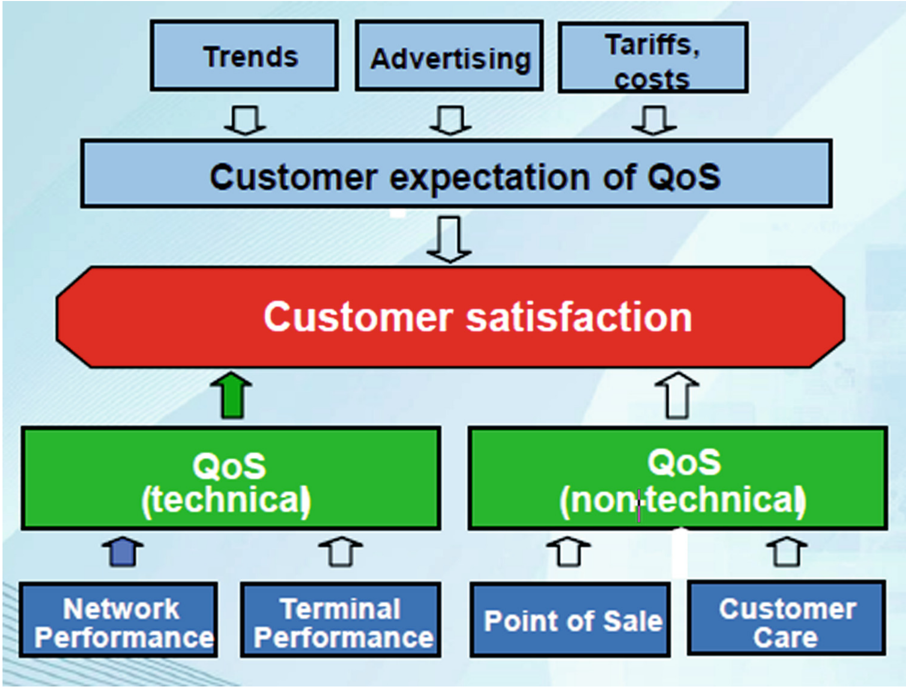


Fig. 2. User perception of end-to-end QoS delivery framework [11,12].

maximum receivable bandwidth, reducing transmission, queueing and processing delay, minimizing jitter and packet loss are the main focus of the design. End-to-end QoS assurance is achieved based on existing resources by using rational scheduling and congestion avoidance methods. DiffServ is used to classify, mark and shape network traffic based on existing SLA agreements. Applying end-to-end QoS using DiffServ can follow the following step by step processes:

- Define access control list (ACL) rules,
- Define traffic classifiers,
- Define traffic classifiers,
- Define traffic behaviors (mark),
- Define traffic policies, and
- Apply traffic policies to interfaces.

2.1 Experimental Resuts

The designed MP-BGP MPLS VPN end-to-end QoS applied on the network architecture depicted in Fig. 1 is fully operational. That means

- All protocols are fully operational,
- Proper implementation of the designed QoS is made,

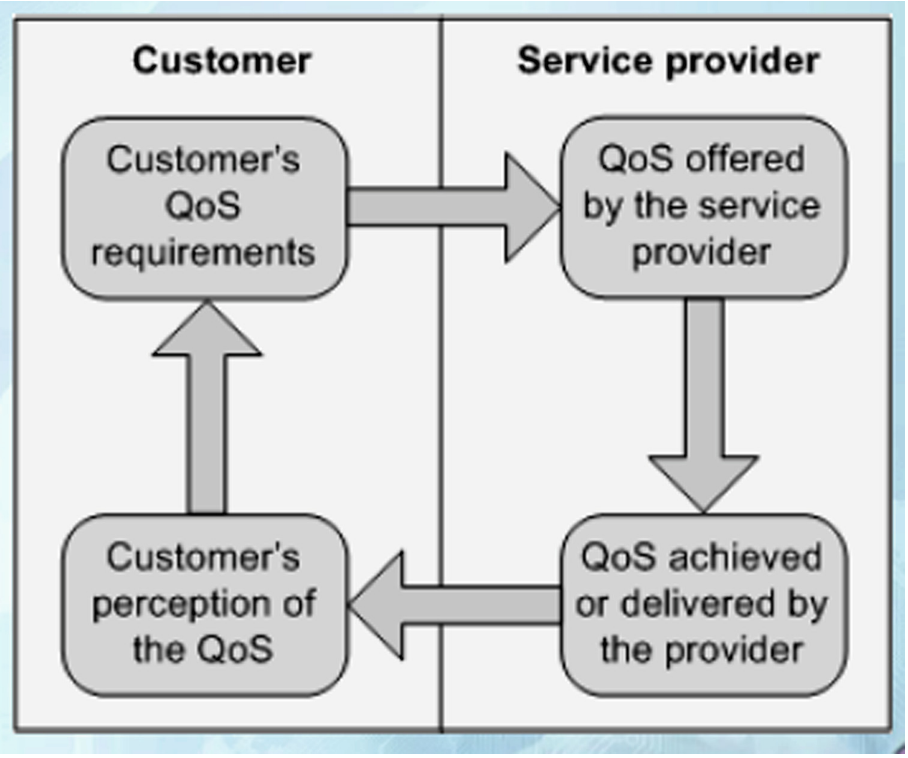


Fig. 3. QoS viewpoints framework [11].

– Provisioning of the required services ensuring MPLS VPN are operational, and
– Redundancy of network resources is made which includes rerouting in case of link or node failure.

The existing and proposed network architectures are the same in devices and physical interconnection. But they have different QoS designs. Table 1 shows the similarities and differences between existing and proposed network architectures.

Table 1. The similarities and differences between existing and proposed network architectures.

Parameters	Existing network architecture	Proposed network architecture
Traffic type	MP-BGP MPLS VPN	MP-BGP MPLS VPN
Service type	MPLS VPN	MPLS VPN
IGP routing protocol	IS-IS	IS-IS
NGN backbone	MPLS	MPLS
QoS model	Best effort model	DiffServ
Congestion management	FIFO	WFQ
Congestion avoidance	Tail drop	WRED

The end-to-end QoS is tested with Wireshark and eNSP tools. A couple of scenarios, based on Fig. 1, are tested with different traffic streams. In the first scenario, the existing network performance is checked. The existing network architecture uses best effort endto- end QoS. All traffic has equal priority. The architecture uses FIFO algorithm for congestion management and tail drop algorithm for congestion avoidance. In the second scenario, which is the proposed network architecture, uses the DiffServ QoS. The traffic has different priorities. The architecture uses WFQ algorithm for congestion management and WRED algorithm for congestion avoidance. In this case, the traffic was classified and prioritized depending on the underlying SLA. Then traffic policies were defined and applied on the aggregation router outbound interface. In this work, the generated traffic consists of two VPN instance application traffic streams. The two VPN instance traffic flows emulate two end nodes connected to the CE routers. Both traffic streams use TCP with speed of 15 Mbps. The existing traffic test is made between CE1 and CE3 and the proposed traffic test is made between CE2 and CE4 routers.

From Fig. 4, one can see that the existing network architecture overturns the bandwidth utilization. This is because the existing network uses the best effort QoS model which cannot isolate the services to guarantee the maximum data transfer. But in the proposed network architecture the bandwidth utilization is respectable. In this case, the network uses the DiffServ QoS model which isolated the network at each aggregation. The isolated aggregate guaranteed to transmit maximum number of data traffic. So, mission critical traffic is transmitted first.

As can be seen from the evaluation testing of Fig. 5, the proposed network architecture, the implementation of DiffServ has many benefits for packet loss compared to the best effort. In the DiffServ model, routers must store traffic and QoS information per aggregation. This creates enough buffer space in the router's queue. A router usually has incoming interface buffers, system buffers, and outgoing interface buffers. In case of congestion, the traffic is remarked and kept in buffer space to avoid the packet loss. But in the case of best effort QoS model, the routers just route packets until they reach the destination. Other packets are dropped causing a higher percentage of packet loss.

DiffServ QoS model minimizes traffic loss. In case of congestion, this model classifies traffic depending on their priority. The classified traffics are marked and shaped depending on the router maximum data transfer rate. Some traffic transmitted, whereas the excess traffics are remarked and transmitted later. This decreases the packet loss ratio.

Latency is the time that a packet waits before being transmitted. As it can be seen from Fig. 6, the proposed network architecture shows lower latency compared to the existing network architecture. The reason for this is that the DiffServ model can guarantee the traffic per aggregation.

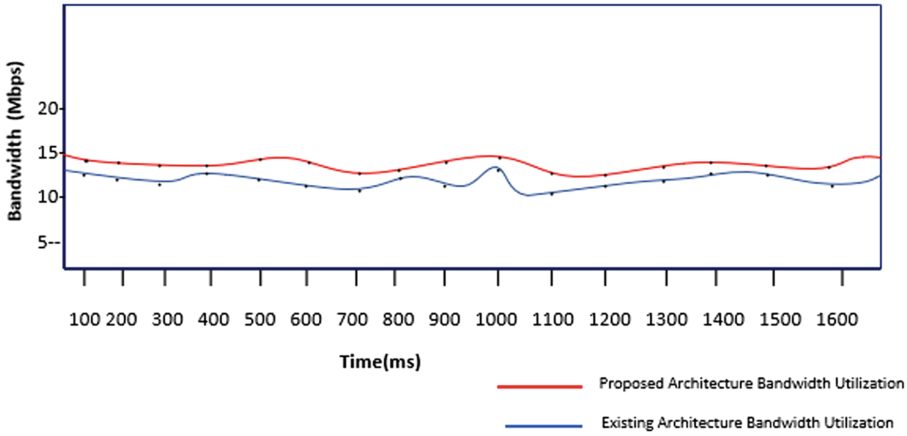


Fig. 4. The bandwidth utilization results of existing and proposed systems.

3 Discussions

The numerical results obtained from the existing and proposed networks are shown in Table 2. Most of the results were as expected. The difference between packet loss and bandwidth in existing and proposed network architecture was visible. But the difference between end-to-end delay and jitter was not that much visible. This happened because we have used ten routers only on both network architectures. This reduces the transmission, serialization, queueing and processing delays. The difference increases as the number of routers (nodes) increases.

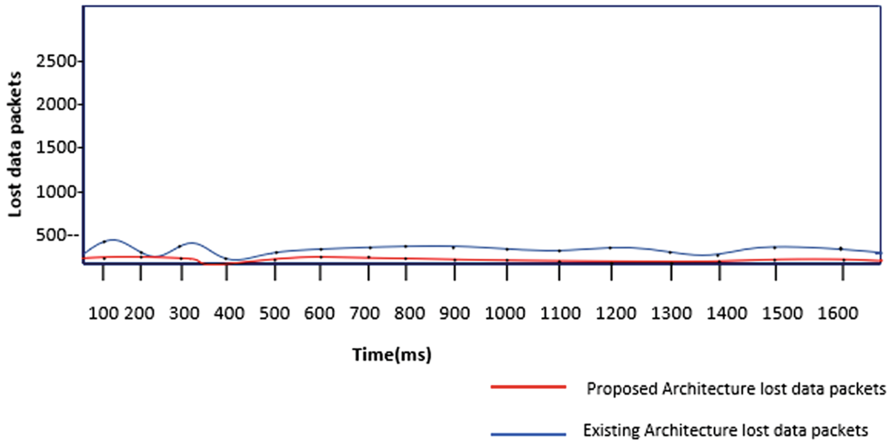


Fig. 5. Packet loss measurement comparison.

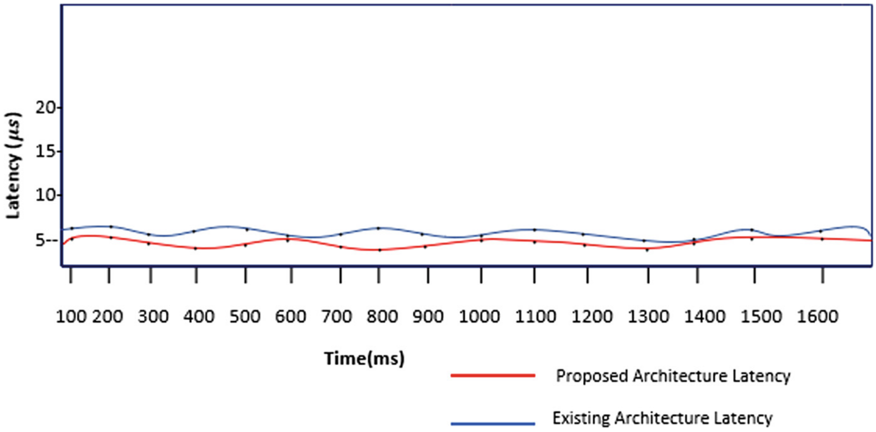


Fig. 6. Latency measurement comparison.

Table 2. Exist and proposed network architecture numerical QoS results.

Parameters	Existing network (Best effort)			Proposed network (DiffServ)		
	Result	SLA targets	ITU threshold	Result	SLA targets	ITU threshold
Packet loss ()	0.169	Within range	Out of range	0.14132	Within range	Within range
Delay (sec)	0.001	Within range	Out of range	0.14132	Within range	Within range
Jitter (sec)	0.001	Within range	Out of range	0.0007747	Within range	Within range
Bandwidth (bps)	15068	Out of range	Out of range	15320	Within range	Out of range

4 Conclusions

In this research, the DiffServ model for the design of MP-BGP MPLS VPN networks with end-to-end QoS was deliberated. This type of networks is suitable for the implementation of QoS for MPLS VPN networks. A simplified network topology was created. Two network architectures were designed, built and evaluated with generic telecommunication equipment. Firstly, the existing BGP MPLS VPN network which used best effort QoS model was implemented and tested. Secondly, the proposed BGP MPLS VPN architecture which uses the DiffServ QoS model was designed and tested. Bandwidth utilization, packet loss, latency and jitter measurements were made for both network models. After the whole evaluations were made, it is observed that the proposed MP-BGP MPLS VPN network architecture has much more benefits than the existing BGP MPLS VPN network architecture. This is due to the opportunity for the class of services and traffic-engineering in the network, which brings better traffic management and provision of suitable end-to-end QoS. The proposed MP-BGP MPLS VPN architecture which uses DiffServ QoS model architecture could be used in many mission-critical applications.

In the proposed DiffServ QoS model better network productivity was achieved. The designed MP-BGP MPLS VPN architecture which uses DiffServ

QoS model network architecture is easy to scale and troubleshoot. The addition of new end devices in the network is simplified and just slight configuration changes are required. In the proposed BGP MPLS VPN architecture, which uses DiffServ QoS model architecture, all services have the required traffic treatment. The designed MP-BGP MPLS VPN network model can easily be used for MPLS VPN services in both centralized and distributed architectures. End-to-end MPLS solutions for the NGN applications are smoothly attended.

Generally, based on the analysis and results gained, it can be concluded that the DiffServ QoS model was more reliable than the best effort QoS model being used by EthioTelecom's MP-BGP MPLS VPN network. The designed QoS uses DiffServ model that can guarantee customers' SLA QoS thresholds. In conclusion, the designed network provides a way of increasing network performance based on the DiffServ QoS model. High network performance indicates better QoS service provision. Better QoS service provision, in turn, creates customer satisfaction and higher quality of experience to customers.

5 Future Works

Based on the scope of this work, the QoS has been guaranteed with respect to SLA QoS targets. But in the future, the network can be extended with more reliability functions. These functions include chassis clustering for access and aggregation devices, implementation of high availability features, implementation of LDP for MPLS label down streaming on demand.

One such future extension is the implementation of self-organizing network architecture, such as self-learning, self-configuration and self-management, self-optimization, prediction of network congestion and traffic loops. To implement advanced extensions there are algorithms for prediction. Algorithms for adaptive training of the network such as the Widrow-Hoff algorithm can be of great use for process predictions in operating networks [13,14]. This way the designed proposed BGP MPLS VPN architecture which uses DiffServ QoS model network architecture can become optimal save operational and maintenance costs. Self-optimization, based on collected data from previous network states and based on predictions can be also be attempted.

References

1. Ahmad, A., Talal, A.: Performance analysis of DiffServ-based quality of service in MPLS networks. Int. J. Sci. Eng. Res. **6**, 15–23 (2015)
2. EthioTelecom. QoS Document on MPLS VPN Services Quality and Customer Experience Related Issues and Complaint Analysis, Version 02 (2017)
3. Huawei technologies. Configuration Guide VPN - Cloud Engine 12800 Series Switches, vol. 6 (2017)
4. Szigeti, T., Hattingh, C.: End-to-End QoS Network Design. Cisco Press, Indianapolis (2005)
5. Lawrence, J.: Designing multiprotocol label switching networks. IEEE Commun. Mag. **39**(7), 134–142 (2001)

6. De Ghein, L.: MPLS Fundamentals epub. Cisco Press, Indianapolis (2016)
7. Pepelnjak, I., Guichard, J.: MPLS and VPN Architectures, vol. 1. Cisco Press, Indianapolis (2002)
8. Mohamed, J.N., Mellachervu, K.R., Ramakrishnan, S.C.: Configuration tool for MPLS virtual private network topologies. Google Patents (2012). (US Patent 8,194,570)
9. Odom, W.: CCNP Route 642–902 Official Certification Guide. Cisco Press, Indianapolis (2010)
10. Grayson, M., Gundavelli, S.: Providing integrated end-to-end architecture that includes quality of service transport for tunneled traffic. U.S Patents (2015). (Patent 9,198,209)
11. Stankiewicz, R., Jajszczyk, A.: A survey of QoE assurance in converged networks. Comput. Netw. **55**(7), 1459–1473 (2011)
12. EthioTelecom. Service Provisioning Manual, July 2014
13. Zhang, G.P.: Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing **50**, 159–175 (2003)
14. Lee, H., Hwang, J., Kang, B., Jun, K.: End-to-end QoS architecture for VPNs: MPLS VPN deployment in a backbone network. In: 2000 Proceedings of International Workshop on Parallel Processing, pp. 479–483. IEEE (2000)

Natural Language Processing Track

A Scalable Text Classification Using Naive Bayes with Hadoop Framework

Mulualem Mheretu Temesgen[1(✉)] and Dereje Teferi Lemma[2]

[1] Collage of Computing and Informatics, Assosa University, Assosa, Ethiopia
mulu2006hu@gmail.com

[2] School of Information Science,
Addis Ababa University, Addis Ababa, Ethiopia
dereje.teferi@aau.edu.et

Abstract. Automated text classification is the labeling of documents to the predefined class label or category using machine learning algorithms. It is one of the important domains in machine learning where the algorithm is applied to classify documents to the appropriate category or genre of the document. For example, the document might be news items and the class/category/genre might be business news, sport news, health news, financial news and social news. Due to the volume of this textual data and its presumed exponential growth, classical data mining techniques may not provide optimal performance in terms of efficiency. To this end, scalable machine learning library apache mahout with hadoop can be used to improve the performance of the algorithm and computation time. In this study Naïve Bayes classification algorithm is implemented on top of hadoop to build automatic document categorizer using Mapreduce programing model. Addis Ababa university institutional repository/Electronic thesis and dissertations text document is used for training and evaluation dataset. The proposed model achieved an accuracy of 79.06%. The result shows that the system can categorize large thesis documents into its predefined class with promising accuracy.

Keywords: Machine learning · Text classification · Naïve Bayes · Mapreduce · Hadoop

1 Introduction

Text classification is the task of classifying document by their contents or by the words that they contain [1]. It automates the task of filing documents into pre-defined categories based on a set of example documents [2]. Text document classification is one of important domains in machine learning. It is the application of learning algorithms to classify documents to their appropriate category or genre. Each instance in the dataset represents a document and the instance class represent the document category. For example, the document might be news items and the class might be business news, sport news, health news, financial news and social news. Every document is characterized by words it contains that will be used to predict the class label or category of the document by applying machine learning algorithm.

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 291–300, 2019.
https://doi.org/10.1007/978-3-030-26630-1_25

According to Tiwary [3], text analytics is one of the most used application area of machine learning that is implemented in automatic email filtering, news article clustering and categorization, and sentiment analytics on social media posts. They contend that the estimated size of text data both public and private could be measured in petabytes[1]. They argue that, due to the volume of this textual data and presumed exponential growth, classical data mining techniques may become inadequate to deal with this data. So, they insist that there is huge opportunity to apply a scalable machine learning algorithm to a text document to build a classification application for unstructured digital texts.

Generally, there are three approaches of machine learning: supervised, unsupervised and reinforcement learning. The first two approaches can be applied in text categorization application. In supervised machine learning algorithm, the class label of the document is known and trained to predict the new unseen class label which is referred to as classification. Whereas, unsupervised machine learning algorithm works on the unlabeled documents. This is the class label for each instance is not known and hence one can only work on grouping the data based on some "distance" criteria. Unsupervised learning is also called clustering. Reinforcement learning is a type of machine learning where an agent learns how to behave in an environment by performing actions and seeing the result. This paper uses Addis Ababa University's categorized thesis documents to build the text classifier model. Since thesis documents are found in unstructured format, preprocessing techniques such as tokenization, stop word removal and steaming of text were applied. Moreover, a vector space model is developed after the preprocessing is performed. Sparse vector is the method used in representing the content baring terms as vector.

Several algorithms are used for text classification such as support vector machine (SVM), Decision tree (C4.5), K-NN (a distance weighted K-nearest neighbor), Naïve Bayes and logistic regression using Stochastic Gradient Descent (SGD). Naive Bayes algorithm is easy and fast to implement and best suited for datasets which have one predictor and one target variable like text document and its class/category. Now a days text classification is mostly used in email spam filtering [4], sentiment classification [5, 6]. and news classification [7]. Once the model is developed, its accuracy is evaluated using confusion matrix and several evaluation metrices.

In this paper text document classification model is developed using Naive Bayes and its scalability for large documents is evaluated using mahout with Mapreduce programing model.

2 Background

2.1 Apache Hadoop Framework

Apache hadoop is an open source apache software foundation project written in java. It is distributed software platform for reliable, scalable and distributed processing of large datasets across clusters of commodity hardware using simple programing model [8]. It

[1] 1 zetta bytes = 10^{21} bytes or 1000 exabyte = 1 million petabytes = 1billion terabyte.

has a distributed processing framework known as Mapreduce that subdivide problems or jobs to a cluster of computers to process in parallel, hadoop distributed file system (HDFS) for storage and Yet Another Resource Negotiator (YARN) which is used as hadoop cluster resource management. There are many projects that complement hadoop. These ecosystems include apache Pig, HBase, Solr and Lucene, Sqoop, flume, mahout and zookeeper etc.

HDFS is distributed file system that provides reliable, scalable and fault tolerant data storage on commodity hardware. It can store large dataset files by dividing it into blocks (default is 128 MB each) and replicate backs in two or more commodity hardware. HDFS has three daemons to store the data and to make HDFS cluster. These are: Namenode, Datanode and Secondary Namenode. Each of which runs when we start distributed file system.

Mapreduce is a programing model for processing large datasets in cluster nodes using distributed and parallel algorithms in side hadoop. Mapreduce refers to two main tasks, Map and Reduce respectively and executed in three steps: Map, shuffle and Reduce [9]. In map phase, the data is partitioned and each partition is given as input to the worker that execute the Map function and produce the key/value pairs. In shuffle phase, key value pairs are grouped by key and each group is sent to the reducer. The reducer receives key/value pairs from multiple map jobs and aggregate these key/value pairs as a final output.

2.2 Apache Mahout

Apache Mahout is a scalable machine learning library from apache run on hadoop that use Mapreduce as a program execution model or independent of hadoop in single machine to process very large datasets. Classification, clustering and collaborative filtering (recommendation engine) are the machine learning tasks that mahout supports. In each task there are algorithms supported by apache mahout that is executed in parallel (in distributed manner in a cluster of computers using hadoop framework) or Sequential (which is run in a single machine without hadoop).

According to Owen et al. [10], mahout begins to be a good choice when the dataset is between 100,000 to 10 million. However, when the data set is below 100,000, traditional machine learning approaches could work well. Apache mahout on the other hand performs well when the dataset is between 1 million to 10 million. Apache mahout has an advantage on larger data sets because as the input data increases the time or memory requirement for training may not increase linearly in non-scalable system. Figure 1 depicts the advantage of using mahout for large datasets.

Mahout include a number of algorithms used to group document into cluster of similar subject area, for example K-means and classification algorithm used to assign category label to text documents for example Naïve Bayes classification algorithm [11].

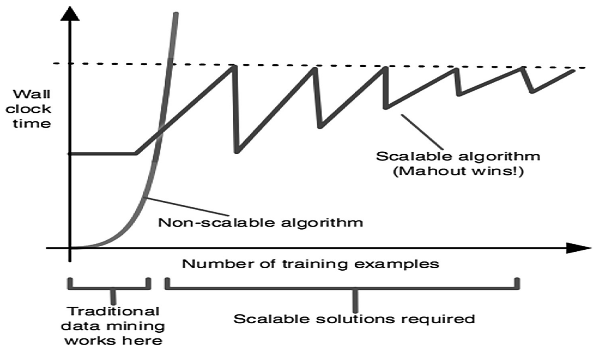


Fig. 1. Advantage of using mahout for large data set and traditional machine learning taken from [10]

3 Naïve Bayes Classifier

Classification is a technique that decides how much a thing/class is or is not a part of some type or category. Naïve Bayes classification algorithm is one of the most widely used supervised algorithm in text classification with mahout. [5] and [12] states that naïve Bayes is proven to be the easy and effective machine learning algorithm for text classification. It is a probabilistic algorithm that build the model based on the relationship between features and categories in training data to predict the category of the class label of the a given new document [13, 14].

Naïve Bayes algorithm is the probabilistic classification algorithm. It uses Bayes theorem on class conditional independence assumption of each features. The training process analyzes the relationship between words in the training documents and categorize and the relationship between categories and the entire training set. Naïve Bayes algorithm doesn't give the absolute probability of the word belonging to a certain class. But it is used to determine that a document containing the word is more likely belonging to a certain category (class label) by comparing the probability assigned to a word for each category [11].

During training the Naïve Bayes algorithm counts the number of times each word appears in a document in the class and divide that by the number of words appearing in that class. This is referred to as conditional probability - that means the probability that a word will appear in particular category which is expressed in Eq. 1.

$$\mathrm{P(word|category)} \tag{1}$$

Bayes Theorem

Bayes theorem states that the probability of a category given a document is equal to the probability of document given the category multiplied by the probability of the category divided by the probability of the document. This is expressed mathematically as in Eq. 2 by [11].

$$\text{P(category|document)} = \frac{P(document|category * P(category))}{P(document)} \quad (2)$$

4 Related Works

Text classification has been one of the main areas of research to apply machine learning. It has been used for search engine, indexing and blog tagging etc. Several researches have been done on text classification by different researchers.

The use of support vector machine (SVM) for training text classifier has explored by [15]. In this experiment four conventional learning method such as, Naïve Bayes, Rachioo (the most popular learning method in information retrieval), Decision tree (C4.5), K-NN (a distance weighted K-nearest neighbor classifier) and SVM on Routers-21578 compiled by David Lewis and Ohsumed corpus Text dataset. From their theoretical and empirical evidences, it can be seen that SVMs are well suited for text categorization as argued by [16]. Similarly, the application of Support Vector Machine with mixture of kernel (SVM-MK) to design a text classification system has been conducted by [16].

As stated by [17], text classification is playing an important role in information extraction and summarization, text retrieval and question answering. The article described text classification process as depicted in Fig. 2, which is mostly used for automated text categorizer. They identified different machine learning algorithms for text classification, such as decision tree, Naïve Bayes, Rule induction, Neural Network, Nearest Neighbors and Support vector machine.

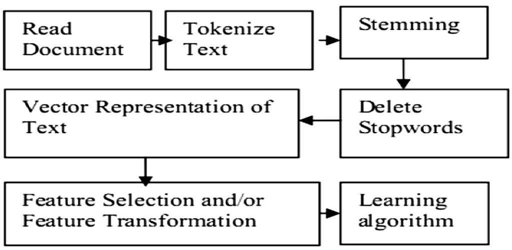


Fig. 2. Text classification process proposed by [17]

Several researches have been done on Amharic news text classification. The use of hierarchical Amharic text classification was explored in [18]. The study presents a comparison of flat and hierarchical Amharic news text classification using SVM and LibSVM multiclass tool. Ethiopian News Agency (ENA) categorized news texts were used to experiment the performance of hierarchical Amharic text classifier. From their research findings, flat news text archive, an accuracy of 68.84% and hierarchical with maximum accuracy of 90.37% is achieved. They have shown that, the performance increases when moving down the hierarchy. Another study has been conducted on the use of machine learning on Amharic news text documents by [19]. The study investigates how to use text classification on Amharic text documents and the effect of stemming or part-of-speech tagging can have on Amharic texts. They have used a corpus of 1065 news documents containing a total of approximately 210,000 word collected form Walta Information Center. They have experimented on their dataset using the RDS system which is rule based machine learning platform. Bagging of decision tree were used as a learning algorithm. Full text representation achieved an accuracy of (69.39%). Artificial Neural network using self-organizing map were used to cluster Amharic texts by using 206 news articles collected from Walta Information Center [20].

Nowadays the size of available data is increasing in volume. Due to the voluminous amount of today's data, it is difficult to use the classical analytics methods and tools for processing. To this end, a number of researches has been conducted to get the best performant classifier. One of the tools developed for this is apache hadoop framework with a scalable machine learning library apache mahout. Mahout includes classification algorithm in its implementation for large datasets [11], described implementation of apache mahout for text classification for large datasets and how it can be integrated using Solr search engine indexing process as document categorizer. Scalable Naïve Bayes algorithm on 1230 news group instances dataset were applied to build the classifier. They produced a classifier that can correctly assign categories with a 79.5% accuracy.

A study conducted by [21], stated that when we have large volume of data, it requires a method such as higher performance parallel computing tools as well as machine learning algorithms. Hadoop is one of distributed file processing and storage and apache mahout for a scalable machine learning algorithm which is run on top of hadoop in parallel. Naïve Bayes classification algorithm were employed to categorize tweets belonging to 3 news group categories. The result revealed that the algorithm classified the tweets with accuracy of 79%.

Robust text classification system that can handle large data by using Mapreduce programing model has developed by [22]. Naïve Bayes classification algorithm have been employed on 2000 articles from Wikipedia and two domains i.e. animal and plant data. The experiment is conducted in single node machine with hadoop configured in pseudo-distributed mode. The classifier achieved an accuracy of 98.8%.

Mapreduce programing model also used in [23], to classify large text documents using multi nominal model using parallel computing. Five categories from 20 newsgroups dataset which has 20,000 texts and 20 categories were used to experiment Naïve Bayes classification algorithm and to test the scalability in five cluster of nodes.

News article classification using Mapreduce programing model with hadoop framework have been used in [24]. They have used Chinese Wikipedia articles for training Naïve Bayes classification algorithm on multimode cluster of computers. The classes are Economy, IT, Health, Sport, Travel, Knowledge, Work, Culture, and Military. Preprocessing of the data by using Chinese word segmentation were applied. Then, features were selected by representing the text as TF-IDF vector. The performance of the classifier was evaluated by precision and recall. Naïve Bayes classifier can also scale up well for large datasets in sentiment analysis [5, 6].

5 Experimental Setup

The proposed algorithm in this research is implemented on the following experimental setup. The type and version of software as well as the hardware used for the experiment is listed in Table 1.

Table 1. Experimental setup

Hadoop version	Hadoop-2.9.2
Programing model	Mapreduce
ML algorithm	Naïve Bayes
File system	HDFS
Operating system	Ubuntu 18.4 which is run on Oracle Virtual box VM
Processor	Intel(R) Core (TM) i5-6200U CPU @ 2.3 GHz
Clustered node	Single node cluster
RAM	4 GB
System type	64 bit operating system
Machine learning library	Mahout -0.12.2
For building mahout code	Maven 3.5.2

6 System Design and Methodology

Hadoop framework is utilized for distributed storage and parallel processing. For the machine learning task, mahout which is a scalable machine learning library that includes classification algorithm in its implementation is used. The dataset for the experiment is gathered from Addis Ababa University[2] digital thesis repository. The total number of thesis documents used for this research are 208. The thesis documents are used as training and test documents to categorize the document based on their genre. Several researchers proposed text classification process model in [7, 17 and 24]. We have proposed text classification process model as shown in Fig. 3. In this study 26.9 MB of size of datasets is used. The documents are first converted into text file and are loaded onto HDFS. Second, the dataset is converted into sequence file by using mahout seqdirectory command on mahout command line interface. This command converts the text file into hadoop SequenceFile by using the file name of the text file as a key and the contents of the text file as a value [25]. Then, to create the TF and TF-IDF vector, one of the mahouts built in utility tool seq2sparse is used. According to Gupta [26], vectors are ordered list of values as defined in two-dimensional space. This TF-IDF vector is used as an input for training the model. The TF-IDF vector dataset is split into training set and testing set with random selection of 60% for training and 40% for testing. Then, training of the Naïve Bayes classification algorithm is performed using HDFS and Mapreduce programing model in single node cluster of computers. After

[2] http://etd.aau.edu.et/.

building the classifier, its performance is measured by different metrics. In this study, accuracy, recall and precision are used to measure the performance of the classifier. Figure 3 depicts the proposed process in this study.

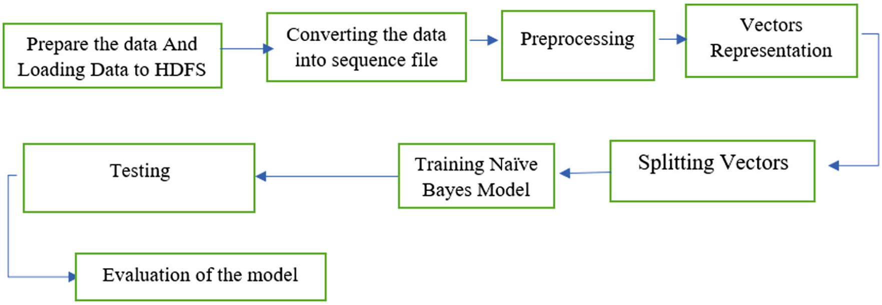


Fig. 3. The proposed process of text categorization

7 Experiment Result and Discussion

In this section, the performance of the automatic thesis document classification to their genre using Naïve Bayes classification algorithm is discussed. Out of the total 208 thesis documents in the dataset, 122 are used for training to build the classifier and 86 are used for testing. Performance is evaluated by using a confusion matrix as shown in Table 2. The model classifies documents with an accuracy of 79.06%. The proposed model also has 85.75% precision and 79.07% of recall. Table 2 discusses the model recall and precision for each class/genre.

Table 2. Confusion matrix

Category		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Total	Recall(%)
Biomedical Engineering	A	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	**100**
Biotechnology	B	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	**100**
Chemistry	C	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**7**	**100**
Computer_science	D	0	0	0	10	0	1	0	0	0	0	1	0	0	0	0	0	0	**12**	**83.3**
Construction_T_managt	E	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	**3**	**100**
Earth_science	F	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	**4**	**100**
Enviromental_Science	G	0	2	1	0	0	0	3	0	1	0	0	0	0	1	0	0	0	**8**	**37.5**
Food_and_nutriation	H	0	0	2	0	0	0	0	5	1	0	0	0	0	0	0	0	0	**8**	**62.5**
General_biology	I	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	**4**	**50**
Hydrolic_engineering	J	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	**3**	**66.6**
Information_Science	K	0	0	0	4	0	0	0	0	0	0	4	0	0	1	0	0	0	**9**	**44.4**
Mathematics	L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**0**
Physics	M	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	**2**	**100**
Sport_science	N	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	**5**	**100**
Stastistics	O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	**6**	**100**
Structural_engineering	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	**5**	**100**
Zoology	Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**1**	**100**
Precision (%)		**100**	**60**	**70**	**71.4**	**100**	**80**	**75**	**100**	**50**	**100**	**80**	**0**	**100**	**71.4**	**100**	**100**	**100**		

As can be seen from the confusion matrix on Table 2, the matrix shows the classifier did an excellent job in classifying biomedical, sport, statistics, earth science, and chemistry text documents accurately. However, the classifier misclassified large number of information science thesis to computer science because the two class labels have almost similar features and unbalanced number of testing data i.e. 12 text documents for computer science and 9 for information science. This indicate that accuracy can be improved by using balanced number of large training datasets.

8 Conclusion and Future Work

In this paper, thesis document repository datasets have been used to build automatic document classification predictive model using a popular machine learning algorithm naïve Bayes on hadoop distributed file system. MapReduce model were employed to process the data in parallel execution mode. We have built the classification model using mahout on top of Hadoop. The classifier gives an accuracy of 79.06%, which is good to classify new documents. The result reveals that, Naïve Bayes algorithm can classify large thesis documents to its category. The experiments were employed on single node cluster of computers, further experiment can be investigated on multi node cluster of computers using large datasets. And also, other classification algorithms can be used to choose highly accurate model on this dataset. Unigram text normalization is used in this paper, it may be effective on bigram and trigrams and accuracy can be improved by taking different parameters.

References

1. Withanawasam, J.: Apache Mahout Essential. Packet, Birmingham (2015)
2. Lucidworks: Text Classification with Mahout and Lucene. https://lucidworks.com/2013/10/30/road-to-revolution-text-classification-powered-by-apache-mahout-and-lucene/. 10 Apr 2019
3. Tiwary, C.: Learning Apache Mahout_Acquire Practical Skills in Big Data Analytics and Explore Data Science with Apache Mahout. Packet, Birmingham (2015)
4. Jiang, E.P.: Content-based spam email classification using machine-learning algorithms. In: Text Mining: Applications and Theory, pp. 37–56. Wiley (2010)
5. Liu, B., Blasch, E., Chen, Y., Shen, D., Chen, G.: Scalable sentiment classification for big data analysis using Naïve Bayes classifier. In: 2013 IEEE International Conference on Big Data, Silicon Valley, CA, pp. 99–104 (2013)
6. Prabhat, A., Khullar, V.: Sentiment classification on big data using Naïve Bayes and logistic regression. In: 2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, pp. 1–5 (2017)
7. Wongso, R., Luwinda, F.A., Trisnajaya, B.C., Rudy, O.R.: News article text classification in Indonesian language. Procedia Comput. Sci. **116**, 137–143 (2017)
8. Ghazi, M.R., Gangodkar, D.: Hadoop, MapReduce and HDFS: a developers perspective. Procedia Comput. Sci. **48**, 45–50 (2015)
9. Kanavos, A., Nodarakis, N., Sioutas, S., Tsakalidis, A., Tsolis, D., Tzimas, G.: Large scale implementations for Twitter sentiment classification. Algorithms **10**, 33 (2017)

10. Owen, S., et al.: Mahout in Action. Manning Publisher, Shelter Island (2012)
11. Ingersoll, G., et al.: Training Naive Bayes using Apache Mahout, pp. 1–10. Mannining Publication, Shelter Island (2015)
12. Kim, S.-B., Rim, H.-C., Yook, D., Lim, H.-S.: Effective methods for improving Naive Bayes text classifiers. In: Ishizuka, M., Sattar, A. (eds.) PRICAI 2002. LNCS (LNAI), vol. 2417, pp. 414–423. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45683-X_45
13. Lewis, D.D.: Effective methods for improving Naïve Bayes text classifiers. Mach. Learn. **1398**, 414–423 (2002)
14. Domingos, P., Pazzani, M.: On the optimality of the simple Bayesian classifier under zero one loss. Mach. Learn. **29**, 103–130 (1997)
15. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 137–142. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0026683
16. Wei, L., et al.: Text classification using support vector machine with mixture of Kernel. J. Soft. Eng. Appl. **5**, 55–58 (2012)
17. Ikonomakis, E.K., Kotsiantis, S., Tampakas, V.: Text classification using machine learning techniques. WSEAS Trans. Comput. **4**(8), 966–974 (2005)
18. Tegegnie, A.K., Tarekegn, A.N., Alem, T.A.: A comparative study of flat and hierarchical classification for Amharic news text using SVM. Int. J. Inf. Eng. **9**, 36–42 (2017)
19. Asker, L., Argaw, A.A., Gambäck, B.: Applying machine learning to Amharic text classification. ResearchGate (2014)
20. Eyassu, B., Gamback, B.: Classifying Amharic text using self organizing map. In: Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages, Ann Arbor (2005)
21. Salur, M.U., Tokat, S., Aydilek, İ.B.: Text classification on mahout with Naïve-Bayes machine learning algorithm. In: International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, pp. 1–5 (2017)
22. Santoso, J., Yuniarno, E.M., Hariadi, M.: Large scale text classification using map reduce and Naive Bayes algorithm for domain specified ontology building. In: 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics, pp. 428–432. IEEE Computer Society, Hangzhou (2015)
23. Chen, H., Fu, D.: An improved Naive Bayes classifier for large scale text. In: 2018 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018) (2018)
24. Feng, M., Wu, G.: A distributed Chinese Naive Bayes classifier based on word embedding. In: International Conference on Machinery, Materials and Computing Technologies. Atlantis Press (2016)
25. Gunarathne, T.: Hadoop Mapreduce v2 Cookbook. Packet, Birmingham (2015)
26. Gupta, A.: Learning Apache Mahout Classification. Packet, Birmingham (2015)

Amharic Question Answering for Biography, Definition, and Description Questions

Tilahun Abedissa[1] and Mulugeta Libsie[2](✉)

[1] Kotebe Metropolitan University, Addis Ababa, Ethiopia
tilaema@gmail.com
[2] Addis Ababa University, Addis Ababa, Ethiopia
mulugeta.libsie@aau.edu.et

Abstract. A broad range of information needs can often be stated as a question. Question Answering (QA) systems attempt to provide users concise answer(s) to natural language questions. The existing Amharic QA systems handle fact-based questions that usually take named entities as an answer. To deal with more complex information needs we developed an Amharic non-factoid QA for biography, definition, and description questions. A hybrid approach has been used for the question classification. For document filtering and answer extraction we have used lexical patterns. On the other hand to answer biography questions we have used a summarizer and the generated summary is validated using a text classifier. Our QA system is evaluated and has shown a promising result.

Keywords: Question Answering

1 Introduction

To successfully operate within an environment, support decisions and solve problems people need precise information. That stimulates a search for information. However, getting precise information from the ocean of electronic information is not simple. QA is proposed to address this problem. In contrast with traditional document retrieval systems/search engines, which return ranked lists of potentially relevant documents that users must then manually browse through, question answering systems try to give users one or more concise answers in the form of sentences or phrases to natural language questions [6].

QA research had a significant boost when it became the subject of interest of the annual Text REtrieval Conferences (TREC), a series of workshops promoted by the US National Institute of Standards (NIST) with the aim of advancing the state-of-the-art in text retrieval [18]. It also attracts a lot of attention due to the availability of public QA datasets such as the Stanford QA dataset [15], WikiQA [22], and the TREC question answering collections [5]. In the earlier days of QA research, non-factoid questions received relatively little attention [16]. [21] have shown that an existing method for factoid question answering is not sufficient for finding answers to non-factoid questions.

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 301–310, 2019.
https://doi.org/10.1007/978-3-030-26630-1_26

Research on Amharic QA focuses on factoid questions, for example [3, 17] answer place, person/organization name, quantity, and date questions, i.e., factoid type questions by extracting entity names for the questions from documents using rule based answer extraction approach. Amharic non-factoid questions like: biographical questions such as "**አጼ ቴዎድሮስ ማናቸው**?" ("Who is Emperor Tewodros?") compel a system to gather relevant encyclopedic facts about a person/organization from multiple documents; and other complex questions like definition, description, how, or why questions with complex answers oblige reasoning and fusion of multiple "information pieces" from different sources [4].

Even though there are other non-factoid QA systems [10, 14, 26] for other languages, due to the language dependency they only answer their respective language's non-factoid questions. Due to Amharic language morphological complexity [12] and the syntax differences of the question and answer of Amharic non-factoid questions with other languages, other language QA systems cannot answer Amharic non-factoid questions.

Thus, we developed an Amharic QA system (see Sect. 4) that can answer Amharic biography, definition, and description questions. The main contributions of this research are:

- Implement a hybrid approach for automatic question classification.
- Develop lexical patterns that can be used to filter documents and extract candidate answers for definition and description questions.
- Paved a way how Amharic summarizers could be used for non-factoid questions.
- Used text classifier to determine a generated answer for a biography question is actually a biography or not.

The remaining part of this paper is organized as follows. Related works and methodology of the study are given in Sects. 2 and 3 respectively. Section 4 presents a detailed description of our proposed QA model along with its implementation. Section 5 states the evaluation, evaluation criteria, and the empirical results of the proposed system with their discussion. Finally, Sect. 6 concludes the paper with a summary and future works.

2 Related Works

Answering non-factoid questions remains a critical challenge in Web QA. One difficulty is the lexical gap between questions and answers. More knowledge is required to find answers to non-factoid questions [24]. The TREC QA tasks started to incorporate certain categories of non-factoid questions, such as definition questions. One of the first English non-factoid QA system [19] was based on web search using chunks extracted from the original question and ranked the extracted candidate answers using a translation model. Keikha et al. [8] used state-of-the-art passage retrieval method and

Yulianti et al. [25] uses summarization. Recently from the literature we observe that neural networks are widely employed, for example [9, 20, 23]. In addition, using community QA is one way to answer non-factoid questions [16, 22].

3 Methodology

In order to achieve the objective of the study and to understand the state-of-the-art in question answering, books, articles, journals and other publications have been thoroughly reviewed. In addition the following methods are employed.

3.1 Data Collection Methods

To develop an efficient Amharic QA system for biography, definition, and description questions analyzing the characteristics of these question types, their respective answers and the data source used for answering the questions is crucial. Thus, for training and testing the question classifier, questions are prepared from different Amharic documents manually, as well for document retrieval and answer extraction evaluation Amharic documents are collected from different web sites including Amharic Wikipedia[1] using WinHTTrack[2] website copier.

3.2 Performance Evaluation Methods

Evaluation for QA system mainly focuses on the accuracy of the question classifier, document retrieval, and answers returned. Thus, as suggested in [1] our QA system's accuracy is evaluated by comparing the answers returned by the system with human-proposed answers to the same set of questions. Hence, we used precision, recall, and F-score as criteria for measuring the performance.

4 Design and Implementation of the Amharic QA System

As shown in Fig. 1, the QA system consists of document pre-processing, question analysis, document retrieval, and answer extraction components.

4.1 Document Pre-processing

The document pre-processing component performs character normalization, short word expansion, stop word removal, stemming, morphological analysis, and indexing. We used the document pre-processing component that was used by [17], HornMorpho [13] tool for morphological analysis, and Lucene[3] to index the pre processed documents.

---

1 https://am.wikipedia.org/wiki/ ዋናው_ገጽ.

2 HTTrack is a free (GPL, libre/free software) and easy-to-use offline browser utility, http://www.httrack.com.

3 http://lucene.apache.org/core/.

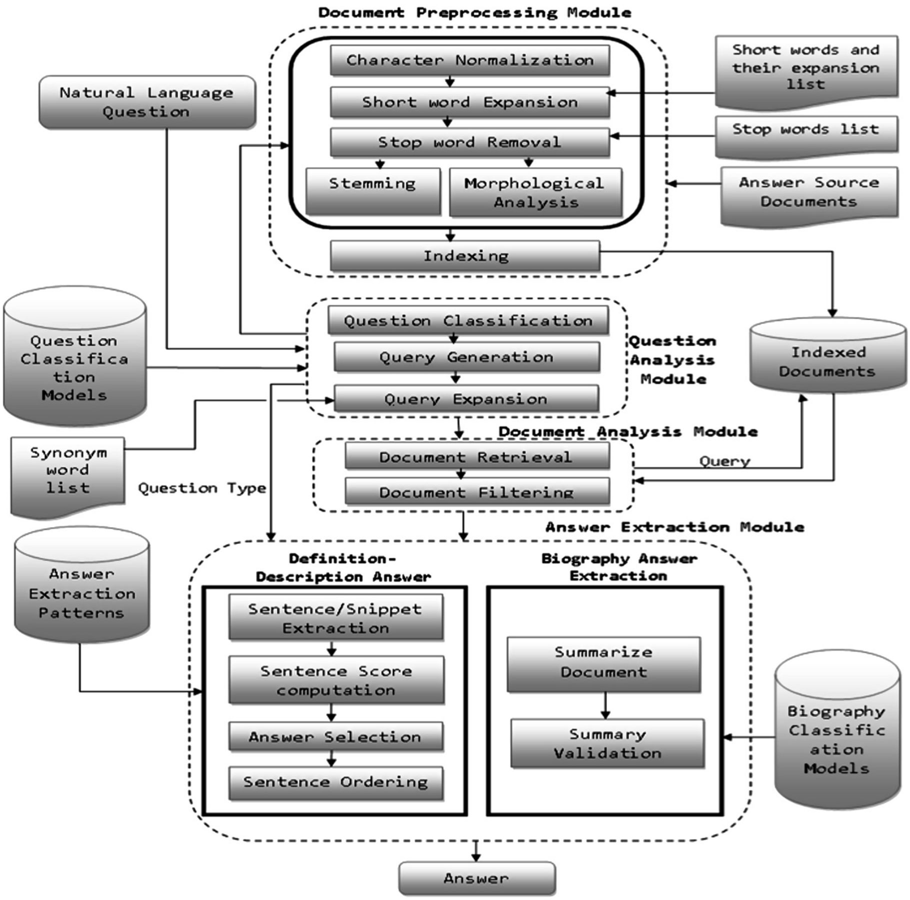


Fig. 1. The architectural design of Amharic biography, definition, and description QA system.

4.2 Question Analysis

This component is dedicated to determine question type, generate query by removing interrogative terms and stop words and expand the query using a synonym list. The question classifier subcomponent determines the type of a question as biography, definition or description. A hybrid approach, i.e., rule based and machine learning techniques are used for the question classification.

A rule based question classifier algorithm determines the question type by using the interrogative terms and question class indicative terms. For example, given the question "በአፍሪካ ውስጥ ድህነትን ለመቅረፍ የትምህርት ጥቅም ምንድን ነው?" ("What is the use of education for eradicating poverty in Africa?") the terms "ጥቅም" (use) and "ምንድን ነው" (what is) indicate that the question is looking for a description. The other approach, machine learning technique predicts the question type of questions based on the training models. The training and classifying tool that we used is Support Vector Machine (SVM) algorithm implementation SVM_light[4] [7]. The svm_learn (learning module) is trained by preparing three question groups (definition versus biography, definition versus description, and description versus biography). Then the svm_classify (classification module) tries to find the class of the question by testing it with the three generated training models for a new question.

4.3 Document Analysis

Once the type of question is identified and the queries have been created and expanded, relevant documents are retrieved from the indexed documents by the document retrieval (implemented using Lucene) sub component. The document retrieval subcomponent simply ranks and extracts documents based on word overlap to the query; it does not consider the question type. That means a document ranked to top might be irrelevant to a question. Due to this reason documents should be filtered before passed to the next module. Thus document filtering subcomponent first locates the question target using target extractor. For biography and definition question types the query itself is taken as a target whereas for description types a target is acquired by removing descriptive indicator terms from the query. Then for definition and description questions if a text is extracted by one of their respective rules (formed as regular expressions), the document is kept; otherwise it is removed. On the other hand, for biography question types, the documents that contain all the target terms in the same order as in the query are taken for further analysis.

4.4 Answer Extraction

Answer extraction is the act of extracting text strings constituting the exact answer(s) to a question from the candidate answer-bearing document(s) retrieved by the document analysis module. Based on the question type, the filtered documents are processed by the biography answer extraction or definition-description answer extraction subcomponents.

4.4.1 Definition - Description Answer Extraction

This sub component finds snippets and/or sentences about the current target then ranks, selects, and orders the candidate answers, finally generates an answer for definition and description questions. First every filtered document is tokenized to sentences, and extracts snippets/sentences by using the lexical patterns (we have prepared about 14 rules for definition and 5 rules for description questions). We observed that many documents, while defining a term they usually state the core concepts about the term in the first two or three sentences. Therefore, for definition questions, if the first sentence of a document is extracted by one of the rules, the next two sentences are included to the candidate answer set.

[4] http://svmlight.joachims.org/.

To select the appropriate sentences/snippets from the candidate answer set we formulate the sentence scoring function given in Eq. 1, i.e., the score of a sentence S is calculated as the sum of the percentage of the query (Q) terms in the sentence, weight of the pattern that identifies the sentence, the reciprocal of the position of the sentence in the document that contains it, and the Lucene score of the document D that contains S.

$$score(S) = \frac{N_{S \cap Q}}{N_Q} + weight(S, P) + \frac{1}{pos(S)} + luceneScore(D, S) \tag{1}$$

Since the position of a sentence does not have any impact for description questions, score of sentence S is computed by the formula given in Eq. 2.

$$score(S) = \frac{N_{S \cap Q}}{N_Q} + weight(S, P) + luceneScore(D, S) \tag{2}$$

At this stage it is likely that some of the snippets or sentences might be redundant and all are not equally important. That is, we have to guarantee that, given sentence A does another sentence B provide any new and relevant information for the purpose of defining or describing a given target? Therefore, one way of determining the similarity of texts is to use word overlap. The more different text fragments share common non-stop words it indicates that they are highly similar [2]. Equation 3 is used to compute the similarity between sentences A and B.

$$sim(A, B) = \frac{|T_A \cap T_B|}{\min(|T_A|, |T_B|)} \tag{3}$$

where *sim(A,B)* is the similarity of the sentences A and B, $|T_A|$ and $|T_B|$ are the number of non-stop tokens in sentences A and B respectively, and $|T_A \cap T_B|$ is the number of common non-stop tokens in A and B.

If the similarity of the sentence S to one of the sentences in the definition/description pool is greater than or equal to 0.7, then skip sentence S; otherwise add it to the definition/description pool. The value of the similarity cut off point is set to 0.7 by doing an experiment. A final answer is generated by selecting 5 top ranked sentences/snippets and then to keep the coherence the candidate answers are ordered in such a way that sentences/snippets beginning with the target term are positioned at the beginning, those that begin with connective terms are in the middle, and sentences which start with other terms are positioned to the end.

4.4.2 Biography Answer Extraction

For biography questions whose focus is a person, important dates in their life (birth, marriage, and death), their major achievements and any other interesting items of note would comprise a ‘correct’ answer. Otherwise if the focus is an organization, the answer should probably include information about when and by whom it was founded, what it serves/ makes/ sells, who owns it, and other organizations it collaborated with. Therefore, to generate an answer the biography answer extraction component first merges[5] the filtered documents, summarizes the merged document using the work of [11], and validates the summary. The evaluation task can be considered as a text classification problem. To do so we trained the SVM_Light by using manually prepared biography and non-biography texts and use the model to classify the summary produced by the summarizer. The summary validation subcomponent returns the summarizer result as an answer for the question if the result of the classifier is greater than or equal to 0.5, otherwise no answer will be displayed. We observed that documents that obtain values greater or equal to 0.5 by the classifier contain the basic information that a biography should contain.

5 Results and Discussion

Precision, recall, and F-score are used as criteria for measuring the performance of the document retrieval and answer extraction components. Since the task of the question classifier is identifying question types, its accuracy is evaluated by the percentage of correctly identified question types. The percentage is computed by taking the ratio of correctly identified questions to the total test questions.

Even though there is no standard Amharic QA dataset for performance evaluation, as shown in Table 1 we have prepared 1500 questions. The question classifier is evaluated using 10-fold cross validation approach. The average result of the question classifier is given in Table 2.

As shown in Table 2 both question classifiers have no problem on identifying biography questions. This is due to the reason that either the definition or the description question types do not have any common interrogative term with biography question. The SVM based classifier performance on definition questions is smaller than that of the other. This is because the interrogative word "ምንድን ነው" (what) is shared by definition and description questions.

The document retrieval performance is measured by 60 queries on two corpuses each containing 600 documents. The first corpus contains the documents that are analyzed by the stemmer and the second contains the same documents as the first but analyzed by the morphological analyzer. The result of the experiment is shown in Fig. 2. From the figure we can see that the recall, precision, and F-score of the document retrieval are greater on the data analyzed by the morphological analyzer than the stemmer.

During our experiment we observed that both the stemmer and the morphological analyzer didn't generate the stem and root of every term in the test documents and queries. Moreover, when the morphological analyzer returns root words it gives in Romanized form, for some words in an attempt to transliterate to Amharic the result contains non Amharic characters. For example, the root returned for the term "ኣይን" (eye) is "'y*" when it is transliterated the result is "ኧይ**". Thus the performance of the stemmer and the morphological analyzer should be improved.

[5] In a biography question an entity may represent different persons/ organizations. To resolve this, our system classifies the filtered documents to different categories and merges the documents in each category separately.

Table 1. Amount of questions used for training and testing.

Type of question	No. of training question set	No. of testing question set	Total
Biography	450	50	500
Definition	450	50	500
Description	450	50	500
Total	1350	150	1500

Table 2. Average performance of the rule based and SVM based question type classifiers on each question type.

Question type	Classification technique	
	SVM based	Rule based
Biography	0.98	0.99
Definition	0.66	0.99
Description	0.86	0.96
Average	0.83	0.98

Biography, definition, and description questions are more difficult to evaluate than factoid questions in that their answers are more complex and can be expressed in different ways. Moreover, answers to these questions can focus on different facets of the concept, entity, person, or organization they are describing. For the evaluation of the answer extraction component a total of 120 questions (40 questions from each type) are prepared. The answers for these questions generated by our system and the answers that are manually constructed are used to compute the precision and recall. The precision, recall, and F-score along with their average value is given in Table 3.

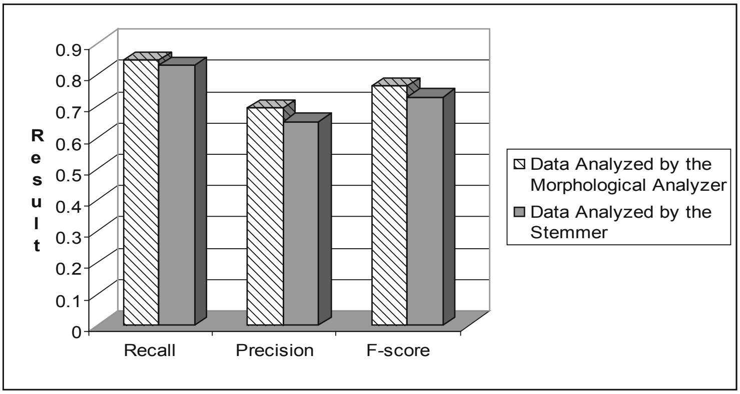


Fig. 2. The recall, precision, and F-score of the document retrieval component on the two data sets.

As shown in Table 3, while conducting the experiment we observed that many documents do not explicitly put the purpose of concepts/entities by using descriptive implication terms rather they put it implicitly. Even in some documents the descriptions are incorporated within their definitions. Due to these reasons the F-score on description questions is less than that of the definition. On the other hand the result that we obtained for the biography questions is highly dependent on the performance of the summarizer.

Table 3. Performance of the answer extraction component

Question type	Recall	Precision	F-score
Biography	0.341	0.895	0.493
Definition	0.725	0.626	0.658
Description	0.762	0.528	0.624
Average	0.609	0.683	0.592

6 Conclusions and Future Works

In this paper we have discussed about an Amharic QA system for biography, definition, and description questions. Question classification, document filtering, and answer extraction patterns are created for their respective purposes. In addition, a machine learning tool has been used for question classification and biography detection. The QA system has been evaluated by preparing a dataset. With minimal amount of tools we were able to achieve a promising result.

Developing co-reference resolution tool, generating lexical patterns automatically, extending this work to more complex question types such as why and how questions, and preparing a standard QA dataset are our future research directions.

References

1. Breck, E., et al.: How to evaluate your question answering system every day and still get real work done. In: Proceedings of the Second International Conference on Language Resources and Evaluation. LREC-2000, Athens, Greece (2000)
2. Manning, C.D., Raghavan, P., Schtze, H.: An Introduction to Information Retrieval. Cambridge University Press, New York (2008)
3. Abebaw, D., Libsie, M.: LETEYEQ (ሌጠየቅ)-a web based amharic question answering system for factoid questions using machine learning approach, M.Sc. thesis, Addis Ababa University (2013, Unpublished)
4. Dietz, L., Verma, M., Radlinski, F., Craswell, N.: TREC complex answer retrieval overview. In: TREC 2017 (2017)
5. Voorhees, E.M., et al.: The TREC-8 question answering track report. In: Text Retrieval Conference, vol. 99, pp. 77–82 (1999)
6. Indurkhya, N., Damereau, F.J. (eds.): Handbook of Natural Language Processing, 2nd edn. Chapman & Hall/CRC, Boca Raton (2010)
7. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 137–142. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0026683
8. Keikha, M., Park, J.H., Croft, W.B., Sanderson, M.: Retrieving passages and finding answers. In: Proceedings of the 2014 Australasian Document Computing Symposium, pp. 81–84 (2014)
9. Lee, K., Salant, S., Kwiatkowski, T., Parikh, A., Das, D., Berant, J.: Learning recurrent span representations for extractive question answering. In: International Conference on Learning Representations (2016)

10. Murata, M., Tsukawaki, S., Kanamaru, T., Ma, Q., Isahara, H.: Non-factoid Japanese question answering through passage retrieval that is weighted based on types of answers. In: Proceedings of the third IJCNLP (2008)
11. Tamiru, M., Libsie, M.: Automatic Amharic text summarization using latent semantic analysis, M.Sc. thesis, Addis Ababa University (2009, Unpublished)
12. Abate, M., Assabie, Y.: Development of Amharic morphological analyzer using memory-based learning. In: Przepiórkowski, A., Ogrodniczuk, M. (eds.) NLP 2014. LNCS (LNAI), vol. 8686, pp. 1–13. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10888-9_1
13. Gasser, M.: HornMorpho: a system for morphological processing of Amharic, Oromo, and Tigrinya. In: Conference on Human Language Technology for Development, Alexandria, Egypt (2011)
14. Trigui, O., Belguith, H.L., Rosso, P.: DefArabicQA: Arabic definition question answering system. In: Workshop on Language Resources and Human Language Technologies for Semitic Languages, 7th LREC, Valletta, Malta (2010)
15. Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P.: Squad: 100,000 + questions for machine comprehension of text. In: Conference on Empirical Methods in Natural Language Processing, pp. 2383–2392 (2016)
16. Savenkov, D.: Ranking answers and web passages for non-factoid question answering, Emory university at trec liveqa (2015)
17. Muhie, S., Libsie, M.: Amharic question answering (AQA). In: 10th Dutch-Belgian Information Retrieval Workshop (2010)
18. Silvia, Q.: Advanced techniques for personalized, interactive question answering on the web. In: Workshop on Knowledge and Reasoning for Answering Questions, Manchester, pp. 33–40 (2008)
19. Soricut, R., Brill, E.: Automatic question answering using the web: beyond the factoid. Inf. Retrieval **9**(2), 191–206 (2006)
20. Wang, W., Yang, N., Wei, F., Chang, B., Zhou, M.: Gated self-matching networks for reading comprehension and question answering. In: 55th Annual Meeting of the Association for Computational Linguistics, pp. 189–198 (2017)
21. Yang, L., et al.: Beyond factoid QA: effective methods for non-factoid answer sentence retrieval. In: Proceedings of the European Conference on Information Retrieval, pp. 115–128 (2016)
22. Yang, Y., Yih, W., Meek, C.: WIKIQA: a challenge dataset for open-domain question answering. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, pp. 2013–2018 (2015)
23. Yu, Y., Zhang, W., Hasan, K., Yu, M., Xiang, B., Zhou, B.: End-to-end answer chunk extraction and ranking for reading comprehension. In: International Conference on Learning Representations (2016)
24. Yulianti, E., Chen, R.-C., Scholer, F., Sanderson, M.: Using semantic and context features for answer summary extraction. In: Proceedings of the 21st Australasian Document Computing Symposium, pp. 81–84 (2016)
25. Yulianti, E., Chen, R.-C., Scholer, F., Croft, W.B., Sanderson, M.: Document summarization for answering non-factoid queries. IEEE Trans. Knowl. Data Eng. **30**(1), 15–28 (2018)
26. Zhang, Z., Zhou, Y., Huang, X., Wu, L.: Answering definition questions using web knowledge bases. In: Dale, R. et al. (eds.) IJCNLP (2005), LNAI, vol. 3651, pp. 498–506. Springer, Heidelberg (2005)

Dictionary Based Amharic Sentiment Lexicon Generation

Girma Neshir Alemneh[1(✉)], Andreas Rauber[2], and Solomon Atnafu[3]

[1] IT Doctoral Program, Addis Ababa University, Addis Ababa, Ethiopia
girma1978@gmail.com
[2] Institute of Information Systems Engineering,
Technical University of Vienna, Vienna, Austria
rauber@ifs.tuwien.ac.at
[3] Department of Computer Science,
Addis Ababa University, Addis Ababa, Ethiopia
solomon.atnafu@aau.edu.et

Abstract. Sentiment analysis is a hot research area with several applications including analysis of political opinions, classifying comments, movie reviews, news reviews and product reviews. To employ rule based sentiment analysis, sentiment lexicon is required. However, manual construction of a sentiment lexicon is time consuming and costly for resource-limited languages. To reduce development time and costs, we propose an algorithm for constructing Amharic sentiment lexicons. The proposed approach transfers sentiment labels from a one language (e.g. English) to resource-limited language (e.g. Amharic) relying on Amharic-English dictionary. Using Bilingual/Monolingual dictionaries as a bridge, two Amharic sentiment lexicons are automatically generated the first based on SO-CAL polarity lexicon, the second on SentiWordNet 3.0. For each Amharic word, the algorithm finds the meaning of the corresponding English word(s). For these English words, sentiment information is searched from the aforementioned sentiment lexicon(s). The weighted average of returned sentiment values, part of speech and gloss information is assigned to the Amharic word. Lexicons of 5683 and 13679 words, respectively, are generated automatically and evaluated subsequently.

Keywords: Lexicon generation algorithm · Sentiment analysis · Amharic sentiment lexicon · Sentiment lexicon generation

1 Introduction

Sentiment Analysis or opinion mining is the process of detecting subjective or emotional indicators in the text. That is finding the attitude of the author towards a particular topic, event, entity, aspect or issue. Nowadays, due to the advancement of web technology, users are not only consumers of online information but also they are producer of information. Most of these data carry opinions that need further analysis to detect its polarity levels (i.e. negative and positive). For instance, political parties want to gather opinions of voters, or a government wants to collect summarized opinions for improvement towards their new policy or existing public services. Thus, a sentiment lexicon is a

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 311–326, 2019.
https://doi.org/10.1007/978-3-030-26630-1_27

valuable resource that should be readily available for opinion mining tasks. So far, sentiment analysis research works were restricted to dominant languages (mostly English). Currently, opinionated language resources are increasingly generated and used in languages other than English. Amharic is one of these resource-limited languages that has started to get attention. But, still it has no sufficient linguistic resources for sentiment analysis such as part of speech taggers and sentiment lexicons. A sentiment lexicon contains a list of opinion words where each term has been assigned positive and negative values, and sometimes, part of speech and glosses definition are included. This work aims to address the following research questions: (a) how do we develop Amharic sentiment lexicons by employing monolingual and cross-lingual resources automatically? (b) how do we evaluate and validate the quality and accuracy of Amharic polarity lexicon's subjectivity clues? (c) how can we measure the accuracy and the coverage of the generated Amharic Sentiment Lexicons for detecting the subjectivity of Amharic text?

The remaining part of the article is organized as follows: in Sect. 2, we provide an overview of related work. Section 3 describes the data sets and the proposed methods. In Sect. 4, we present results and discussions, followed by a summary of our conclusions in Sect. 5.

2 Related Work

This section briefly presents a few key related works. Gebremeskel [6] manually built Amharic polarity Lexicon of 900 terms. This was, again manually, extended by Tilahun to 1,000 terms [7]. In these works, the polarity value of each term in this Amharic sentiment lexicon was labeled either negative 2 or positive 2 without using numerical value in between or no nominal fine-grained levels of sentiment strength granularities or no part of speech information in the lexicon. The authors used this lexicon for lexicon based (rule based) Amharic opinion mining tasks. However, the effort required to further extend these lexicons manually is prohibitive, calling for automated methods, benefiting from the efforts invested in building such dictionaries in other languages. Taboada et al. [1] propose a semantic orientations calculator (SO-CAL) for lexicon based subjectivity classification. SO-CAL utilizes sentiment dictionaries and also includes intensification and negation to calculate and assign semantic orientation of words in the text. The major strength of this work is that the lexicon performed consistently in any domain to extract word level/sentence level opinion in text and also presented the process of creating and evaluating sentiment dictionaries using Mechanical Turk. SO-CAL performed well in sentiment analysis of user's blog postings. We will use SO-CAL as one of the baseline sentiment lexicons for assigning sentiment values to our translated terms.

Medagoda et al. in [2] built SentiWordNet 3.0 for Sinhala language by mapping English SentiWordNet 3.0 relying on the online English-Sinhala dictionary. The English words in the dictionary were used as search key to generate a lexicon for the corresponding translated Sinhala sentiment lexicon. If the translated Sinhala word is found, then it is inserted with its polarity value and POS tag into the Sinhala sentiment lexicon, otherwise search proceeds for the next English word in the dictionary. The final Sinhala Sentiment Lexicon contains 5973 adjectives and 405 adverbs. This lexicon was evaluated using 2,083 manually classified news article opinions collected from Sinhala online newspaper. Based on the different classification methods using this lexicon, the accuracy achieved was between 56–60%. As this evaluation result is below the baseline, handling negations, multiword with negations and context sensitivity were suggested to be addressed in the forthcoming works to improve the performance of the lexicon generation. We will use a similar approach to generate Amharic Sentiment Lexicon, considering the idiosyncrasies of the Amharic language.

3 Methods

3.1 Data Sets and Lexical Resources

This subsection describes the main data sets and lexical resources used for building and evaluating Amharic Sentiment Lexicons.

(a) English SentiWordNet 3.0: SentiWordNet is automatically built based on synsets of wordnet version 3.0 [3]. This lexicon contains 72,092 terms with part of speech, id number, PosScore, NegScore, SynsetTerms and Gloss. The pair (POS, ID) uniquely identifies a WordNet (3.0) synset. The values PosScore and NegScore are the positivity and negativity scores (in the range of 0 to1) assigned to each entry of the synset. The objectivity score is computed as: ObjScore = 1 − (PosScore + NegScore) SynsetTerms column reports the terms, with sense number, belonging to the synset (separated by spaces). We selected to port this lexicon into Amharic because of its extensive coverage.

(b) SO-CAL Polarity Lexicon: SO-CAL refers to the Semantic Orientation CALculator, a tool to extract sentiment from text. It has a long history of development. We use this lexicon [1] as a baseline. SO-CAL contains 10126 words with polarity value ranges from −5 to +5. The lexicon is categorized into adjectives, Adverbs, nouns, verbs and intensifier word lists. This lexicon has been extensively tested showing good performance in different domains.

(c) Amharic-English Dictionary: This is a dictionary that works in one direction (i.e. Amharic to English). For each Amharic word, it contains part-of-Speech tag(s) and corresponding meanings in English. The English text might be a single word, phrase, or list of synonyms. The total size of this dictionary is more than 31000 terms which are obtained by merging Amharic-English dictionary (12700 Amharic words) [4], Amsalu_Aklilu Amharic-English Dictionary (16231 words) and Amharic-English dictionary by SelamSoft Plc (1,075 words). This dictionary serves as a bridge to propagate sentiment from English sentiment lexicon to Amharic sentiment lexicon.

(d) Amharic-Amharic Dictionary: From more than 30 thousand entries of Amharic lexical wordlist in [4], we built 33965 Amharic-Amharic dictionary automatically, where this dictionary is served as gloss source for terms in the generated Amharic Sentiment Lexicon.

(e) Facebook Comments Data Set: This dataset consists of 2500 sentence/phrase level sentiment annotated facebook news users' comments collected from the Government Office Affairs Communication (GOAC) between 2008 and 2010. News that received high view counters/frequent comments were selected as "hot topics" and the associated comments labeled by professionals into either positive or negative sentiment.

(f) Amharic Web Corpus: This is an automatically crawled collection of web documents consisting mainly of news, politics and religion documents. It was collected and tokenized, automatically part-of-speech tagged and published under the HaBiT project [5]. The file is a single xml file of size 421 MB. It contains 20 million tokens. We will use this corpus for evaluation of the sentiment annotations.

3.2 Automatic Sentiment Score Calculation

To build an Amharic Sentiment Lexicon relying on bilingual dictionaries, we tried to transfer subjectivity or polarity information from polarity lexicon of resource rich language (i.e. English sentiment lexicon) to resource limited language (i.e. Amharic). Specific to this approach, an algorithm is developed to transfer labels from English SentiWordNet 3.0 and SO-CAL to Amharic Sentiment Lexicon.

As shown in Fig. 1 below, the Amharic-English bilingual dictionary is employed as a bridge between the two languages. The proposed approach of building Amharic Sentiment Lexicon is similar to Sinhala language in [2]. However, our approach is different at

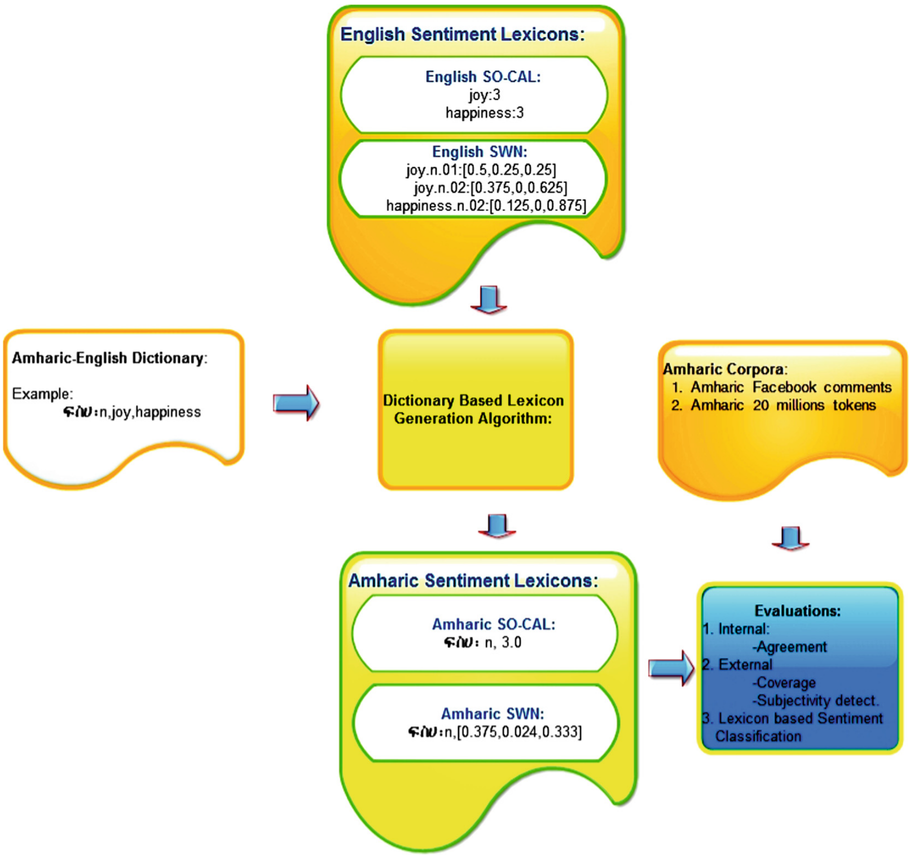


Fig. 1. Dictionary based Amharic Sentiment Lexicon Building Framework

least in three features: (1) algorithm searches for each Amharic word as key rather than English words in the Amharic-English bilingual dictionary. That is why Amharic-English dictionary is used instead of English-Amharic dictionary. (2) our algorithm uses average weighting of synonymous of English to tokens to assign the sentiment score for the corresponding Amharic term in the bilingual Amharic-English dictionary and (3) the approach in this work will also handle negation words that appear in the Amharic text.

Consider Amharic word w_{a_i} in the Amharic-English Dictionary d_{ae} has corresponding meanings in English word(s), $w_{e\_i\_j}$. The Amharic Sentiment Lexicon is denoted by s_a, the English language sentiment lexicon by s_e. As part of preprocessing, we normalized all Amharic words in the Amharic-English dictionary by replacing varied alphabets of the same sound with identical symbols. Moreover, a stemmer is applied during evaluations. Then the algorithm is represented:

BuildAmharicLexicon(*s_e*)
Input: *s_e (EnglishSentiment Lexicon i.e. either SWN or SO-CAL)*
Output: *s_a (generated Amharic Sentiment Lexicon)*
Format for s_a is {w_{ai} as key: scores, part of speech, gloss as list} pair.
Initialize: *d_{ae}–load Amharic-English Dictionary from File and initialize other vars*

1. *For each w_{a_i} in the d_{ae}:*
   a. *Search its {$w_{e\_i\_1}$, $w_{e\_i\_2}$,..., $w_{e\_i\_j}$,..., $w_{e\_i\_n-1}$, $w_{e\_i\_n}$}*
   b. *For each synonym word $w_{e\_i\_j}$, search it in s_e*
      i. *If found, return all the polarity values(+,-,objective) of the synsets*
      ii. *Keep them in a list L*
   c. *Find length of +,-, objective list in L*
   d. *Compute the average weighted polarity in L using equation (1)*
   e. *Assign the average weighted polarity, part of speech, and gloss to w_{a_i} in s_e. That is,*
      $s_a[w_{a\_i}]$=[[Average weighted polarity], part of speech, gloss]

return s_a

Listing-1. Dictionary based Algorithm

Algorithm Description: In Listing-1, for each Amharic word in the bilingual dictionary, the aggregated sentiment values of corresponding English term(s) in SentiWordNet 3.0 (or SO-CAL) are assigned to the Amharic word if terms are found in the source lexicon. Finally, the Amharic Sentiment Lexicon contains each Amharic term, its average weighted polarity values, its corresponding part of speech and gloss definitions are assigned [8].

For each Amharic word w_{a_i} in the Amharic-English dictionary, there are list of English words($w_{e\_i\_1}$, $w_{e\_i\_2}$,. $w_{e\_i\_j}$., $w_{e\_i\_n-1}$, $w_{e\_i\_n}$) in the dictionary. The assumption is that the first English word $w_{e\_i\_1}$ has more closer in meaning to Amharic word w_{a_i} than the second word $w_{e\_i\_2}$ and in turn $w_{e\_i\_2}$ is more prominent to w_{a_i} than $w_{e\_i\_3}$ and so on. So more weight is given to the first word $w_{e\_i\_1}$ and least weight is given to the last word $w_{e\_i\_n}$. On the basis of this assumption, the weighted average sentiment of English words is computed and then assigned the resulting sentiment value to the corresponding Amharic word w_{a_i}. Then, the sentiment score (w_{a_i}) is given by:

$$\begin{aligned} &= (\mathrm{score}(w_{e\_i\_1})X(n-1+1) + \mathrm{score}(w_{e\_i\_2})X(n-2+1) + ..\,\mathrm{score}(w_{e\_i\_j})X(n-j+1).. \\ &\quad + \mathrm{score}(w_{e\_i\_n})X(n-n+1))/n \\ &= \frac{\sum_{j=1}^{n} score_{eij}(w_{eij})x(n-j+1)}{n} \end{aligned} \tag{1}$$

where n is the number of English words $w_{e\_i\_1}, w_{e\_i\_2}, .., w_{e\_i\_n-1}, w_{e\_i\_n}$ and score_{eij} is in turn refers to:

$$score_{eij} = \left\{ \begin{array}{ll} sentimentScore(w_{eij}), & if\ the\ lexicon\quad is\quad SO-CAL \\ sumOfSentimentScore(Synset(w_{eij}), & if\ the\ lexicon\quad is\quad SWN \end{array} \right\}$$

It should be noted that the gloss definitions are not found in Amharic English dictionary; rather it is propagated from another Amharic–Amharic dictionary. However, there are some terms without gloss definitions as they were not found in the Amharic–Amharic Dictionary [4].

The Amharic sentiment lexicons are generated and used under a few assumptions. (1) The senses of words in the bilingual dictionaries of the two languages are considered to be same. Thus, the sentiment score of an English word in sentiment analysis of English text is assumed to be identical to the sentiment score of an Amharic word in Amharic texts. (2) The parts of speech in both languages in the bilingual dictionary are assumed to be equivalent [2].

From English SentiWordNet 3.0 and SOCAL, two corresponding Amharic Sentiment lexicons are automatically generated. The sizes of these newly built Amharic lexicons are 13679 term entries and 5683 terms, respectively. The experimental results and evaluations are discussed in the subsequent section.

4 Results and Discussion

In this section, the experimental results relying on the generated Amharic sentiment lexicons are discussed. Each of the generated Amharic lexicons is evaluated in two ways using external data as well as intrinsic evaluation and mutual comparison of agreement. Finally the usefulness of the Amharic lexicons is tested in lexicon based sentiment classification of Amharic News Comments. To increase the string matching accuracy in generating the evaluation scores, we apply basic preprocessing: tokenization, stopword removal, punctuation mark removal and stemming.

(i) Tokenization: Amharic words are separated by space as delimiter. Thus, Amharic text is tokenized into words if space is found in between tokens.

(ii) Punctuation mark removal: Amharic punctuation marks such as ።(full stop), ፤ (double colon), ፣ (single quote), ፡ -(preface colon), ‹‹ ››(double quotes) and so on should be removed.

(iii) Normalization: Ethiopic script contains redundant symbols representing the same sound. These symbols are substituted by the symbol on the right. E.g. ('ሀ,ሃ,ሐ,ሓ,ኀ,ኃ,ኻ=>ሀ)፣ (ሰ,ሠ=>ሰ)፡ (ፀ,ጸ=>ፀ)፡ (ዐ,አ,ኣ,ዐ,ዓ=>ዐ)፡ (ቆ,ቈ,ቖ=>ቆ)፡ (ቊ,ቍ=>ቊ)፡(ኮ,ኰ=>ኮ)፡ (ጎ,ጐ=>ጎ)፡(ኋ,ዃ,ሗ=>ኋ)). Where, the arrow(=>) means "replaced by". That is, if either of the symbols on the left side found in the text, then it is replaced by the symbol right of the arrow(=>).

(iv) Stopword removal: For stopword removal, stopwords identified in [5] are used.

(v) Stemming: As Amharic is highly morphologically rich language, we develop a shallow stemmer by specifying some conditions to enforce and normalize the morphological variations of Amharic words to a common base form/stem. The stemmer is composed of a set of rules of the regular expressions to remove prefixes and suffixes. During stemming, if Amharic word has a match to one of the patterns of regular expressions, then the word is reduced to its corresponding stem by removing the matched suffix or/and prefix. For example, for the Amharic words such as 'ቆንጆው'/the one who is handsome/masculine/, 'ቆንጆውን' /someone who is handsome/masculine/, object form/, 'ቆንጆዎቹን'/those who are beautiful, ' ቆንጆዎች'/ those who are beautiful, 'ለቆንጆዋ'/to the beautiful one/feminine/and soon. Finally, these Amharic words are reduced to the common stem/root by converting the symbols to corresponding consonant of Amharic/'ቅ-ን-ጅ'/by our stemmer. More sample of Amharic stemmed words are attached in Appendix A of this report.

4.1 External to Lexicon Evaluation

In this research, external evaluation is carried out in two ways.

(a) Subjectivity detection: Each term in the lexicons is counted only if it is present in a particular Amharic news comments. If more terms of the lexicon found in the Amharic comment, then the comment is subjective, otherwise objective. The usefulness of automatically generated lexicons is evaluated by their accuracies in detecting subjectivity in Amharic texts. These lexicons are used to identify subjective (positive or negative) Amharic text from objective (neutral) Amharic text. The accuracy of subjectivity detection increases as we apply stemming along with the sentiment lexicons. As Amharic language is morphologically rich, the variations of words derivations are normalized to the same stem. Usually, the resulting stem will be in root form. This decreases the string mismatches which arise due to morphological variation of Amharic texts.

Table 1. Lexicons' subjectivity detection rate in percent on Facebook comments data

Sentiment Lexicons	Detection with No Stem (%)	Detection With Stem (%)
Amharic Manual (baseline)	43.23	93.56
Amharic SOCAL	31.28	96.65
Amharic SWN	75.83	99.33

Table 1 shows that the generated Amharic Sentiment SentiWord Net Lexicon outperforms the other lexicons, resulting in 99.33% correct subjectivity detection. The rate of detecting the subjectivity of Amharic news comments using Amharic Sentiment SO-CAL is, in turn, outperforming the manual Amharic sentiment lexicon (93.56%). This subjectivity detection result shows that each of the Amharic sentiment lexicons are outperforming above the subjectivity detection result of the baseline lexicon (the manually generated Amharic sentiment lexicon) by 5.77%. One of the reasons for this might be the size of SWN larger than the manual one. Thus, the degree to which the entries of SWN to appear in the Amharic news comment is higher than the other lexicons.

(b) Evaluation of lexicons by its coverage count (or in percent): This way of evaluating the generated lexicons is to measure the size of lexicons in terms of their coverage count in general corpus containing 20 millions tokens and 2500 Amharic facebook news comments.

Table 2. Lexicons' coverage (positive/negative count and in percent) on 2500 Amharic Facebook comments and 20 million tokens of Amharic web corpora

	2500 Amharic comments		20 millions Amharic tokens	
Lexicons	coverage(+,−) count	%	coverage(+,−) count	%
Manual	[4399, 2995]	31.95	[2713167, 2161501]	25.01
SOCAL	[5738, 3953]	41.87	[4169817, 3391213]	38.88
SWN	[9447, 4803]	61.57	[6645592, 4006072]	54.77

Discussion: The results in Table 2 verify that the coverage of the sentiment lexicon based on SWN is larger than the other lexicons on both data corpora. The other aspect that we can verify from this analysis is that the number of positive and negative opinion words are balanced in both SWN and SOCAL. However, the coverage of the Amharic lexicons in both corpuses is below the average benchmark semantic lexical coverage in other languages Welsh (25%) to Arabic (88%) [9], even though, it is very difficult to compare the coverage of Amharic sentiment lexicons with the coverage of these general purpose semantic lexical resources. Moreover, there is difference in languages intrinsic characteristics.

4.2 Internal to Lexicon Evaluation

Internal evaluation of the lexicon is the process of counting common positive and negative opinion terms in the two lexicons. The number of common terms in the two lexicons is expressed in percentage to show the extent of the agreement (overlap) between the generated lexicons and the manual lexicon (baseline). To evaluate the results of the proposed approach, we computed the agreement between lexicons.

Table 3. The Agreement (in percent) between Lexicons

	Amharic Sentiment Lexicons	Agreement (%)
1	SOCAL and Manual Lexicon	70.80
2	SWN and Manual Lexicon	59.49
3	SOCAL and SWN	66.40

Discussion: Table 3 depicts that the Amharic sentiment lexicon from SOCAL generated through dictionary approach has agreement or overlap of 70.80% with manual Amharic sentiment and it overlaps 66.40% with Amharic sentiment generated from SentiWord Net. On the other hand, Amharic Sentiment from SWN agrees for 59.49% of all terms with the manual Amharic Sentiment lexicon. Although the size of Amharic Sentiment from SWN is much larger (more than double) than Amharic Sentiment from SOCAL, the latter is more consistent in its sentiment scores with the other lexicons. The main purpose of finding agreement between lexicons is to know the extent to which the lexicons overlap.

The disagreement level of English Sentiment lexicons are compared in [10]. In this comparison, the agreement level of SWN is better with Harvard General Inquirer (77%) than other English Sentiment lexicons (MPQA, Opinion Lexicon, LIWC). Amharic generated sentiment lexicon is below the agreement levels reported for the English language resources. The reason for this is that it is very difficult to compare Amharic Lexicon with English lexicon as the two languages are very different in morphology apart from difference in cultural connotations. So, as Amharic is morphologically rich where there could be more variations of terms in the lexicons than the terms in English lexicons. This might decrease the extent of agreement of the Amharic sentiment lexicons.

4.3 Lexicon Based Sentiment Classification

Besides the evaluations in the earlier subsections, we will also evaluate the usefulness of the generated lexicons and their combinations for sentiment classification of Amharic facebook news comments. Prior to sentiment aggregations of Amharic texts, we apply basic text preprocessing (tokenization, punctuation mark removal, normalizing Amharic script symbols, stopword removal, spelling corrector, stemming, etc.). The effect of stemming and negation detection technique on Amharic text is investigated to increase the accuracy of lexicon based Amharic sentiment classification.

Table 4. The accuracy (in percent) of Lexicons for Sentiment Classification

Amharic Senti.Lexicons	Accuracies (%)		
	NoStem+NoNeg.	*Stem+NoNeg.*	*Stem+Neg.*
Manual (baseline)	16.7	42.9	52.7
SOCAL	14.6	46.3	50.8
SWN	30.9	50.1	54.7
Manual + SOCAL	37.2	63.4	72
Manual + SWN	49.9	65.9	74.6
SOCAL + SWN	43.7	66.6	73.5
Manual +SOCAL + SWN	53.7	75.8	**86.2**

Discussions: The usefulness of lexicons is evaluated in terms their accuracies of classifying sentiment of Amharic facebook news comments as shown in Table 4. In general, the results in Table 4 also reveal the effect of applying stemming and negation handling on Amharic texts to boost the performance of sentiment classification. The automatically generated lexicon from English SWN outperforms the performance of the manual (baseline) for classifying sentiment of Amharic texts. On the other hand, the manual lexicon (with stemming and negation handling) in turn outperforms the automatically generated lexicon from English SO-CAL. However, Amharic sentiment lexicon (with stemming and without negation handling) from English SO-CAL relatively performs better than the Amharic manual sentiment lexicon in classifying sentiment of Amharic texts. The combination of the automatically generated sentiment lexicon (SWN) with Manual sentiment lexicon (baseline) outperforms (with accuracy of 74.6%) the other combinations of lexicons for sentiment classification of Amharic texts. The automatically generated lexicons (SO-CAL + SWN) perform well (with accuracy of 73.5%) for classification sentiment in Amharic texts. Yet another combinations of the three lexicons(SO-CAL + SWN + Manual) outperform (with accuracy of 86.2%) the other lexicons or their combinations.

4.4 Error Analysis

Unfortunately, it is challenging to trace the sources of errors in the automatically generated lexicons. Let us try to point out some of the causes for the subsets of generated errors in the automatically translated Amharic sentiment lexicons. We present manually identified errors in this subsection:

(a) On-spot analysis of errors detected in lexicons: In the automatically generated lexicons, the generated errors in sentiment transfer from source language terms to target language terms are mainly caused by the following issues: (i) We identified mistranslation of Amharic terms in SWN by the bilingual dictionary. For example, in Amharic SWN, the word ሸጋ("SHEGA") is incorrectly translated into target 'hibernation' and it is wrongly assigned with negative sentiment. But, the correct meaning of the word ሸጋ("SHEGA") means "nice" which has positive sentiment value. (ii) We got correctly translated terms which are assigned opposite in polarity to source terms. For example, in Amharic SWN, the word ቆራጥ("QORAT") correctly translated into'courageous decisive in manner'. The translation is correct, but wrong sentiment value (opposite) is assigned. (iii) We discovered few terms are correctly translated, the same sentiment polarity but different sentiment strength. For example, ሰይጣን("SEYTAN") means "devil" in Amharic SWN. It is correctly translated and assigned negative sentiment value but the sentiment strength is small. (iv) We found some terms in SWN which are correctly translated but different sentiment due to cultural connotations. For example, the term እርካሽ("ERKASH") means "cheap" and it is assigned positive sentiment. In English, this might be correct, but in Amharic the word እርካሽ("ERKASH") has negative connotation in that it refers to something which is sinfulness and lower in quality. The level of these type errors can be minimized by incorporating context dependent lexicon generated from Amharic corpus. This leads to another venue of future researches.

(b) Analysis of Incorrectly Detected Amharic News Comments: We identified the associated reasons why the sentiment/subjectivity of Amharic News Comments are wrongly detected. The subsets of reasons which cause wrong subjectivity/sentiment detection include: (i) We discovered some Amharic comments which contains sarcasm and idioms. The nature of these comments is difficult to detect its sentiment relying on ordinary lexicon. For example, 'የራሷ አሮባት የሰው ታማስላለች/solve your own problem before talking about others/is an Amharic proverb that cannot be translated directly relying on ordinary dictionary. That is why it is wrongly classified by the sentiment lexicons. (ii) Besides handling negation, we discovered that further formulations of linguistic rules (e.g. intensifiers, contrast rules, conjunction rules) are required to handle wrongly detected Amharic news comments. For example, 'ኃይሌ ገ/ስላሴ በሩጫ ጎበዝ እንጂ ስለኢትዮጵያዊ መንፈስ ሊሰበክ አይችልም'/H/G/Silasie is the best runner, however, he cannot preach about Ethiopianism/. In this example, the sentiment computation is failed as the semantic orientation of the text is diverted by contrasting word እንጂ /however/. The text next to this contrasting word has dominant sentiment than the phrase before it. (iii) We identified some Amharic comments which are wrongly annotated by human annotator. For example, 'በጣም አሳዛኝ ነው'/it is very tragedy/. The sentiment of this comment is labeled wrongly as positive by the human annotator. Thus, the labeled data should be reviewed for correction. (iv) We also found some Amharic news comments which are detected wrongly as their meanings are implicit where their interpretations are connected to pragmatics. Such context dependent text connotations are difficult to handle by explicitly using our generated lexicons. For example, the Amharic news comments ''ፍትህ ለአማራ''/Justice to Amhara/is assigned negative sentiment by annotator, but lexicon based classifier wrongly detected it as positive. The reason is that the lexicon based classifier is limited in handling the contextual meaning of the comment in place. This comment is primarily connected to the context and implicitly associated to the meaning behind the original news post.

5 Conclusions and Recommendations

Amharic sentiment lexicon is one of the resources required for Amharic sentiment analysis. Yet, extensive lexical resources are expensive to build. To remedy this problem, we propose a dictionary based approach for generating Amharic Sentiment lexicon. It requires a bilingual dictionary to propagate polarity information from sentiment lexicon of source language to target language. This approach can be used to generate large scale sentiment lexicons. However, it generates general sentiment lexicons that may lack accuracy for sentiment analysis. It is unable to handle cultural and language specific connotations in a particular language. So to address these issues, we proposed another approach which is a corpus based approach to be done in the forthcoming work in our project. Then once processed and refined, dictionary based approach generates two Amharic lexicons: Amharic Sentiment Lexicon_SOCAL(5683) and Amharic Sentiment Lexicon SWN(13679).

The lexicons generated using dictionary based approach are evaluated by agreement (internal), coverage (external), subjectivity detection rate (external) and its performance in lexicon based Amharic text sentiment classification. The Amharic sentiment lexicon generated from SWN is not only good in internal evaluation (i.e. it has acceptable agreement rate with both the manual and SOCAL lexicons) but it also has higher coverage in both test corpora than the other two sentiment lexicons. Moreover, the lexicon based sentiment classification of Amharic Sentiment lexicon from SWN outperforms the other Amharic Sentiment lexicon(including the baseline).

This work demonstrates that it is possible to automatically generate sentiment lexicons relying on available bilingual dictionaries to minimize time and labor cost of manual sentiment lexicon preparation. The generated lexicon expected to get sufficient sentiment lexicon size by porting from resource rich language. Some of the contributions of this work are briefly summarized below:

– Being an automatic approach, the algorithm developed reduces cost and time of labeling terms in sentiment lexicon.
– The approach developed is generic enough that it can be adapted to generate sentiment lexicon to other resource limited languages.
– The entries in our lexicons contain part of speech information that can be applied or utilized in other linguistic tasks of interest.
– The approach can also be adapted to other tasks of natural language processing including information extraction, multilingual semantic lexicons, question and answering, just to name a few.
– The combination of all generated lexicons with the manual lexicon achieves best sentiment classification performance on Amharic texts.
– The code and related resources will be available online for research communities.

Yet, the generated sentiment lexicons may lack accuracy for sentiment analysis of a particular language context where the approach potentially does not sufficiently capture the cultural and language specific connotations in a particular language. To address these issues, we may need to consider corpus based approaches that capture and incorporate semantic information on the specific meaning of a term in a given context to provide higher precision in sentiment score assignment for each particular instance.

Appendix A. Sample of Amharic Stemmed Words and Its Variant Word Forms

Three sample words are selected. These are from verb, noun and adjective categories. The base forms of these sample words include ሰበረ//seBeRe/means ‘he break something’, ቤት/Beet/means ‘home’ and ቆንጆ/qonJo/means ‘beautiful’, respectively. Table 5 below shows the different forms of these words and their corresponding stems or roots.

Table 5. Sample of the different variant word forms and their corresponding stems.

Surface Word	SERA form	Stem	Root
'ሰበርኩ'	seBeRku	'ስብር'	'ስብር'
'ሰበርክ'	seBeRk	'ስብር'	'ስብር'
'ሰበርሽ'	seBeRX	'ስብር'	'ስብር'
'ሰበረ'	seBeRe	'ስብር'	'ስብር'
'ሰበረች'	seBeRec	'ስብር'	'ስብር'
'ሰበርን'	seBeRn	'ስብር'	'ስብር'
'ሰበራችሁ'	seBeRachu	'ስብር'	'ስብር'
'ሰበሩ'	seBeRu	'ስብር'	'ስብር'
'ሰበርኩት'	seBeRkut	'ስብር'	'ስብር'
'ሰበርከው'	seBeRkeW	'ስብር'	'ስብር'
'ሰበርሽው'	seBeRXW	'ስብር'	'ስብር'
'ሰበረው'	seBeReW	'ስብር'	'ስብር'
'ሰበረችው'	seBeRecW	'ስብር'	'ስብር'
'ሰበርነው'	seBeRneW	'ስብር'	'ስብር'
'ሰበራችሁት'	seBeRachut	'ስብር'	'ስብር'
'ሰበሩት'	seBeRut	'ስብር'	'ስብር'
'እሰብራለሁ'	IseBRaLehu	'ስብር'	'ስብር'
'ትሰብራለህ'	tseBRaLeh	'ስብር'	'ስብር'
'ትሰብሪአለሽ'	tseBRiaLeX	"ትሰብሪ'አለ"	'ስብር'
'ይሰብራል'	YseBRaL	'ስብር'	'ስብር'
'ትሰብራለች'	tseBRaLec	'ስብር'	'ስብር'
'እንሰብራለን'	InseBRaLen	'ስብር'	'ስብር'
'ትሰብራላችሁ'	tseBRaLachu	'ስብር'	'ስብር'
'ይሰብራሉ'	YseBRaLu	'ስብር'	'ስብር'
'አልሰብርም'	aLseBRM	'ስብር'	'ስብር'
'አልሰበርኩም'	aLseBeRkuM	'ስብር'	'ስብር'
'አትሰብርም'	atseBRM	'ስብር'	'ስብር'
'አትሰብሪም'	atseBRiM	'ስብር'	'ስብር'
'አንሰብርም'	anseBRM	'ስብር'	'ስብር'

'አትሰብሩም'	atseBRuM	'ስብር'	'ስብር'
'አይሰብርም'	aYseBRM	'ስብር'	'ስብር'
'አልሰበሩም'	aLseBeRuM	'ስብር'	'ስብር'
'አልሰበረችም'	aLseBeRecM	'ስብር'	'ስብር'
'አይሰብሩም'	aYseBRuM	'ስብር'	'ስብር'
'ሰባበሩ'	seBaBeRu	'ስብር'	'ስብር'
'ሰባበርን'	seBaBeRn	'ስብር'	'ስብር'
'ሰባበረች'	seBaBeRec	'ስብር'	'ስብር'
'ሰባበርሽ'	seBaBeRX	'ስብር'	'ስብር'
'ሰባበርህ'	seBaBeRh	'ስብር'	'ስብር'
'ሰባበሩ'	seBaBeRu	'ስብር'	'ስብር'
'ሰባበረ'	seBaBeRe	'ስብር'	'ስብር'
'ሰባበርኩ'	seBaBeRku	'ስብር'	'ስብር'
'አልሰባበሩም'	aLseBaBeRuM	'ስብር'	'ስብር'
'አልሰባበረችም'	aLseBaBeRecM	'ስብር'	'ስብር'
'አልሰባበርንም'	aLseBaBeRnM	'ስብር'	'ስብር'
'አልሰባበረም'	aLseBaBeReM	'ስብር'	'ስብር'
'አልሰባበርክም'	aLseBaBeRkM	'ስብር'	'ስብር'
'አልሰባበርንም'	aLseBaBeRnM	'ስብር'	'ስብር'
'አትሰባብሩትም'	atseBaBRutM	'ስብር'	'ስብር'
'አንሰባብረውም'	anseBaBReW	'ስብር'	'ስብብር'
'አይሰባብሩትም'	aYseBaBRutM	'ስብር'	'ስብር'
'ያልሰበረ'	YaLseBeRe	'ስብር'	'ስብር'
'ያለሰበሩ'	YaLeseBeRu	'አለሰበር'	'ስብር'
'ያለሰበረች'	YaLeseBeRec	'አለሰበረች'	'ስብር'
'ያልሰባበረ'	YaLseBaBeRe	'ስብር'	'ስብር'
'ሳትሰባብረ'	satseBaBRe	'ስብር'	'ስብር'
'ሳንሰብረ'	sanseBRe	'ስብር'	'ስብር'
'ሳንሰባብር'	sanseBaBR	'ስብር'	'ስብር'
'ሳልሰባብር'	saLseBaBR	'ስብር'	'ስብር'

'ሳልሰብር'	saLseBR	'ስብር'	'ስብር'
'ሳይሰበር'	saYseBeR	'ስብር'	'ስብር'
'እየሰባበሩ'	IYeseBaBeRu	'ስብር'	'ስብር'
'ቤቴ'	Beetee	'ቤት'	'ብት'
'ቤትህ'	Beeth	'ቤት'	'ብት'
'ቤታችሁ'	Beetachu	'ቤት'	'ብት'
'ቤታችን'	Beetacn	'ቤት'	'ብት'
'ቤትሽ'	BeetX	'ቤት'	'ብት'
'ቤትሽን'	BeetXn	'ቤት'	'ብት'
'ቤቱ'	Beetu	'ቤት'	'ብት'
'ቤቱን'	Beetun	'ቤት'	'ብት'
'ቤታቸው'	BeetaceW	'ቤት'	'ብት'
'ቤታቸውን'	BeetaceWn	'ቤት'	'ብት'
'ቤታችሁን'	Beetachun	'ቤት'	'ብት'
'ቤታችሁ'	Beetachu	'ቤት'	'ብት'
'ቤትዎ'	BeetWo	'ቤት'	'ብት'
'ቤትዎን'	BeetWon	'ቤት'	'ብት'
'ቤቶች'	Beetoc	'ቤት'	'ብት'
'ቤቱ'	Beetu	'ቤት'	'ብት'
'ቤቱን'	Beetun	'ቤት'	'ብት'
'የቤቱ'	YeBeetu	'ቤት'	'ብት'
'ለቤቱ'	LeBeetu	'ቤት'	'ብት'
'ቤት'	Beet	'ቤት'	'ብት'
'ቤትን'	Beetn	'ቤት'	'ብት'
'የቤት'	YeBeet	'ቤት'	'ብት'
'ለቤት'	LeBeet	'ቤት'	'ብት'
'ቤቶቹ'	Beetocu	'ቤት'	'ብት'
'ቤቶቹን'	Beetocun	'ቤት'	'ብት'
'የቤቶቹ'	YeBeetocu	'ቤት'	'ብት'
'ለቤቶቹ'	LeBeetocu	'ቤት'	'ብት'

'ቤቶች'	Beetoc	'ቤት'	'ብት'
'ቤቶችን'	Beetocn	'ቤት'	'ብት'
'የቤቶች'	YeBeetoc	'ቤት'	'ብት'
'ለቤቶች'	LeBeetoc	'ቤት'	'ብት'
'ቆንጆው'	qonJoW	'ቆንጆ'	'ቅንጅ'
'ቆንጆውን'	qonJoWn	'ቆንጆ'	'ቅንጅ'
'ቆንጆዎቹን'	qonJoWocun	'ቆንጆ'	'ቅንጅ'
'ቆንጆዎች'	qonJoWoc	'ቆንጆ'	'ቅንጅ'
'ለቆንጆዋ'	LeqonJoWa	'ቆንጆ'	'ቅንጅ'

References

1. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Comput. Linguist. **37**(2), 267–307 (2011)
2. Medagoda, N., Shanmuganathan, S., Whalley, J.: Sentiment lexicon construction using sentiwordnet 3.0. In: Proceedings of the 11th International Conference on Natural Computation (ICNC), pp. 802–807, IEEE (2015)
3. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Language Resources and Evaluation (LREC), vol. 10, pp. 2200–2204 (2010)
4. Lexical Data Repository of the Ge'ez Frontier Foundation. https://github.com/geezorg/data. Accessed 15 Feb 2017
5. Project teams from Addis Ababa University, Masarykova univerzita, Norges teknisk-naturvitenskapelige universitet, The University of Oslo, Hawassa University, 7F14047 HaBiT - Harvesting big text data for under-resourced languages. http://habit-project.eu/wiki/HabitSystemFinal. Accessed 05 May 2018
6. Gebremeskel, S.: Sentiment mining model for opinionated Amharic texts, Unpublished Masters thesis, Department of Computer Science, Addis Ababa University, Addis Ababa (2010)
7. Tilahun, T.: Linguistic localization of opinion mining from Amharic blogs. Int. J. Inf. Technol. Comput. Sci. Perspect. **3**(1), 890 (2014)
8. Denecke, K.: Using SentiWordNet for multilingual sentiment analysis. In: Data Engineering Workshop. pp. 507–512, IEEE (2008)
9. Piao, S., Rayson, P.: Lexical coverage evaluation of large-scale multilingual semantic lexicons for twelve languages. In: European Language Resources Association (ELRA), pp. 2614–2619 (2016)
10. Christopher Potts: Sentiment Symposium Tutorial Lexicons, Stanford Linguistics. http://sentiment.christopherpotts.net/lexicons.html. Accessed 10 Jan (2019)

Information Extraction Model for Afan Oromo News Text

Sisay Abera[1(✉)] and Tesfa Tegegne[2(✉)]

[1] College of Engineering and Technology, Adigrat University, POBOX 50, Adigrat, Ethiopia
aberasisay79@yahoo.com

[2] Faculty of Computing, Bahir Dar Institute of Technology, Bahir Dar University, POBOX 26, Bahir Dar, Ethiopia
tesfat@gmail.com

Abstract. Information Extraction (IE) concerned with the automatic extraction of facts from text and stores them in a database for easy use and management of the data. As the first research work on IE from Afan Oromo text, we designed a model that deals with Infrastructure news domains in the Oromo language. The proposed model has document preprocessing, learning and extraction and post processing as its main components. In this work recall, precision and F-measure are used as evaluation metrics for Afan Oromo Text Information Extraction (AOTIE). Being trained and tested for the dataset of size 3169 tokens, AOTIE performed 79.5% precision, 80.5% recall and 80% F-measure. These results are used as a baseline to experiment on AOTIE. We set up two main experimentation scenarios to experiment on AOTIE. The first scenario is conducted by developing a gazetteer. The second scenario is aimed at observing the influence of Afan Oromo grammatical structure. Both scenarios showed that, the performance of AOTIE is mostly dependent on grammatical structure of Afan Oromo.

Keywords: Afan Oromo Text Information Extraction · Machine learning approaches to information extraction · Information extraction

1 Introduction

Information Extraction (IE) is an automatic extraction of facts from text [1]. It is based on analyzing natural language in order to extract information. The process takes texts (and sometimes speech) as an input and produces structured data as an output. This data may be used directly by users, or may be stored in a database for later analysis, or may be used for indexing purposes of information retrieval (IR) applications such as web search engines [1–3]. There are more than 80 languages in Ethiopia. Afaan Oromo is one of the language with large number of speakers under Cushitic family [4]. As the Ethiopia‘s statistical report of 2007 [5] shows there are more than 27 million speakers of Afaan Oromo in Ethiopia and this fact shows that, the language has the largest speaker followed by Amharic language.

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 327–340, 2019.
https://doi.org/10.1007/978-3-030-26630-1_28

There are a number of Media that uses Afaan Oromo as primary language; for instance, Oromia Television and Radio (Web news), Kallacha Oromiyaa, Bariisaa, Yeroo, Voice of America (VOA) (web news), and different academic and recreational medias to mention a few [6]. Those Media present different information (textual) in structured, unstructured and semi-structured forms. Cowie and Wilks [3] noted, manually extracting information from such an often unstructured or semi-structured text is a very tiresome and time consuming job. Thus, getting the right information for decision making from existing abundant of unstructured text is a big challenge. In addition, the unavailability of tools for extracting and exploiting the valuable information, which is effective enough to satisfy the users for Afan Oromo language has also been a major problem. The main objective of this experimental study is to design a model of information extraction system from Afan Oromo news texts (Fig. 1).

Information extraction has three different components regardless of the language and domain on which it is developed for, which are linguistic preprocessing, the learning and application stage and semantic post processing to do the extraction of data from a given text [2]. The linguistic preprocessing uses different tools to make the natural language texts ready for extraction. The learning and the application component learns a model and extracts the required information from the preprocessed text and the last component, which is the semantic post processing assigns the extracted information to their predefined attribute category and manages the normalization and duplication problem with the extracted data [7]. The Afan Oromo Text IE model designed in this work has also linguistic preprocessing, the learning and application stage and semantic post processing components. These three main components of AOTIE also contain different subcomponents which are language specific and general subcomponents that are required in IE. The preprocessing component of AOTIE is responsible for tokenization and parsing of news texts whereas the learning and extraction component extracts candidate texts from the news text and learns a classification model that will be used to predict the category of the candidate text. The post processing component is responsible for the formatting of the extracted data.

2 Methodology

In this paper, we follow an experimental science research methodology. We have conducted three experiments to evaluate the performance of AOTIE. The first experiment is conducted using the dataset that is tagged according to BIO tagging format whereas the second and third experiments are conducted using a gazetteer and a data set tagged based on the Afan Oromo grammatical structure. The information extraction approaches can be divided into two primary categories:

- Knowledge engineering approaches and
- Machine learning approach.

The Knowledge Engineering Approach is characterized by the development of the grammars used by a component of the IE system by a "knowledge engineer," i.e. a person that is familiar with the IE system, and the formalism for expressing rules for that system, who then, either on his own, or in consultation with an expert in the domain of application, writes rules for the IE system component that mark or extract the sought-after information [8]. In this approach, the skill of the knowledge engineers plays a large factor of the level of performance achieved by the overall system. In addition to requiring skill and detailed knowledge of a particular IE system, the knowledge engineering approach usually is laborious (is very intensive) [8]. Therefore, developing an IE system using knowledge engineering approach will be very difficult as there are no experts that know IE and the Afan Oromo language. Additionally, the NLP resources available in Afan Oromo language are almost none.

The Machine Learning approach is quite different. Following this approach, it is not necessary to have someone on hand with detailed knowledge of how the IE system works, or how to write rules for it. It is necessary only to have someone that knows enough about the domain and the task is to take a corpus of texts, and annotate the texts appropriately for information being extracted [8]. The machine learning approach to IE is then focused on the automatic acquisition of the extraction patterns that can be used to extract the information relevant to the particular task from each single document. Rather than focusing on producing rules, the automatic training approach focuses on the training data. Corpus probabilities or rules are then derived automatically from the training data, and used to process novel or unseen data. As long as someone familiar with the domain is available to annotate texts, systems can be customized to a specific domain without intervention from any developers. For this reason, machine learning approaches are used to develop a model for Afan Oromo Text Information Extraction (AOTIE).

3 Design and Implementation of AOTIE System

3.1 Data Source

The data we used to train and test AOTIE system is obtained from Oromia Radio and Television Organization (ORTO). The choice of the organization is the availability of the electronic news items for the Afaan Oromo. The news items from September 2007 to June 2007 E.C. (Ethiopian Calendar) are used as data source for the experiment of this research. One of the challenge is the collected data is not classified. Since, information extraction is domain specific, the news items must be classified into different categories. Therefore, after analyzing the collected news documents, we classify the news items into 11 main categories as presented in the following table (Table 1).

Table 1. News main categories

No.	Category name
1	Infrastructure
2	Sport
3	Economy
4	Social Affairs
5	National Politics
6	Law and Justice
7	Associations
8	Defense and Security
9	Others (Different) Affairs
10	Foreign
11	Governmental organizations

Among the different categories from ORTO news the Infrastructure category is used as a data source for the training and testing for our AOTIE system. The reason for selecting Infrastructure news category, among the other categories is the availability of factual information which can be extracted. During selection of news category as a data source, we consider all the eleven categories and analyze which one of these categories contain facts that can be extracted and stored in the database that can be used by different application and individuals after extraction. Most of the other news categories are more subjective which has less factual data that need to be extracted. The other reason is the importance of the data after extraction. An Infrastructure news contain many vital information that investors, financial organizations or any individuals can use [9] and extracting this information and using them for later stage or by other application will be vital. Due to these reasons and the nature of IE which is very domain specific, the infrastructure category news texts are used as training and testing data set in this research work.

3.2 Components of AOTIE

IE system has three basic components which are: preprocessing, learning and extraction and post processing regardless of the approach, language and domain on which the IE system is developed for [2]. In addition to these three components other subcomponents are also included in each of the main components. The Afan Oromo Text IE model designed in this work has also these three components. These main components also contain different subcomponents, which are language specific and general subcomponents that are required in IE. The main components interact with each other to make information extraction from Afan Oromo news text possible.

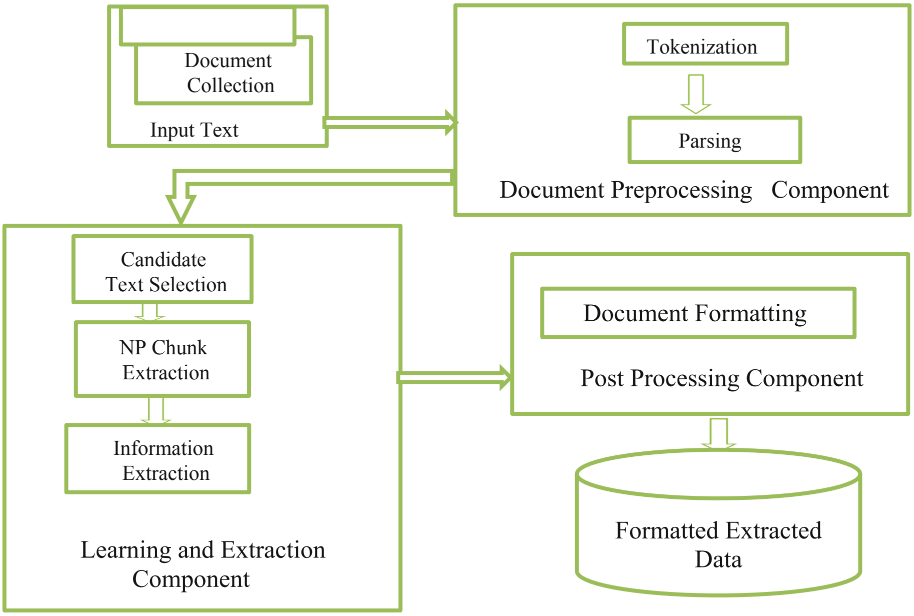


Fig. 1. Afan Oromo Text Information Extraction Model

3.3 Document Preprocessing

The document preprocessing component handles the different language specific issues that are imposed by the nature of the language to make the data ready for remaining phases. The data collected from ORTO requires different preprocessing before it is used by AOTIE system. The document preprocessing component consists of tokenization and parsing.

Tokenization. Tokenization is the task of chopping up the given text into pieces, called tokens [10], which are then used by IE system. Since IE uses tokens and there POS as a candidate text for training and testing classifiers; tokenization is crucial component. Tokenization can occur at a number of different levels: a text could be broken up into paragraphs, sentences or words. For any given level of tokenization, there are many different algorithms for breaking up text depending on the intention of the system and the language under consideration. For example, at the word level, it is not clear how to treat such strings as "can't," "New York," and "so-called". Having these problems, most languages like English, tokenization at the word level is done based on white space.

In Afan Oromo, tokenization is a trivial problem as its writing system looks like English. In fact, there are circumstances in which two words are treated as a single word when they are joined to each other by a hyphen (-) and they are commonly called "*jecha tishoo*". For example "bar-kumee" and "bu'a-qabeessa" are treated as a single word. The tokenizer developed in this research work will treat double words that are connected to each other by a hyphen (-) as a single token.

Parser. During the development of the corpus, we first went through all the obtained news articles and selected those sentences that are Infrastructure news. The obtained sentences are then tokenized and each token was tagged with its BIO tag in accordance with the CoNLL"s 2002 standard. Since the corpus is developed manually it may be vulnerable to errors. One of the errors that may occur is the illegal tagging of the words. BIO scheme's legal tag sequence is described in Table 2 where the first column lists the representation of the tags and the second column lists the tags that may follow them, with the variables ranging over all feature types.

Table 2. BIO legal tag sequence

Tag	Legal following tags
O	O, B-X
B-X	O, I-X, B-Y
I-X	O, I-X, B-Y

In addition to illegal tagging, other errors like: the distance between a token and its tag might be larger than a single white space, there might not be a space between a token and its tag or a token might be left untagged. The corpus has to be parsed and checked against all of the above-mentioned issues. Thus, parsing is a crucial algorithm for checking occurrence of the aforementioned errors and report if errors are found. The parsing algorithm implemented in AOTIE is presented in Fig. 2.

3.4 Learning and Extraction

The main task of learning and extraction component is to extract candidate texts and train and use classifier model for predicting the category of the extracted candidate texts. To train and use the classifier model for learning and extraction component, we use the training and testing phase. The training phase uses the manually annotated news data to train a classifier model which will be used for extraction while testing phase evaluates the performance of the trained model on the unseen news texts.

Training Data Preparation. Training data set preparation to develop an IE system is tedious and time-consuming task. To carry out the experiment of AOTIE system, the infrastructure news category is used. The predefined attributes that we set to extract from infrastructure news are the Infrastructure name, place where the infrastructure is built, the amount of money spent to build the infrastructure, the source of money for the infrastructure development, the number of users that can be benefited from the infrastructure, the person who give the information to the news agency. These six attributes are selected to analyze different infrastructure news and the common facts which exist on most of the infrastructure news and which we thought are relevant facts that should be extracted.

```
    For all lines of text in the file
      Read a line
        If not tagged
          If last index is ".O"
            Assign "." To token
            Assign "O" to tag
           Else
               Report illegal line error
        Else
         Assign index of 0 to token
         Assign index of 1 to tag
      End If
      If the space b/n a word & its tag is more than a single space
        Report Error
      If  the previous tag is O
        Check if the current tag is one of O & B-X
       Else
          Report error
      If the previous tag is B-X
         Check if the current tag is one of O, I-X, B-Y
       Else
         Report error
      If the previous tag is I-X
        Check if the current tag is one of O, I-X, B-Y
       Else
         Report error
End For
```

Fig. 2. AOTIE parsing algorithm

In training data preparation, all the text segments, which are numbers and names are tagged accordingly. If the name is the fact to be extracted, it is tagged according to the tag set for the predefined attribute otherwise it is tagged by other tags.

Paarkiin B-INF Industirii I-INF Adaamaa B-LOC bulchiinsa I-LOC magaalaa I-LOC Adaamaatti I-LOC bajata B-FSOUR mootummaa I-FSOUR qarshii B-CAP miliyoona I-CAP 500 I-CAP oliin O ijaaramaa O jiru O yeroo O xumurametti O lammiilee B-NUS 2000 I-NUS ol O ta'aniif O carraa O hojii O yeroo O fi O dhaabbataa O kan O uumu O ta'uu O obbo O Tasfaayee B-REP Abbabaa I-REP beeksisan O.O

The main purpose of tagging all text which is not facts considered for extraction to train classifier is to identify texts that are part of the predefined attributes and those that is not. Attribute tagging is done based on the BIO tagging format as:

- B-INF for infrastructure name
- B-LOC for place where infrastructure is built
- B-CAP for the amount of money spent for the development of the infrastructure
- B-REP for the name of the person that give the information
- B-NUS for the number of users after the infrastructure development
- B-FSOUR for the financial source of the infrastructure development
- O for any attribute found in the document which is not considered as facts.

NP Chunk Extraction. Classifier model generation requires a corpus of chunks. A special type of chunk used in this work is NP chunk which creates noun phrases known as chunks from a sequence of tokens. It is done by NP Chunker. NP Chunk Extraction in AOTIE is implemented in such a way that first the token/tag sequence will be generated based on BIO tagging format. Then tokens are converted into character sequence with each character in the sequence given a position number. Changing the tokens into character sequence helps to compute the range covered by a phrase. Once the input tokens are represented in character sequence, those tokens with BI (Beginning In) tag sequence will be combined to create a chunk. This means from the sequence of tokens, those tokens following each other with tag sequence of the form B-X I-X I-X … will be combined together to form a chunk of phrases.

Since the extractions are done at phrase level, the task requires deep linguistic analysis. Therefore, developed algorithms are different from an algorithm developed for the same domain but different language. For instance, in Afan Oromo noun phrase formation, nouns come before adjectives which is inverse for English language. For example, in the noun phrase: '*konkolaataa diimaa*' (red car), *konkolaataa* (car) is noun which is succeeded by adjective *diimaa* (red). With regard to our work, NP chunks can be defined as phrases belonging to one of six predefined attribute categories. The obtained chunks will be assigned a tag that was taken from the extension of the constituent BI sequences. The Chunker stores the entity type of the chunk. If a token sequence is not of the form BI sequence, i.e. it has the tag sequence of the form B-X B-Y, then a single token is treated as an NP chunk. Since our concern is extracting NP chunks based on the facts to be extracted, those tokens with tag O are not extracted. The algorithm used to extract NP chunks in AOTIE is presented in Fig. 3.

```
For all token/tag sequence
      Change the token sequence into character sequence
      If a token has O tag
         Ignore;
      For those tokens with tags of the form BI sequence
         Read the tokens until the next tag is B
         Combine the tokens and create an NP chunk
         Store the start and end+1 of the chunk in the sequence
         Assign the chunk a tag of the extension of B
         Create an NP chunk set of the form (start,end+1):tag
      End For
      For those tokens with tags of the form BB sequence
         Return token as an NP chunk
         Store the start and end + 1 of the chunk in the sequence
         Assign the chunk a tag of the extension of B
         Create an NP chunk set of the form (start,end+1):tag
      End For
End For
```

Fig. 3. Algorithm for NP Chunk Extraction

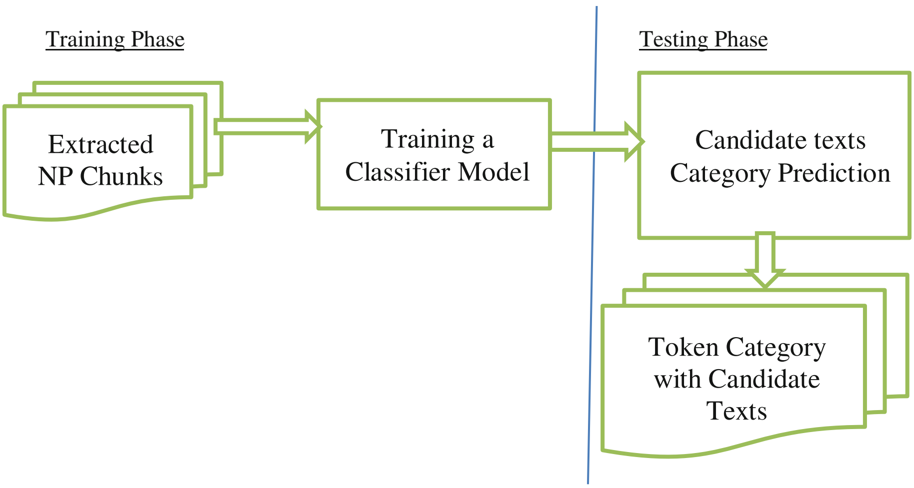


Fig. 4. Information Extraction sub component

Information Extraction. Information Extraction using machine learning approach requires training and test data set to examine the accuracy of the extractor. Since AOTIE is developed using machine learning approach, it has also training and testing phase as presented in the following figure (Fig.4).

In training phase, the classifier model is generated based on the extracted NP chunks. The generated classifier model is used for predicting the category of the candidate texts during the testing phase. The outputs of the information extraction subcomponent are candidate text with their respective predicted category, which is used as an input to the post processing component.

Training A Classifier Model. After the data is preprocessed, it is then passed to the classifier learners in LingPipe. The **Dynamic Language Model Classifier (DynamicLMClassifier)** algorithm is selected among different available machine learning algorithms that LingPipe supports. A DynamicLMClassifier is selected because it is a language model classifier that accepts the training events of categorized character sequences. Training is based on multivariate estimator for the category distribution and dynamic language models for per-category character sequence estimators. It is also used in tag-a-little, learn-a-little supervised learning without retraining epochs. We select this algorithm to see the performance of the classifier and evaluate its efficiency in predicting category for the candidate tokens. Once the classifier model is generated using the training data; the next step is to use the trained classifier model to work on the unseen news text. But the problem is the trained classifier model doesn't directly work for the plain news texts, it rather works on the extracted NP Chunks from the news texts. To use the classifier model for extraction, first the candidate texts should be identified and the NP Chunks of the candidate texts must be extracted using the NP Chunk extractor. In order to achieve this, there is a candidate text selection component which tags all the candidate tokens accordingly.

Candidate Text Selection. The candidate texts for extraction for this research are names and numbers. Names and numbers are considered as candidate texts because, names and numbers are the most important facts for extraction. And also since extracting relationship between the entities requires different NLP tools which are not publicly available for Afan Oromo, names and numbers are considered as the candidate texts in this research work. Name of person is candidate text for Reporter attribute, place name is for a place where the infrastructure is built on, numbers in news text are considered as a candidate text for number of users and for amount of money spent for infrastructure development, organization names are used as candidate texts of attribute financial source for infrastructure development and names used for infrastructure are the candidate texts. The task of candidate text selection is to tag all these candidate texts according to their predefined attributes.

Extraction of Data. After extracting the NP chunks, we predict the category of candidate texts using the trained classifier; texts which match the correct category of the predefined attributes are stored in the database, whereas those which do not match the category are discarded.

4 Post Processing

Different attributes that are extracted from the news texts are finally formatted and stored in the database according to the predefined format of the database slot. The main function of the post processing component is to arrange the format of the extracted data.

5 Results and Discussion

The AOTIE System is trained and tested on the dataset developed by the researchers. The dataset contains 155 news texts which contain a total of 3169 tokens. When trained and tested on the original dataset (dataset which contains 3169 tokens), AOTIE score ***Precision*** value 79.5%, ***Recall*** value 80.5% and an F-***measure*** value 80% (Table 3).

Table 3. Detailed accuracy by attribute category

Attribute category	Precision	Recall	F-measure
Infrastructure name	65.5%	73.5%	69.2%
Location	79.5%	73.8%	76.5%
Money spent	89.2%	100%	94.3%
Reporter	80%	66.7%	72.7%
Number of users	100%	57.1%	72.7%
Financial source	95.2%	90.9%	93%
Combined cross val.	79.5%	80.5%	80%

These values are used as references to experimentation scenarios to observe effects those scenarios have on the performance of AOTIE.

5.1 Evaluation Scenarios of AOTIE

To evaluate the performance of AOTIE system, we have used two evaluation scenarios: using gazetteer and grammatical structure of Afan Oromo.

Scenario 1: Using Gazetteer In order to examine performance of AOTIE system, we prepared a gazetteer or a set of dictionaries of the six attribute categories. The reason for developing a gazetteer is to reduce the burden of tagging all attributes accordingly. Therefore, the gazetteer which consists of names for different places in Ethiopia, different names that can be used for identification of persons, infrastructure names, and different governmental and nongovernmental organization lists, which are found in Ethiopia are used. We have then trained and tested AOTIE model on the dataset using the gazetteer. The experimental results of scenario 1 is presented in the following table.

Table 4. Detailed accuracy by attribute category for Scenario 1

Attribute category	TP rate	FP rate	Precision	Recall	F-measure
Infrastructure name	70.2%	29.8	70.2%	85.1%	76.9%
Location	62.5%	37.5%	62.5%	78.9%	69.8%
Money spent	91.2%	8.8%	91.2%	100%	95.4%
Reporter	44.4%	55.6%	44.4%	66.7%	53.3%
Number of users	80%	20%	80%	57.1%	66.7%
Financial source	77.3	22.7%	77.3%	89.5%	82.9%
Combined cross val.	72%	28%	72%	85.1%	78%

Table depicts that both precision and F-measure are decreased by 7.5% and 2% respectively while recall increases by 4.6%. From this experiment, we understood that using dictionary reduces the performance of information extraction from Afan Oromo news text.

Scenario 2: Influence of Afan Oromo grammatical structure In order to examine impact of grammatical structure of Afan Oromo writing system on AOTIE system, we prepared a dataset which is tagged based on Afan Oromo grammatical structure. While preparing the dataset, we have also increased the size of our dataset of 3169 tokens to 5187 tokens.

The reason for developing a dataset based on the grammatical structure of the language is that in Afan Oromo writing system a single word (token) can be used as an attribute of different attribute category. For example in the text:

Waarshaan Simintoo Daangootee [1] dureessi Afrikaa Haji Aliko Daangootee [2] godina shawaa kaabaatti ijaaran dargaggoota hoji-dhabeeyyif carraa hojii uumuun ibsame.

Daangootee [1] is in an Infrastructure name category, but depending on the contextual meaning *Daangootee [2]* is in the category of a person who gives a financial support for infrastructural development (Financial Source). The model is then trained and tested on the new dataset. As Table 4 shows the evaluation result increases as the size of the dataset increases depending on grammatical structure of the language (Table 5).

Table 5. Detailed accuracy by attribute category for Scenario 2

Attribute category	TP rate	FP rate	Precision	Recall	F-measure
Infrastructure name	75%	25%	75%	76.6%	75.8%
Location	88.9%	11.1%	88.9%	84.2%	86.5%
Money spent	91.2%	8.8%	91.2%	100%	95.4%
Reporter	80%	20%	80%	66.7%	72.7%
Number of users	100%	0	100%	57.1%	72.7%
Financial source	84.2%	15.8%	84.2%	84.2%	84.2%
Combined cross val.	84.2%	15.8	84.2%	83.1%	83.7%

As it can be seen in the Table – using grammatical structure have improved results of AOTIE. Precision, recall and F-measure have increased by 4.7%, 2.6% and 3.7% respectively.

6 Conclusion

As the growth of unstructured and semi structured documents in an electronic media is increasing from time to time, a tool that can extract relevant data to facilitate decision making is becoming crucial. Information Extraction (IE) is concerned with the automatic extraction of facts from text and stores them in a database for easy use and management of the data. As the first research work on IE from Afan Oromo text, we designed a model that deal with Infrastructure news domains in the Oromo language. The proposed model has document preprocessing, learning and extraction and post processing as its main components. The preprocessing component is responsible for tokenization and parsing of news texts. The learning and extraction component extracts candidate texts from the news text and learns a classification model that are used to predict the category of the candidate text. The post processing component is responsible for the formatting of the extracted data.

In this work Recall, Precision and F-measure are used as evaluation metrics for Afan Oromo Text Information Extraction (AOTIE). Being trained and tested on the dataset of size 3169 tokens, AOTIE performed 79.5% Precision, 80.5% Recall and 80% F-measure. These results are used as a baseline to experiment on AOTIE. We set up two main experimentation scenarios to experiment on AOTIE. The first scenario is conducted by developing a gazetteer. The second scenario is aimed at observing the influence of Afan Oromo grammatical structure. Both scenarios showed that, the performance of AOTIE is mostly dependent on grammatical structure of Afan Oromo.

7 Contribution of the Work

- A model is developed for Afan Oromo text information extraction system using a machine learning approach.
- An algorithm is developed for language specific issues which can handle parsing.
- An algorithm for automatic NP Chunk extractor is developed.

8 Future Work

Information Extraction is a new study for Afan Oromo and also for Ethiopian languages. As IE requires different NLP processing tools for extraction, the task is very complex for such under resourced languages. IE system proposed in this research work is not complete it requires different improvements. The following are our recommendations for future works.

- In this research work names and numbers are the only attributes to be extracted. It does not consider about the relation between attributes. Therefore the inclusion of relation extraction in the further study will be vital for the development of full-fledged information extraction.
- For recognizing names, we try to tag all names in the news which is very complex and time consuming task. Using automatic named entity recognition in later stages might minimize the burden of selecting the named entities.
- Incorporating Afan Oromo spell checkers to minimize the spelling problems which mostly happen in the news text might also have an impact as we manually modify the spelling errors as they have impact for named entity recognition.
- In this work, we used a single algorithm. Interested researchers can use more algorithms and compare the results to obtain best algorithm for AOTIE.
- As we proposed a machine learning approach for the system, interested researchers can develop a rule based approach and compare the performance of the system.

References

1. Corro, L.D.: Methods for open information extraction and sense disambiguation on natural language text, Doktor der Ingenieurwissenschaften (Dr.-Ing.), Naturwissenschaftlich-Technischen, Saarlandes (2015)
2. Fabio Ciravegna, C.G., Kushmerick, N., Lavelli, A., Muslea, I.: Adaptive text extraction and mining (ATEM 2006). In: 11th Conference of the European Chapter of the Association for Computational Linguistics Proceedings of the Workshop, Trento, Italy, 4 April 2006
3. Wilks, J.C.A.Y.: Information extraction, Lecture note on Information extraction
4. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval, vol. 1. Cambridge University Press, Cambridge (2008)
5. Ababa, A.: Population Census Commission, Summary and statistical report of the 2007 population and housing census, population size by age and sex (2008)
6. Etnologue: Show Language. http://www.ethnologue.com/web.asp. Accessed 15 Jan 2016
7. Siefkes, C., Siniakov, P.: An overview and classification of adaptive approaches to information extraction. J. Data Semantics IV, 172–212 (2005)
8. Masche, P.: Multilingual information extraction, Master's thesis, University of Helsinki, Faculty of Science, Department of Computer Science (2004)
9. Tsedalu, G.: Information extraction model from Amharic news texts, Addis Ababa University (2010)
10. Stanford tokenization. https://nlp.stanford.edu/IR-book/html/htmledition/tokenization-1.html. Accessed 24 Apr 2017

Design and Development of Sentence Parser for Afan Oromo Language

Hailu Beshada Balcha[1,2](✉) and Tesfa Tegegne[1,2](✉)

[1] ICT for Development (ICT4D) Research Center,
Bahir Dar Institute of Technology, Bahir Dar University,
P.O. Box. 26, Bahir Dar, Ethiopia
beshadahailu@yahoo.com, tesfat@gmail.com
[2] Faculty of Computing, Bahir Dar Institute of Technology,
Bahir Dar University, P.O. Box. 26, Bahir Dar, Ethiopia

Abstract. Parser is an efficient and accurate enough to be useful in many natural language processing systems, most notably in machine translation [1]. Previously many sentence parsers are developed for foreign languages such as English, Arabic, etc. as well as for Amharic language from local languages of Ethiopia. However, to the best of the researcher's knowledge concerned, there is no Afan Oromo sentence parser for simple and complex sentences. Thus, we proposed to develop a sentence parser for Afan Oromo language. Parsing Afan Oromo sentence is needed and a necessary mechanism for other natural language processing applications like machine translation, question answering, knowledge extraction and information retrieval, particularly for Afan Oromo language. Rule-based parser using a top-down chart parsing algorithm for Afan Oromo sentences presented in this paper. Context Free Grammar (CFG) is used to represent the grammar. 500 sentences were prepared for sample corpus and CFG rules are extracted manually from sample tagged corpus. We also developed simple algorithm of a lexicon generator to automatically generate the lexical rules. Python programming language and NLTK are used as an implementation tools for this study. From the total of sample dataset 70% is simple sentence type because of we considered four different types of simple sentences (declaratives, interrogatives, imperatives and exclamatory sentences) and the rest 30% is complex sentence type. The parser was trained on 400 sentences of training dataset with the accuracy of 98.25% and tested on 100 sentences of testing dataset with the accuracy of 91%. The experimental results on a parser is an encouraging result since it is the first work for simple and complex sentences of Afan Oromo language.

Keywords: NLP · Parser · Context Free Grammar · Top-down chart parser · Lexicon generator · Lexicon

1 Introduction

One of the fundamental aspects of human behavior is a natural language. Natural language is a language that is spoken by the people. Theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at

F. Mekuria et al. (Eds.): ICT4DA 2019, CCIS 1026, pp. 341–358, 2019.
https://doi.org/10.1007/978-3-030-26630-1_29

one or more levels of linguistic analysis for the purpose of achieving human like language processing for a range of tasks or applications is natural language processing (NLP) [2]. NLP runs different applications, namely tokenization, lexical analysis, syntactic analysis, semantic analysis, and pragmatic analysis. Among these applications our focus is on parsing (Syntactic analysis), which provides an order and structure of each sentence in the text. According to [3] parsing is the process of assigning structural descriptions to the sequences of words in natural language processing. Mainly two components are used in parsing process: a parser and a grammar. Parser is a procedural component and a computer program, whereas, grammar is a declarative component. Both the parser and grammar depend on the grammar formalism.

Sentence parsing is one of the steps to design a functional NLP application and which can work in cooperation, and as input to other many NLP applications like grammar checker, machine translation, and etc. It is also called syntax analysis [4], which is the process of identifying how words can be put together to form correct sentence and determining what structural role of each word plays in the sentence and what phrases are subparts of what other phrases or what other words modify, which words of the central point of the whole sentence constructed. Parsing has an important role in semantic processing operation on that of sentence constituents. If there is no syntactic parsing step, then the semantics system must decide on its own constituents [5]. If parsing is done, however, it contains the number of constituents that semantic can consider. The present work is a contribution towards designing and developing natural language processing applications for Ethiopian Languages, particularly, for Afan Oromo language. Hence, this study is to develop a sentence parser for Afan Oromo language using top-down chart parsing algorithm and Context Free Grammar (CFG) formalism to represent Afan Oromo grammar rules.

To this end, the researcher has gone through different literatures to find if there is any sentence parser, which can parse simple and complex sentences in Afan Oromo. Thus, to the best of the researcher's knowledge, there is no Afan Oromo sentence parser for simple and complex sentences. However, there is one attempted by [4] on automatic sentence parser for Afan Oromo language using supervised learning technique for simple declarative Afan Oromo sentences. In his study, the chart algorithm has been used. In addition, the unsupervised learning algorithm was designed to guide the parser in predicting unknown and ambiguous words in a sentence. It also adopts an intelligent (Rule-Based learning module) approach to develop a prototype. The result obtained was 95% on the training dataset and 88.5% on the test dataset. The parser was developed purely based on an Intelligent (hybrid of Rule-based and supervised learning) System approach and tagger were not included, which could have been used as a preprocessor to the parser. It was developed only for simple declarative sentences of Afan Oromo language. Due to this fact, we are motivated to develop a parser for simple and complex Afan Oromo sentences. Moreover, as the nature and structure of sentences parsing (syntactic parsing) in Afan Oromo is different from English, Amharic or other languages, sentence parser developed for such languages could not be functional for Afan Oromo language. It (Afan Oromo) has different syntactic and morphological nature, and they have also their own grammatical and word formation technique that is different from other languages. As a result, sentence parser developed for other languages could not be used for Afan Oromo language, which results in the need for the independent sentence parser. So, sentence parser for Afan Oromo language is needed independently.

Even though there are several works on sentence parsing using various approaches, there is only one study for Afan Oromo language regarding sentence parsing. The first parser for an automatic Afan Oromo sentence parser was developed by [4] which was aimed to parse declarative simple sentences. Some works have been done on Amharic sentence parsing. However, it is very few works when compared to the number of works dealing with other foreign natural languages such as Arabic, English, etc. A simple automatic parser for Amharic sentences was developed by [6], to address the problem of developing systems that can automatically process Amharic text. The study was tried to combine probabilistic formalism and rule-based reasoning for developing automatic sentence parser. The total size of sample corpus was 100 Amharic sentences only from simple declarative sentence. The results achieved based on the first set of sample sentences was very high, 100% on the training set and approximately 96% on the test set. As a researcher state in her work, this high accuracy is obtained partly due to the small number of words considered for the experiment. Another reason is that all the sentences have identical constructions, and the highest probability parses were almost always the correct ones. The second attempted was by [7]. The work was the integration of the ideas and outputs of previously attempted by [6], to develop an automatic sentence parser, particularly for complex Amharic sentences. The third work in an Amharic sentence parser was done by [8], which is a hybrid approach to Amharic base phrase chunking and parsing. Its main objective was to extract different types of Amharic phrases by grouping syntactically correlated words, which are found at a different level of the parser using Hidden Markov Model (HMM) model and to transform the chunker to a parser. Another important work and similar approach and parsing strategy that we have been proposed in our study was done by [5]. The researcher was developed top-down chart parser and designed to parse all types of Amharic sentences using a top-down chart parsing algorithm using Context Free Grammar to represent the Amharic grammars. There are also many works of sentence parser systems that have been done globally with different approaches. For an English language sentence parser, the researchers in [9] and [28] are developed a parser. For an Arabic language, some of researchers in [10], [11] and [12] are developed a sentence parser. The researcher in [13] was develop sentence parser for Indian language. The researchers in [14] and [15] are also developed sentence parser for Myanmar language and Chinese language sentences respectively.

2 Research Methodology

Parsing is the step in which a flat input sentence is converted into a hierarchical structure that corresponds to the units of meaning in the sentence. There are two basic approaches in order to parse the sentences efficiently, such as stochastic and rule-based approaches. Stochastic approach is called corpus-based approach, which is based on the use of text corpora. The rules are assigned with probabilities in stochastic sentence parsing approach. On the other hand, rule-based approach is based on a core of solid linguistic knowledge. The advantage of the rule-based approach over the corpus-based approach is for less-resources languages [16]. There are three ways in which this approach can be applied, such as top-down, bottom up, and chart-based approach.

Top-down approach starts with the largest point. It breaks down from the largest into the smaller segments. The advantage of this approach is that the only rules are applied, which can be useful in proving that the sentence is grammatical, and its disadvantage is that the rules are tried blindly without any regard to the lexical material present in the sentence [17]. Unlike top down parsing, the bottom up parser only checks the input sentence once, and only builds each constituent exactly once [18]. This is because a bottom-up parser works from left to right. It has an advantage that the choice of the grammar rules that are applied depends on the words present in the sentence and on analyses for sub-strings of the sentence. Analyses for sub-strings which do not contribute to the overall analysis of the sentence is the disadvantage of bottom-up approach [17]. Chart-based approach combines some of the advantages of top-down and bottom-up approaches. According to [19], in chart parsing approach the process of parsing is an n-word sentence consists of forming a chart with n + 1 vertices and adding edges to the chart one at a time. There is no backtracking, everything that is put in the chart stays there, and chart contains all information needed to create a parse tree. Chart parser is driven by an agenda of completed constituents and the arc extension, which combines active arcs with constituents when they are added to the chart [10]. The technique of extending arcs with constituents can be applied by using both bottom-up and top-down approach. However, the difference is in how new arcs are generated from the grammar. In bottom-up approach, new active arcs are generated whenever a completed constituent is added that could be the first constituent of the right-hand side of the rule, whereas in the top-down approach, new active arcs are generated whenever a new active arc is added to the chart [5, 10]. On the basis of concepts, [5] states that, the number of constituents generated using a top-down chart parser is less than the number of constituents which are generated using bottom-up chart parser. Therefore, [5] suggests the top down chart parser is considerably more efficient for any reasonable grammar. Thus, we are motivated to develop Afan Oromo Sentence Parser by using top-down chart parsing approach.

3 Structures of Afan Oromo Language

Ethiopia is one of the multilingual countries. It constitutes more than 80 ethnic groups with diversified linguistic backgrounds. The country comprises the Afro-Asiatic super family (Cushitic, Semitic, Omotic and Nilotic) [20]. Afan Oromo belongs to an East Cushitic language family of the Afro-Asiatic language super family. It is the most widely spoken language in Ethiopia. Afan Oromo language is the official language of Oromia National Regional State. It is used in offices, schools, colleges, universities and in media. The writing system of this language is nearly phonetic since it is written the way it is spoken, i.e. one letter corresponds to one sound. The language uses Latin alphabet **"Qubee"** which was formally adopted in 1991 G. C [21], and it has its own consonants and vowels sounds. Afan Oromo has thirty- three consonants, of these seven of them are combined consonant letters: **ch, dh, ny, ph, sh, ts** and **zh**. The combined consonant letters are known as **'qubee dachaa'**. Afan Oromo has five short and five long vowels. The Afan Oromo alphabet is characterized by capital and small letters like English alphabet. In Afan Oromo, as in English language, vowels are sound makers and do stand by themselves.

Afan Oromo is morphologically rich language, each root word can combine with multiple morphemes to generate a huge number of word forms [22]. The grammatical system of Afan Oromo is quite complex and exhibits many features common to other Cushitic languages, this means it is an inflected language that uses post-positions more than prepositions [21]. For the purpose of supporting such inflectionally rich languages, the structure of each word has to be identified. Thus, we present about the grammar of Afan Oromo language starting from its word orders, word classes, phrase types and sentence types in the following sections due to their importance for our study.

3.1 Word Order

Words are combined in different orders to form sentences and phrases. One of the primary ways in which languages differ from one another is in the order of constituents or word order. For instance, Afan Oromo and English have differences in their syntactic structure. In Afan Oromo, the sentence structure is subject-object-verb (SOV). SOV is a sentence structure where the subject comes first, then the object and the verb next to the object. For example, if we take Afan Oromo sentence **"Dagaagaan nyaata nyaate"**, "Dagaagaan" is the subject, "nyaata" is the object and "nyaate" is the verb of the sentence. In case of English, the sentence structure is subject-verb-object (SVO). If the above Afan Oromo sentence is translated into English it will be **"Degaga ate food"** where "**Degaga**" is the subject "**ate**" is the verb and "**food**" is the object, however, Afan Oromo follows the Subject-Object-Verb (SOV) format. But nouns change depending on their role within the sentence, word order can be flexible, though verbs always come after their subjects and objects. Typically, indirect objects follow direct objects.

3.2 Word Categories

In Afan Oromo language based on their contextual and formation in the sentences, word classes are categorized into five major groups. These are noun, adjective, verb, adverb and adposition (pre- and post-position). However, this paper adopts the trend that conjunctions and adposition appear in the same category, which is adposition category.

Noun: Nouns are names that are used to name or identify things, people, animals, places or abstract ideas. In Afan Oromo noun, we can have nouns, adjectives and pronouns.

Adjective: An adjective modifies a noun or a pronoun by describing, identifying, or quantifying words. In Afan Oromo, an adjective usually follows the noun or the pronoun which it modifies. Adjectives are also categorized under different categories, like nouns. Afan Oromo adjectives can be either primitive or derived. It comes after the nouns they describe.

Verb: Verb is the most important part of a sentence that says something about the subject to a sentence, expresses an action, events or states of being. In Afan Oromo, verb occurs within the final positions of a sentence. It is not the case that verbs constitute a distinct, open word class in all languages.

Adverb: Adverbs are words, which are used to modify a verb, an adjective, another adverb, or a clause. Adverbs usually precede the verbs they modify or describe in Afan Oromo sentences. An adverb indicates time, manner, place, cause, or degree and answers questions such as, how? when? where? and how much? In the following example, each of the bold words is an adverb.

Adposition: Adpositions are traditionally defined as words that link to other words, phrases, and clauses and express spatial or temporal relations. Adpositions are almost universal part of speech. It is a cover term for prepositions and postpositions. Afan Oromo has both prepositions and postpositions, though postpositions are more common.

Conjunction: A conjunction is the word that is used to connect words, phrases, clauses or sentences. Conjunctions in Afan Oromo are coordinating or subordinating. In this study, conjunction and adpositions are used as the same category.

Numeral: Numerals are words representing numbers, and they can be cardinal or ordinal numbers [2]. Afan Oromo cardinal numbers refer to the counting numbers, because they show quantity. Ordinal numbers, on the other hand, tell the order of things and their rank. In Afan Oromo, the ordinal numbers are formed from the cardinal numbers by suffixing the suffix **{–affaa}** [2, 4].

3.3 Phrases Categories

A phrase can be defined as a syntactic combination of a word with one or more other words. A phrase is constrained or restricted by two things [4]. These are in terms of the constituents' and the lexical categories like nouns, verbs, etc. Based on the type of lexical categories in Afan Oromo, there are five phrase types in the language [23]. They are briefly presented as following.

Noun Phrase: A noun phrase is made of one noun and one or more other lexical categories, including the noun itself. For example, in the phrase **'mana citaa'** [**thatched house**], there are two nouns, which make the noun phrase: **mana** [house] and **citaa** [thatched] [4].

Verbal Phrase: A verb phrase (VP) is composed of a verb as head and other constituents such as complements, modifiers and specifiers. Afan Oromo verb phrases can be captured by dividing verbs into three categories. These are intransitive verbs, strictly transitive verbs and ditransitive verbs [4].

Adverbial Phrase: Adverbial phrase is made up of one adverb as head word and one or more other lexical categories, including adverbs itself as modifiers and specifiers in Afan Oromo. For example, in Afan Oromo it is possible to have two adverbs in an adverb phrase like in a phrase 'kaleessa galgala' [yesterday night]. As indicated above adverbs and their phrases are used to modify verbs. Hence, they precede verbs in a sentence.

Adjectival Phrase: An adjective phrase is a group of words that describe a noun or pronoun in a sentence. In adjectives nouns, can act as adjectives like **'Mana Magarsaa'** '**Megersa's house'** or verbs as adjectives like '**farda bite'** 'bought horse'.

Adpositional Phrase: Adpositional phrases are the combination of nouns and adpositions. They usually specify a verb phrase. This phrasal category sometimes is called adpositional objects [4, 23].

4 Design of Afan Oromo Sentence Parser (AOSP)

In this section, the main components of Afan Oromo sentence parser and its proposed architecture is presented.

4.1 Components of Afan Oromo Sentence Parser (AOSP)

Rule-based sentence parser approach has the three basic components such as grammar rule, lexicon and the parsing algorithm. The grammar component is responsible for storing grammar rules written in one of the grammatical formalisms. To learn a set of rules automatically based on the given strings, the parser should enable by grammar rules and then the parser parse sentences based on those rules. The lexicon component is used as a dictionary for the parser by storing lexical rules which are separated from grammar rules. According to [5] states in his work, the rule based parser comprises of the structure of the grammar rules, the lexicon and the parsing algorithm differs from system to system and from one language to the other. Hence, additional components to the basic components such as Sentence Tokenizer and Lexicon Generator are required. Sentence Tokenizer is used to break down or split the input sentence into individual words. whereas, Lexicon Generator is used to avoid the manual preparation of the lexical rules and generate the lexical rules automatically from sample corpus in the form of **tag** → **'word'**. We used the corpus which is manually annotated. However, the POS tagger is not included in our system. Grammar rule component of Afan Oromo sentence parser is required to store Afan Oromo Context Free Grammar rules. The grammar rules are identified in a way that they can represent the structure of Afan Oromo sentences in terms of what phrases and word categories, in case we used Context Free Grammar formalism. The lexicon component stores a list of lexical rules, which specify the possible categories of each word. The parser engine accepts input sentence, by considering the grammar rules of Afan Oromo sentence, and retrieves the POS tag of each word in the sentence from the lexicon and finally returns parsed sentence as an output (Fig. 1).

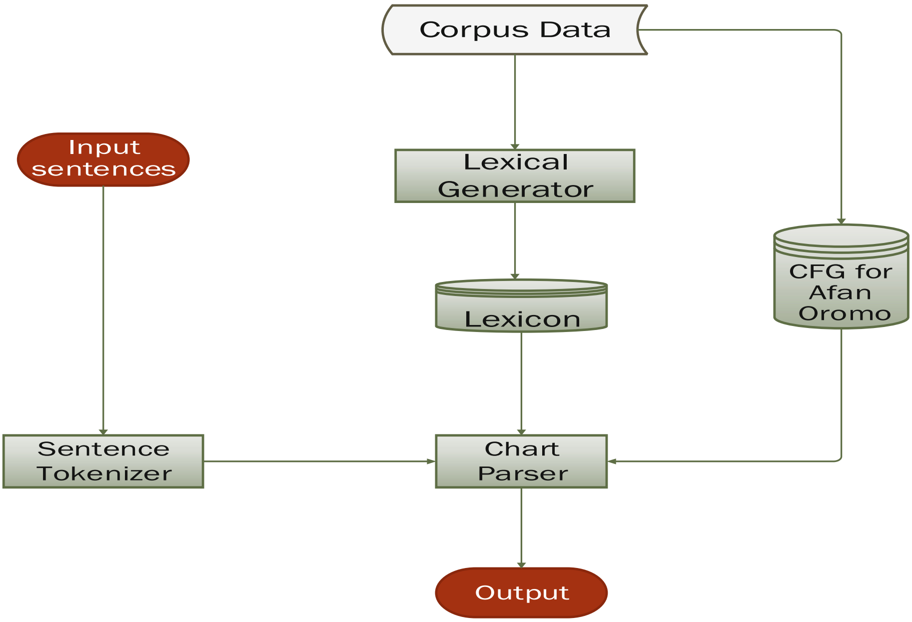


Fig. 1. Architecture of sentence parser for Afan Oromo language

4.2 Context Free Grammar (CFG)

A grammar in human language represents understandable specification of language syntax. The grammar is a collection of words that describes well-informed sentences in a language. An efficient parser can be constructed automatically from a properly designed grammar. The idea of a context-free grammar should be familiar from formal language theory. Furthermore, Context Free Grammar in natural languages represents a formal system which describes a language by specifying how any legal text can be derived from a distinguished symbol called the syntactic symbol [24]. CFG rules for this paper are extracted from sentences collected from Afan Oromo grammar books and previous research papers, which are already tagged manually by researchers. It is used to represent the grammatical structure of valid sentences as much as possible. As we have discussed earlier, there are two main types of sentences in Afan Oromo language, such as simple and complex sentences. Hence, context free grammar rule incorporates both sentences to represent their grammatical structures in this paper.

4.3 Sentence Tokenizer

Tokenization is an early step of processing to divide the input text into units called tokens where each is either a word or something [18]. It also stated in [25] as one of the more basic operations that can be applied to a text to breaking up a stream of characters into words, punctuation marks, numbers and other discrete items. For this reason, we need to have a tokenizer that is responsible to break words of the sentence. The tokenizer spans through the sentence from the beginning to the end and whenever it gets a space gap it considers the text before the space as one word. This process was done by writing python code which is used to split the input sentence into its words.

4.4 Lexicon Generator

A lexicon is the vocabulary of a person, language, or branch of knowledge. It is also thought to include bound morphemes, which cannot stand alone as words, such as most affixes. Because words tend to follow regular morphological patterns, many forms of words are not explicitly included in the lexicon. For example, for the verb 'deem-'(dependent root form), there are different forms of the verb **'deemte, deemuu, deeme, etc.'** [4]. From the sample corpus we have collected, the lexicon was prepared which is a list of words with their POS tag name. Preparing the lexicon manually by typing the word and its word category, especially when there is a large size corpus it is an error prone and time taking. Therefore, it is better to have an automatic lexicon generator that outputs the result correctly and within a short time. We have prepared simple algorithm, then wrote small python code to develop a lexicon generator that output lexical rules from tagged sentences automatically for later used in parsing Afan Oromo sentences. The lexicon generator reads the POS tagged sentence, then generate the result as **(tag name - > 'word')**.

4.5 AOSP Chart Parser

This component is the main part of the proposed system. According to Zhu [26], general search(top-down or bottom-up) methods are not best for syntactic parsing, the reason is the same syntactic constituent may be re-derived many times as a part of larger constituents due to the presence of the local ambiguities of the grammar. So, we have considered the idea of chart parsing. A chart is a form of well-formed substring table. Chart parsing is a common context free parsing algorithm which uses dynamic programming techniques to avoid duplication of effort by ignoring differences in derivation where they have no effect [27]. Chart parsing doesn't throw away any information. This means it keeps a record (a chart) of all the structure you have found so far [10]. Chart parsing has two forms: *passive* and *active*. In case of *passive chart parsing*, the chart is simply a record of all constituents that have been recognized. Whereas, *active chart parsing* is to keeping track of a record of complete constituents that we have found, so we record what we are actually looking for and how much of it we have found so far. Such information is recorded in active edges or active arcs. [i.e. **S → NP. VP**]. In this production rules, the arc label as S → NP. VP means: "we are trying to build an S consisting of an NP followed by a VP. So far, we have found an NP arc, and we are still looking for the VP". So, the insignificant looking "." Or *dot* symbol in the middle of the rule is very important, it marks the boundary between what we have found so far, and what we are still looking for, i.e. a boundary between active and passive arcs. Therefore, constituents before the dot are passive edges, whereas active edges can be combined with these passive edges to create new edges. According to [5], the fundamental rule for combining of the passive edges and an active edge can be performed as S → **A. cB.** Suppose that there is the passive edge going from where the active edge ends and has category *c* on the left side as c → **Z.** The dot in the active edge is now moved one category forward (i.e., S → **Ac. B**).

Chart parsing has two main approaches to apply the parsing algorithm, such as top-down and bottom-up approach. Bottom-up chart parsing checks the input sentence and builds each constituent exactly once. It can also avoid duplication of effort. However, bottom-up chart parsing may build constituents that cannot be used legally, whereas only grammar rules that can be legally applied will be put on the chart in top-down chart parsing [26]. In addition, the algorithm reads (bottom-up) the rules right-to-left, and starts with the information in passive edges. However, Top-down parsing reads the rules left to right and starts with the information in active edges in the case of top-down chart parsing. Besides, top-down searching will be used to use rules to make active edges. The agenda will have at least one active edge to start the parsing process. The active edge starts at the position zero from the sentence S. Hence, the active edge will be taken from the input sentence. The chart will remain empty until an active edge is added to it. Thus, in our work, we proposed, Context Free Grammar and top-down chart parsing for the grammar rules and for parsing algorithm respectively. The reason why CFG grammar formalism is selected in our work, it is easier to maintain, easy to understand and to add new language features. It also imparts structure to language and builds an efficient parser automatically [24]. On the other hand, the reason we have chosen top-down chart parsing is because it does well if there is useful grammar-driven control [5], and it has the advantage of both top-down and bottom-up parsing.

5 Experiment

In this section the details implementation of Afan Oromo sentence parser is presented.

5.1 Development Environment

We have developed a sentence parser which takes an input sentence from the user and parses the sentence according to the CFG and lexicon based on the parsing method and finally deliver the output for the user. The parser has also automatic lexicon generator as its components. The lexicon generator takes manually tagged sentences from local storage and produces lexical rules automatically. We used python programming language for implementation purpose.

5.2 Corpus Preparation

500 simple and complex sentences have been collected from different sources. Most of the sentences are simple sentences which around 70% and the rest 30% is complex sentences. In this study we have used declarative, exclamatory, interrogative and imperative simple sentence types and complex in Afan Oromo sentences. Some of the sentences are taken from previous research work by [4], some others are taken from Seer-luga Afan Oromo (Afan Oromo grammar) book, and the rests are taken from different Afan Oromo written documents. Sentences are tagged manually using tag set

which was developed by [28] based on Afan Oromo language rule and verified by linguistics of Afan Oromo language.

5.3 Grammar Rules Extraction

CFG rules are used to train the parser with a set of grammar rules and enable it to parse sentences based on rules. In order to extract the grammar rule from manually tagged sentences we have reviewed and identified the morphological property of the language and word order. Then, Context Free Grammar rules are extracted manually from the collected corpus after studying the grammar of Afan Oromo language and verified by the linguists.

Our sample corpus contains manually tagged sentences (POS) in order to identify the word category of each word in the sentence. Sentences (looked at the sentences) are used to construct CFG rule and identify part of speech tag of a phrase (Noun Phrase and Verb Phrase). A number of sentences have similar phrase structure, and some others share common sub-phrases so that a single CFG rule can represent many sentences. Once we have the structure of the sentence, we transformed it into the proper format of the CFG, which is a non-terminal followed by an arrow and then terminal or other non-terminals, which can replace the non-terminal before the arrow, i.e., like S - > NP VP, NP - > NN JJ and VP - > NN VB. The CFG begins from the non-terminal S, which represents sentences, and then phrases, which can form S, most of the time noun phrase (NP) and verb phrase (VP). Each phrase which is on the right side of S will be expanded or expressed by other non-terminals (like NN, JJ, VB, AV, etc.) in next rules. As we have discussed in our previous section in detail there are five phrases in Afan Oromo language (Table 1).

Table 1. Tag name of Afan Oromo phrases

Name phrase	Tag name
Noun phrase	NP
Verb phrase	VP
Adverb phrase	ADP
Adjective phrase	JJP
Adposition phrase	APCP

5.4 Generating Lexical Rules

The same corpus that is used in CFG extraction is used for lexical rules generation. The construction of the lexicon is a one-time process that is done at the very beginning of the parser implementation. However, the result will be needed whenever there is parsing. During the lexicon construction, the lexicon generator reads the corpus from local disk and goes through each sentence. While scanning each sentence, the generator identifies the POS tag name and the word which will be associated with it in lexical rules. The output of the lexicon generator is expected to be formal lexical rules that will be stored in the lexicon in the form of *tag* - > *"word"*.

Input: Afan Oromo tagged sentences
Read tagged sample corpus from local disk
Scan each sentence of sample corpus
Identify the part of speech (POS) tag name and word that will be associated with it
For each word in each sentence of sample corpus
  Call str2tuple () built-in function from a python library
  If the words and its tags are split in the form of (word, tag)
    Reverse the form into (tag, word)
    Return the reversed value
    generate the result with proper format
    return the result
  End If
End For
Output: lexical rules

Algorithm 1: Lexical Generator Algorithm

5.5 Implementation of Chart Parser

In order to apply a top-down chart parsing algorithm, fundamental rule and other three rules are considered, such as: Top-Down Initialization Rule, Top-Down Expand Rule and Top-Down Match Rule. The main focus of this is to parse the user sentence of Afan Oromo language using top down chart parser. The parser interacts with other components in order to parse the user sentences. Initially, the parser accepts the input sentence resulted from the tokenizer. Before the sentence undergoes through the parsing process, the parser checks the grammar rule and lexical rules of sentence. Then the parser scans through the sentence and asks for the POS tag of each word from the lexicon. Then after, the chart will be initialized by an active edge, which is a grammar rule that has S symbol at the left-hand side. Active chart parsing is to keeping track of a record of complete constituents that have found, so record what actually looking for and how much of it we have found so far. The fundamental rule for combining of the passive edges and an active edge can be performed as $S \rightarrow NP.\ VP$. Suppose that there is the passive edge going from where the active edge ends and has category VB on the left side as VP → VB AV. The dot in the active edge is now moved one category forward (i.e., VP → VB. AV, which mean S → NP VB. AV). Our intension is working with active chart parser to make use of an agenda. Then the agenda also initialized by a grammar rule that has a non-terminal at its left-hand side. The left-hand side terminal is similar with that of the non-terminal which is immediately after the arrow in a grammar rule in the chart. The grammar rule in the agenda will move to the chart to replace the first non-terminal, in the right-hand side of S, if it can replace unless another grammar rule is added to the agenda till the grammar rule that can replace the non-terminal is found. If there is a terminal that can replace the non-terminal, the parser will replace it and continues to the next non-terminal. However, if the non-terminal in the chart can't be replaced by the terminal, it looks for other non-terminals which can replace it and which can be replaced by terminals later on. This process will continue until all non-terminals in S are replaced by the terminals and the grammar structure of the sentence S is recognized. This means when new edges are created, the parser has to look at them to see whether they can be combined with other edges to create another new edge in any way. It stores them in the agenda to later remember. The agenda contains edges (grammar rules that can replace non-terminals in the chart and creates new active edge or new grammar rule). The parser will then take one edge at a time from the agenda, add it to the chart and then use it to build new edges. The top down chart parsing algorithm we have adopted from [5, 10] with a few modification is shown in Algorithm 2.

Input: Afan Oromo Sentence from the user
Scan and tokenize the input sentence
Check the words of the sentence whether it is in the lexicon or not
If the word of the user sentence in the lexicon
Take the sentence
Make initial the chart and the agenda
Repeat the following until the agenda becomes empty:

a. *Take the first arc (grammar rule) from the agenda*
b. *Add the arc to chart (if the edge is not already on the chart)*
c. *Combine this arc with arcs from the chart and add the obtained edges to the agenda*
d. *Make a hypothesis about new constituents based on the arc and the rule of the grammar. Add these new arcs to the agenda.*

End repeat
See if the chart contains passive edges from the first node to the last node that has labeled S.
if the chart contains the passive edges that represent all nodes of the sentence then the parsing process succeeds,
if not, the input sentence has a syntax error with respect to the grammatical production rules in the CFG.
Return the parsed sentence (parse tree).
Output: Parse tree or the sentence is not parsed

Algorithm 2: Top Down Chart Parsing Algorithm for Afan Oromo Sentences

5.6 Evaluations

In this section, the evaluation result of lexicon generator and chart parser is discussed. simple lexicon generator algorithms that constructs lexical rules automatically from a sample tagged sentences of Afan Oromo language was developed. The correctness of the lexical rules was inspected manually by checking whether the words are categorized in their proper word classes or not by comparing with manually tagged sentences. Hence, our lexicon generator generates correct lexical rules from manually tagged sentences as expected. Thus, lexicon generator performed correctly as expected without any error.

The sample result of the lexicon generator for five Afan Oromo Sentences from the sample corpus are as follow: “Tulluun nama gurraacha dha” ‘Tullu is a black man’, “Inni gara manaa deeme” ‘He went to home’, “Inni kaleessa galgala dhufe” ‘He comes yesterday at evening’, “Tolosaan mana guddaa qaba” ‘Tolosa has a big house’ and “Abdiisaan mana citaa ijaare” ‘Abdisa build the thatch house (Fig. 2).’

```
PNS -> "Tulluun"
NN -> "nama"
JJ -> "gurraacha"
CP -> "dha"
PPN -> "Inni"
APC -> "gara"
NN -> "manaa"
VB -> "deeme"
PPN -> "Inni"
AD -> "kaleessa"
AD -> "galgala"
VB -> "dhufe"
PNS -> "Tolosaan"
NN -> "mana"
JJ -> "guddaa"
AV -> "qaba"
PNS -> "Abdiisaan"
NN -> "mana"
NN -> "citaa"
VB -> "ijaare"
```

Fig. 2. Screenshot of lexical rules generated by the lexicon generator

The input sentence in the parsing process after it is a tokenized word by word, and the word of input sentence is checked whether it exists in the generated a lexicon in the corpus. The sentence tokenizer did not encounter any error in the parser, so it was perfect for tokenize the input sentence into words. Based on this, the system was trained on the training dataset repeatedly and after correcting the man-made error on manually tagged sentences by using CFG rules which manually extracted from sample corpus was obtained the accuracy of 98.25%. In addition, in order to test the effectiveness of the parser, 100 sentences are selected from the corpus as a test set. On average 20 sentences are from each type of the sentences in the corpus. The correctness of the parser is examined by inspecting its result manually. The output can be checked with respect to the right categorization of words in their proper word class, the right identification of sub phrases and main phrases, the right order of sub phrases in building main phrases, and whether all words and phrases are involved during construction of the sentence S. Therefore, before testing, the sentences are parsed manually on paper based on linguist's suggestion and comments, and then compared the results of the chart parser for the same sentences with manually parsed sentences. Any one of the results, which doesn't satisfy one of the criteria is considered as wrong output or if the result or the parse tree doesn't display at all. The result obtained when the parser was trained and run on the same data is shown in Table 2 as follow.

Table 2. Parsing a result on training set before making number of error correction

Dataset	Number of sentences	Number of correctly parsed sentences	Accuracy in %
Training set	400	350	87.5%

The accuracy achieved should be high when a parser is trained and tested on the same data is expected. But, due to man-made errors during the manual tagging of the sentences and manually extracted context free grammar rules from sample sentences the accuracy was not as high as it was expected. To ensure the accuracy of the parser, discussion with linguistics was takes place in order to check the correctness of tagged sentences as well as the extracted grammar rule. Then, the final accuracy obtained on training set after the error were identified and corrected is the result is improved (Table 3).

Table 3. Parsing a result on training set after making most of error correction

Dataset	Number of sentences	Number of correctly parsed sentences	Accuracy in %
Training set	400	393	98.25%

The result obtained when testing the parser on declarative, interrogative, imperative and exclamatory simple sentence type is approximately 95%, 100%, 100% and 90% accuracy respectively and 70% accuracy on complex sentence type. The result obtained from complex sentences was due to man-made errors during the manual parsing process on complex sentences and incorrectly extraction context free grammar rules of the language. These points were identified to be the cause for wrong parse assignments. On average 91% of accuracy was obtained on 100 sentences of test dataset from all type sentences used in sample corpus (Fig. 3).

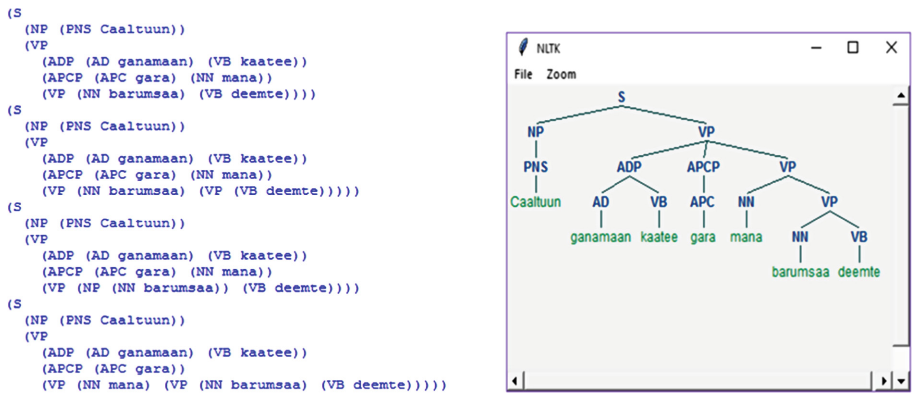


Fig. 3. Screenshot of parsed complex sentence

In general, result shows that the chart parser which is developed to parse Afan Oromo sentence could obtain the accuracy of 98.25% on training set after the correction of errors faced during the first training and 91% on test set, which is an encouraging result.

6 Conclusion

Sentence parsing is one of the steps to design a functional NLP application and which can work in cooperation, and as input to other many NLP applications like grammar checker, machine translation, and etc. Thus, an efficient parsing approach which is top-down chart parsing approaches is developed for Afan Oromo simple and complex sentences is presented in this paper. CFG rule used to represent the grammar of Afan Oromo sentences and simple automatic lexical generator is developed to generate the lexical rules from tagged sample corpus. The total size of 500 Afan Oromo sentences are used and these are collected from previous research works, Afan Oromo grammar book and from other written documents. Developed parser has shown encouraging results in terms of covering simple and complex sentences and automatic construction of the lexical rules from the given sample corpus. An efficient parsing approach which is top-down chart parsing approaches is used in this study rather than using only traditional parsing methods like top-down and bottom-up approaches.

The developed parser is rule-based, so, the parser needs to have components, which are used to enable the system to learn how to parse from grammar rules. This part is composed of lexicon generator, context free grammar rules and lexicon. Lexicon generator is the component used for the automatic construction of the lexical rules. It uses the same POS tagged corpus, which is used for the extraction of context free grammar rules. The parser was trained on 400 sentences of training dataset with the accuracy of 98.25% and tested on 100 sentences of testing dataset with the accuracy of 91%. The experimental results on a parser is an encouraging result since it is the first work for simple and complex sentences of Afan Oromo language.

In general, the study has been designed the general architecture of top-down chart parser for Afan Oromo language. We have been developed how to construct lexical rules automatically from tagged corpus. Another contribution of this study was developed rule-based top-down chart parser, which does not require a tree bank from which the parser learns how to parse through iterative trainings for Afan Oromo simple and complex sentences. This study is the first work in its scope, particularly for parsing simple and complex sentences for Afan Oromo Language, hence, it is still an open issue to develop a full-fledged parser for Afan Oromo language sentence. Besides, compound-complex sentence can be another open research area.

References

1. Katz-Brown, J., et al.: Training a parser for machine translation reordering. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011), pp. 183–192 (2011)
2. Genemo, A.S.: Afaan Oromo Named Entity Recognition Using Hybrid Approach, M.Sc. thesis, Department of Computer Science, School of Graduate Studies, Addis Ababa University, Addis Ababa (2015)
3. Chomsky, N.: Syntactic Structures, 2nd edn. New York (2002)

4. Megersa, D.: An automatic sentence parser for Oromo language using supervised learning technique, M.Sc.thesis, Department of Information Science, School of Graduate Studies, Addis Ababa University, Addis Ababa (2002)
5. Mohammed, A.D.: A top-down chart parser for Amharic sentences, M.Sc. thesis, Department of Computer Science, School of Graduate Studies, Addis Ababa University, Addis Ababa (2015)
6. Alemu, A.: Automatic sentence parsing for Amharic text an experiment using probabilistic context free grammars, M.Sc. thesis, Department of Information Science, School of Graduate Studies, Addis Ababa University, Addis Ababa (2002)
7. Agonafer, D.G.: An integrated approach to automatic complex sentence parsing for Amharic text, M.Sc. thesis, Department of Information Science, School of Graduate Studies, Addis Ababa University, Addis Ababa (2003)
8. Ibrahim, A.: A hybrid approach to Amharic base phrase chunking and parsing, M.Sc. thesis, Department of Computer Science, School of Graduate Studies, Addis Ababa University, Addis Ababa (2013)
9. Sleator, D.D.K.: Parsing English with a Link Grammar, National Science Foundation under grant CCR-8658139, Oline Corporation. R. R. Donnelley and Sons, New York (1991)
10. Al-Taani, A., Msallam, M., Wedian, S.: A top-down chart parser for analyzing Arabic sentences. Int. Arab. J. Inf. Technol. **9**(3), 109–116 (2012)
11. Khoufi, N., Aloulou, C., Hadrich, L., Anlp, B.: ARSYPAR : a tool for parsing the Arabic language. In: International Arab Conference on Information Technology, ACIT, University of Science & Technology (2013)
12. Bataineh, B.M., Bataineh, E.A.: An efficient recursive transition network parser for Arabic language. In: Proceedings of the World Congress on Engineering 2009, vol. II, pp. 1307–1311 (2009)
13. Hambir, N.: Hindi parser-based on CKY algorithm, vol. 3, no. 2, pp. 851–853 (2012)
14. Thant, W.W., Htwe, T.M., Thein, N.L.: Context free grammar based top-down parsing of Myanmar sentences. International Conference On Information Technology, Pattaya, December 2011, pp. 71–75 (2011)
15. Lian, H.: Chinese language parsing with maximum-entropy-inspired parser maximum-entropy-inspired parser, M.S. thesis, pp. 1–6 (2005)
16. Ouersighni, R.: Robust Rule-based Approach in Arabic processing. Int. J. Comput. Appl. **93** (12), 31–37 (2014)
17. Erbach, G.: A flexible parser for a linguistic development environment. In: Herzog, O., Rollinger, C.-R. (eds.) Text Understanding in LILOG. LNCS, vol. 546, pp. 74–87. Springer, Heidelberg (1991). https://doi.org/10.1007/3-540-54594-8_53
18. Weikum, G.: Foundations of statistical natural language processing. ACM SIGMOD Rec. **31** (3), 37 (2002)
19. Jason: Parsing. https://www.cs.cornell.edu/courses/cs4740/2012sp/lectures/parsing-intro-4pp.pdf. Accessed 03 Feb 2017
20. Thompson, B.I.: Afro Asiatic Language Family (2017). http://aboutworldlanguages.com/afro-asiatic-language-family. Accessed 04 Feb 2017
21. Gamta, T.: The Oromo language and the latin alphabet. J. Oromo Stud. 10–13 (1992). http://www.africa.upenn.edu/Hornet/Afaan_Oromo_19777.html. Accessed 06 Feb 2017
22. Ganfure, G.O., Midekso, D.: Design and implementation of morphology based spell checker, vol. 3, no. 12, pp. 118–125 (2014)
23. Yimam, B.: The phrase structures of ethiopian oromo, Ph.D. Dissertification, Addis Ababa University (1986)
24. Alqrainy, S., Jordan, S., Alkoffash, M.S.: Context-free grammar analysis for Arabic sentences. Int. J. Comput. Appl. **53**(3), 7–11 (2012)

25. Kibble, R.: Introduction to natural language processing undergraduate study in computing and related programmes (2013)
26. Zhu, S.C.: Ch 4 classic parsing algorithms chart parsing in NLP pp. 1–51
27. Fox, H.J.: Lexicalized, edge-based, best-first chart parsing, M.Sc. thesis, Department of Computer Science, Massachusetts Institute of Technology, Brown University (1999)
28. Nedjo, A.T., Huang, D., Liu, X.: Automatic part-of-speech tagging for Oromo language using Maximum Entropy Markov Model (MEMM), vol. 10, pp. 3319–3334 (2014)

Author Index