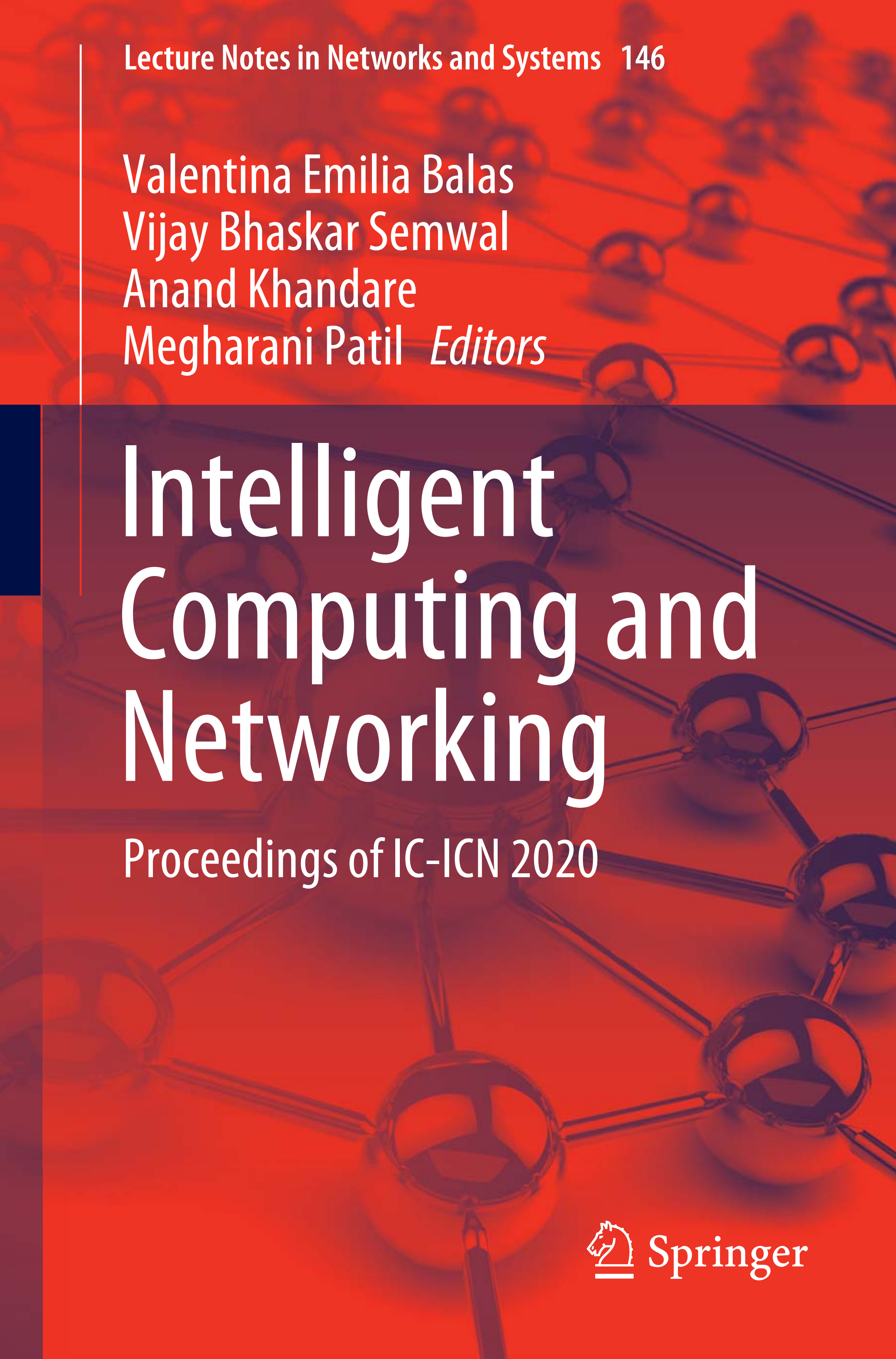
Lecture Notes in Networks and Systems 146

Valentina Emilia Balas
Vijay Bhaskar Semwal
Anand Khandare
Megharani Patil *Editors*

Intelligent Computing and Networking

Proceedings of IC-ICN 2020



Springer

Lecture Notes in Networks and Systems

Volume 146

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

**** Indexing: The books of this series are submitted to ISI Proceedings, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/15179

Valentina Emilia Balas · Vijay Bhaskar Semwal ·
Anand Khandare · Megharani Patil
Editors

Intelligent Computing and Networking

Proceedings of IC-ICN 2020

Springer

Editors
Valentina Emilia Balas
Aurel Vlaicu University of Arad
Arad, Romania

Vijay Bhaskar Semwal
National Institute of Technology Rourkela
Rourkela, Odisha, India

Anand Khandare
Thakur College of Engineering
and Technology
Mumbai, India

Megharani Patil
Thakur College of Engineering
and Technology
Mumbai, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-15-7420-7 ISBN 978-981-15-7421-4 (eBook)
https://doi.org/10.1007/978-981-15-7421-4

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Organizing Committee

Chief Patron

Mr. V. K. Singh, Chairman

Patrons

Mrs. Karishmma V. Mangal, Secretary
Mr. Karan V. Singh, CEO

Organizing Program Chairs

Dr. B. K. Mishra, Principal, TCET
Dr. Deven Shah, Vice-Principal, TCET

Program Co-chairs (IC-ICN 2020)

Dr. R. R. Sedamkar, Dean, Academic
Dr. Kamal Shah, Dean, R&D

Convenor

Dr. Sheetal Rathi, Professor, HOD-COMP

Joint Convenor

Dr. Rajesh Bansode, Professor, HOD-IT

Technical Program Committee

Dr. Sateesh Kumar, Associate Professor, Department of CSE, IIT, Roorkee
Dr. Vijay Bhaskar Semwal, Assistant Professor, NIT, Bhopal, Madhya Pradesh
Dr. Padmaja Joshi, Director, CDAC, Mumbai, India
Dr. Suneeta Agarwal, Professor, NIT, Allahabad
Dr. Dusko Lukac, CCEC—Centre of Competence, Germany
Dr. Alveen Singh, Durban University of Technology, South Africa
Prof. Valentina Emilia Balas, Ph.D., Aurel Vlaicu University of Arad
Dr. Selwyn Piramuthu, Professor, Information Systems, University of Florida

Overall Coordinator

Dr. Anand Khandare, Associate Professor, COMP

Preface

The International Conference on Intelligent Computing and Networking (IC-ICN 2020) is an international conference scheduled on February 28–29, 2020.

This is the eleventh event in the series of international conferences organized by Thakur College of Engineering and Technolgy (TCET), Mumbai under the umbrella of MULTICON-W since the first event ICWET 2010. This event is organized in view of strengthening the research culture among its stakeholders which include students, faculty, and industry. The conference provides a platform to the authors and attendees for collaborations and networking among universities and institutions from India and foreign for promoting trending technologies and research. It aims to promote research which is basic as well as advanced to convert applied investigation into practice. Moreover, need of the time is technological development in the areas of intelligent systems and communication engineering which can simplify our life in the eco-friendly environment with better connectivity and security. In the conference, one can present research papers, technical papers, case studies, best and innovative practices, engineering concepts and designs. The Eleventh Annual International Conference, IC-ICN 20 has affiliation with Scopus Indexed journal for intelligent systems, leading publication house like Springer, Tata McGraw Hill, IOSR, and Conference proceeding with ISBN number. It serves as a premier platform that gathers all academicians, researchers, and professionals, in the relevant engineering disciplines and domains, to showcase their research contributions. Not just inculcating the research culture, the conference also provides a flavor of lectures by eminent speakers from different areas and panel discussion by industry people. IC-ICN 20 has gained wide publicity through Web site, social media coverage, and visits done to various colleges by the team of faculty members and our well-wishers. As a result, this event has got overwhelming response.

TCET has strong belief in quality and relation building. A lot of care has been taken for branding the event such as logistic support required for the event, compilation and printing of conference proceeding, and souvenir. Effort has been taken to make the delegates feel at home though away from home with the confidence of doing something for the development of nation.

Till date, under the banner of ICWET and Multicon-W, more than 50 conferences, over 250 tracks and 4000 papers have been presented and published in the proceedings with ISBN numbers. The event will comprise four conferences and three workshops with multiple tracks. Around 5000 participants and delegates have attended this event. During these two days, there were about 235 presentations from national as well as international researchers and industrial personnel and idea presentations with deliberation by the delegates. Our best wishes and good luck to all.

We thank all the members of the organizing and editorial committee for supporting the event and extending their cooperation to make it a grand successful event.

Team-IC-ICN 2020
TCET
Mumbai, India

Contents

Editors and Contributors

About the Editors

Valentina Emilia Balas is currently a Full Professor in the Department of Automatics and Applied Software at the Faculty of Engineering, “Aurel Vlaicu” University of Arad, Romania. She holds a Ph.D. in Applied Electronics and Telecommunications from Polytechnic University of Timisoara. Dr. Balas is author of more than 300 research papers in refereed journals and international conferences. Her research interests are in intelligent systems, fuzzy control, soft computing, smart sensors, information fusion, modeling and simulation. She is the Editor-in-Chief to the International Journal of Advanced Intelligence Paradigms (IJAIP) and to the International Journal of Computational Systems Engineering (IJCSysE), is a member in Editorial Board of several national and international journals and is the Director of Intelligent Systems Research Center in Aurel Vlaicu University of Arad. She is a member of EUSFLAT and SIAM, senior member of IEEE, member in TC – Fuzzy Systems (IEEE CIS), member in TC – Emergent Technologies (IEEE CIS) and member in TC – Soft Computing (IEEE SMCS).

Vijay Bhaskar Semwal is working as an Associate Professor (CSE) at NIT Bhopal since 5 February 2019. Before joining NIT Bhopal, he was working at NIT Rourkela. He has also worked with IIIT Dharwad as an Assistant Professor (CSE) for 2 years (2016–2018), and he has also worked as an Assistant professor (CSE) at NIT Jamshedpur. He has earned his doctorate degree in robotics from IIIT Allahabad (2017), M.Tech. in Information Technology from IIIT Allahabad (2010) and B.Tech. (IT) from the College of Engineering Roorkee (2008). His areas of research are bipedal robotics, gait analysis and synthesis, artificial intelligence, machine learning and theoretical computer science. He has published more than 15 SCI research papers. He has received early career research award by DST–SERB under the government of India.

Dr. Anand Khandare is working as an Associate Professor in Thakur College of Engineering and Technology, Mumbai University. His total experience is 15 years in teaching. He has completed Ph.D. in Computer Science and Engineering in the domain Data Clustering in Machine Learning from Sant Gadge Baba Amravati University, 2019, Master of Engineering (M.E.) in Computer Engineering from Vidyalankar Institute of Technologies, Mumbai University, 2011, and Bachelor of Engineering (B.E.) in Computer Science and Engineering from Amravati University with 64.5 percentages, 2005. He has published a total of 50 papers in national and international conferences and journals. Two patents are also in his name. He has worked in various committees in conferences and workshops. He taught various subjects such as C, C++, JAVA, R PYTHON, Machine Learning, Mobile Computing, OS, Data Structures and Software Engineering. He has knowledge of RDATABASE such as SQL Server, MYSQL, and Oracle. He worked on various OS PLATFORMS such as Windows, LINUX and Androids. He guided 30+ undergraduate projects and 3 postgraduate projects. His area of interest is machine learning and networking. His interests also include web application development and mobile application development. He is a lifetime member of ISTE professional body.

Dr. Megharani Patil has graduated from Konkan Gyanpeeth College of Engineering, Mumbai University, in Computer Engineering, in 2003. She completed her Master's Degree from Shivaji University in the academic year 2010–2011. She completed Ph.D. (Technology) from Thadomal Shahani Engineering College, Mumbai University. Currently, she is working as an Associate Professor at Thakur College of Engineering and Technology, Mumbai University. She has 13.5 years of teaching experience in the institute and total 15.5 years of teaching experience. She has published more than 44 papers in international journals and conferences, 2 books C++ and Java programming and 1 patent in Indian patent journal. In institute, she is Domain Incharge for Intelligent System Design & Development. She guided 30+ undergraduate projects and 4 postgraduate projects. Her area of interest is software engineering, user experience design and artificial intelligence. Her interests also include web application development and mobile application development. She is a lifetime member of ISTE professional body.

Contributors

Mushtaq Ahmed Malaviya National Institute of Technology, Jaipur, Rajasthan, India

Rajanikanth Aluvalu Vardhaman College of Engineering, Hyderabad, India

A. S. Alvi Information Technology, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India

Salim Y. Amdani Department of Computer Science and Engineering, BNCOE, Pusad, India

Nazneen Ansari Department of Information Technology, SFIT, Borivali, Mumbai, India

Abhishek Anurag CSE, Pune, India

Namdeo Baban Badhe IT Department TCET, Mumbai University, Mumbai, India

G. R. Bamnote Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India

Priyanka Bandagale Finolex Academy of Management and Technology, Mumbai University, Ratnagiri, Maharashtra, India

Rajesh Bansode Mumbai University, Mumbai, India;
Department of Information Technology, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India

Harish Barapatre Department of Computer, Mumbai University, Mumbai, India

Fabian T. R. Barreto Department of Electronics and Telecom, Xavier Institute of Engineering, Mumbai, India

Urjita Bedekar Vivekanand Education Society's Institute of Technology, Mumbai, India

Simran Bhagwandasani Vivekanand Education Society's Institute of Technology, Mumbai, India

Vinayak Ashok Bharadi IT Department FAMT (Finolex Academy of Management and Technology), Mumbai University, Ratnagiri, Maharashtra, India

Asha Bharambe Vivekanand Education Society's Institute of Technology, Mumbai University, Mumbai, India

Gresha S. Bhatia Vivekanand Education Society's Institute of Technology, Mumbai, India

Rahul Bhatia Vivekanand Education Society's Institute of Technology, Mumbai, India

Anita Caudhari Mumbai University, Mumbai, India

Amogh Chaudhari MGM Institute of Health Sciences, Mumbai, India

Krishna Keerthi Chennam Muffkham Jah College of Engineering and Technology, Hyderabad, India

Vaibhav D. Dabhade Computer Engineering, MET's BKC, IOE, Nashik, India

Anushree Deshmukh Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, India

Amol S. Dudhe Department of Information Technology, BNCOE, Pusad, India

Ashish Gangaramani Vivekanand Education Society's Institute of Technology, Mumbai University, Mumbai, India

Madhuri N. Gedam Department of Computer Engineering, Veermata Jijabai Technological Institute, Matunga, Mumbai, India

Saroj Ghadle National Institute of Technology, Raipur, Chhattisgarh, India

Uttara Gogate Associate Professor, Shivajirao S. Jondhale College of Engineering, Mumbai, India

V. S. Gulhane Sipna College of Engineering, Amravati, Maharashtra, India

Shiwani Gupta TCET, Mumbai, India

D. T. Ingole Department of Electronics & Telecommunication, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India

Manik D. Ingole Department of Electronics & Telecommunication, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India

M. A. Jabbar Vardhaman College of Engineering, Hyderabad, India

Swati Jayade Department of Electronics & Telecommunication, Government Polytechnic, Washim, India

Sanjay Kumar National Institute of Technology, Raipur, Chhattisgarh, India

Samidha Kurle Computer Engineering Department, Atharva College of Engineering, Malad West, Mumbai, India

Mamta Meena Computer Engineering Department, Atharva College of Engineering, Malad West, Mumbai, India

Rupali A. Meshram Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India

B. B. Meshram Department of Computer Engineering, Veermata Jijabai Technological Institute, Matunga, Mumbai, India

Vandana Munde TCET College, Mumbai, Maharashtra, India

Pranit Naik Vivekanand Education Society's Institute of Technology, Mumbai, India

Deepa Parasar Amity University, Panvel, India

Ashwini P. Parkar Department of Computer Engineering, SLRTCE, Mumbai, India

Suhasini Parvatikar Mumbai University, Kharghar, Navi Mumbai, India

Aruna Animish Pavate Department of Information Technology, Thakur College of Engineering, Mumbai University, Mumbai, India

Priti Rumao Computer Engineering Department, Atharva College of Engineering, Malad West, Mumbai, India

Megha Sahu Vivekanand Education Society's Institute of Technology, Mumbai University, Mumbai, India

Sunny Sall Department of Computer Engineering, St. John College of Engineering and Management, Palghar, Maharashtra, India

Julli Sarode Department of Computer, Mumbai University, Mumbai, India

Vidya Sarode Department of Electronics and Telecom, Xavier Institute of Engineering, Mumbai, India

Farook Sayyad Department of Mechanical Engineering, Dr. D Y Patil School of Engineering, Lohegaon, Pune, India

Shabnam Sayyad Department of Computer Engineering, AISSMS College of Engineering, Pune, India

R. R. Sedamkar TCET, Mumbai, India

Priyank Shah Fr. Conceicao Rodrigues College of Engineering, Mumbai, India

Mohammed Umraan Shaikh Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India

Saim Shaikh Fr. Conceicao Rodrigues College of Engineering, Mumbai, India

Akram H. Shaikh Department of Computer Engineering, Veermata Jijabai Technological Institute, Matunga, Mumbai, India

Yogesh Kumar Sharma Department of Computer, JJTU University, Rajasthan, India

Vaishali Shinde Department of Computer, Mumbai University, Mumbai, India

Deval Srivastava Fr. Conceicao Rodrigues College of Engineering, Mumbai, India

Swati Suchak PG Scholar, Alamuri Ratnamala Institute of Engineering & Technology, Mumbai, India

Nitesh M. Tarbani Sipna College of Engineering, Amravati, Maharashtra, India

Shanthi Therese Department of Computer Engineering, TSEC, Bandra, Mumbai, India

Garima Tripathi Fr. Conceicao Rodrigues College of Engineering, Mumbai, India

Rakesh Tripathi National Institute of Technology, Raipur, Chhattisgarh, India

V. Uma Maheswari Vardhaman College of Engineering, Hyderabad, India

Daitule Omkar Vilas Malaviya National Institute of Technology, Jaipur, Rajasthan, India

Deepali Vora Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India

Sunil B. Wankhade Department of Information Technology, Rajiv Gandhi Institute of Technology, Mumbai, India

Akash Yadav Malaviya National Institute of Technology, Jaipur, Rajasthan, India

Exploiting Processor Variability for Efficient Application Mapping

Akash Yadav, Mushtaq Ahmed, and Daitule Omkar Vilas

Abstract Increasing on-chip integration with technology scaling in which more transistors are available on a chip than that can be simultaneously powered on will lead to the dark silicon era. Process variation gradually impacts the performance (power, speed, frequency etc.) of system over its usage life. In this paper, by adapting the application allocation algorithm at the software layer, reduction of the impact of process variation in multicore architecture is suggested and tested. Process variation causes core-to-core variations in power and frequency. To gain the maximum performance within the power limits, best subset of cores should be selected from the available cores to run an application. When the power constraints are used as cost parameter, simulation of adaptive allocation algorithm on Sniper multicore simulator promises twice the better performance compared to conventional core allocation strategy.

Keywords Process variations · Dark silicon · Efficient application mapping · Sniper multicore simulator

1 Introduction

Process variation [1] is naturally occurring variation in attributes of transistors, which results in measurable variance in the output performance and behavior of all circuit. We tried to execute processes in a way that it will choose cores based on its timing and maximum power constraints, which will be adaptive to the changes in the features of core like frequency. The basic idea here is to optimize the execution in such a way that the inefficient cores in the pool are turned off instead of the ones which are fast and which can give us better performance. This process of picking better cores

A. Yadav (✉) · M. Ahmed · D. O. Vilas
Malaviya National Institute of Technology, Jaipur, Rajasthan, India
e-mail: 2019rcp9150@mnit.ac.in

M. Ahmed
e-mail: mahmed.cse@mnit.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_1

out of available cores is called cherry picking [2]. The primary focus of this paper is on exploiting process variations in dark silicon cores for performance improvement under a power budget so as to solve the problem of variation-aware core selection. We also present a performance parameters (i.e., throughput and power consumption) based comparison between different topologies such as MESH, TORUS, and FAT-TREE.

2 Literature Survey

With the advancement of the technology, transistors are becoming faster and smaller, but transistor power consumption does not scale with the same level [3]. Due to supply voltage limits, power densities rapidly increase on the chip. So to prevent thermal emergencies, a significant amount of on-chip resources needs to stay dark, This phenomenon is known in the literature as dark silicon [4, 5]. Process variation and dark silicon are the most important factors impacting the performance of multicore systems in today's world.

2.1 Process Variation

Mostly post-manufacturing tuning is used to reduce the impact of process variation [6]. This conventional method often leads to system overdesign relative to their specification. Less leaky processing units are more energy efficient than their leakier counterparts at a given supply voltage and frequency. Algorithm suggested in this paper takes advantage of this observation, to shift the processes to execute on less leaky ones, maintaining performance while reducing total power consumption.

2.2 Cherry-Picking: Process Variation Aware Core Selection

Increased number of cores on a single chip led to dark silicon era, it refers to the amount of silicon that cannot be powered-on at the nominal operating voltage for a given thermal design power (TDP) constraint. Certain cores of the system remain unpowered to avoid overheating. According to recent studies at 8 nm technology nodes, the amount of Dark Silicon may reach up to 50–80% depending upon the processor architecture, cooling technology and application workloads. Some early work has been done in this field which uses a heterogeneous approach of using Hardware accelerators to get better performance out of the system [7]. When accounted for Process Variations, the approach becomes quite difficult.

3 Proposed Work

An adaptive application allocation algorithm is proposed which can be used at any level of the core to application allocation. It uses two algorithms, one for estimation of execution cost and another for allocation of a particular core to a particular application so as to minimize the total cost of running these applications. Input to these algorithms are application statistics, core frequencies and cost of running each application on different available cores.

It first gets the application statistics like maximum runtime allowed for each application to maintain the quality and minimum frequency requirements for each application. Minimum required frequency is stored as $f_{\min}$ for all i applications. It then estimates cost (here taken from Sniper- runtime dynamics values) of using a particular core for a particular application based on estimation algorithm. Core frequencies are read and stored in f_{core} for all the j cores. It then checks whether a particular core can be used for a particular application by checking it against maximum runtime requirements. If it doesn't satisfy the constraints of the application then its cost is multiplied with a large value so that this particular core can never be used for this application.

After all applications have been checked against all cores, it allocates a particular core to a particular application by using an allocation algorithm (Munkres allocation algorithm is used in this paper) [8]. It then displays the optimal mapping of each core with a certain application so that it results in a minimum cost. The estimation algorithm is as follows:

Algorithm: Estimation Algorithm

```
for i ← 1 to n do                              \\ for app
   for j ← 0 to n do                           \\ for core
      if (f_min[i] > f_core[j]) then
         costmatrix[i][j] ← ∞;
      else
         Execute app[i] on core j;
         Get execution time and cost;
         if execution time > timing constraint then
            cost[i][j] ← ∞;
         else if power budget > max power constraint for chip then
            cost[i][j] ← ∞;
         else
            update cost[i][j] to estimated cost value
```

Apply allocation Algorithm [8] on cost matrix. Display output as one to one core allocation.

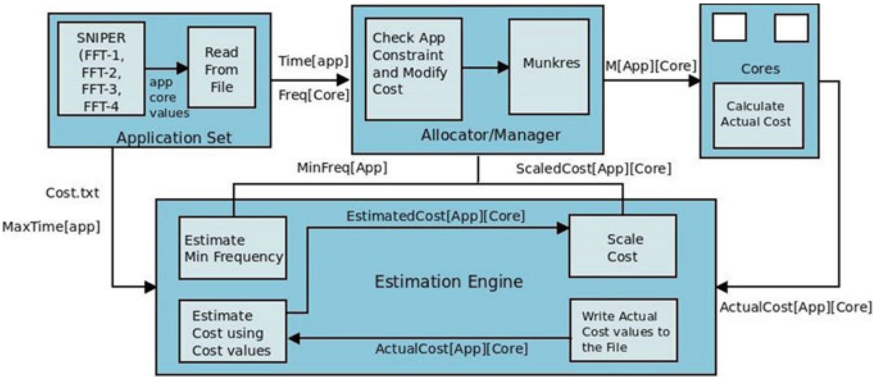


Fig. 1 Adaptive application mapping architecture

4 Adaptive Application Mapping Architecture

Figure 1 shows the flow of data from one component to another in adaptive application allocation algorithm. Initially data was obtained from Sniper using its test application FFT with different split radices. This data is stored in a text file and will be fed to two components namely allocator/manager and estimation engine. Estimation engine estimates minimum frequency requirements for each application and estimate the cost using previous cost values and actual cost which it gets from a component called Cores. It also saves the actual cost values in an array and text file. After estimating the cost, it scales it to an integer value which is then fed to allocator along with a minimum frequency requirement of each application. Allocator checks whether a particular core can be used for an application and if it can’t be used then it multiplies the cost with a large number. It then uses an allocation algorithm (Munkres used in this case) which gives an optimal mapping i.e., a suitable core for each application which is fed to cores so that the mapped core can be used for the application it has been mapped with.

Cores then output the actual cost values for the cores which are used for a particular value and send it to estimation engine so that these values can be used for estimation of costs in future.

5 Experimental Setup

A test application FFT of Sniper simulator with different split radices are used as independent processes. Sniper is a parallel, high-speed multi-core simulator [9, 10]. These applications are compiled and then scheduled on different core frequencies like 2, 2.5, 3 and 3.66 GHz. The total execution time of an application to run on core, power leakage values and runtime dynamics are estimated with the help of McPAT

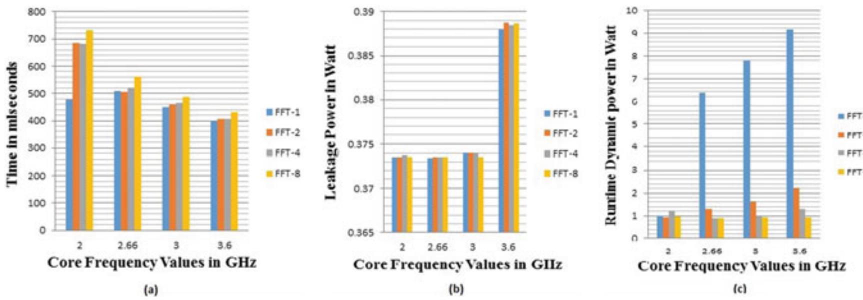


Fig. 2 **a** Total execution time for processes on different core frequencies, **b** leakage power values for processes on different core frequencies, **c** runtime dynamic power values for processes on different core frequencies

Table 1 Cost matrix

Application	Core-1 (2 GHz)	Core-2 (2.66 GHz)	Core-3 (3 GHz)	Core-4 (3.66 GHz)
FFT-1	1.0035	6.36631	7.79809	9.1548
FFT-2	0.940722	1.31966	1.619870	2.1909
FFT-4	1.20623	0.916404	0.980522	1.2567
FFT-8	0.967971	0.898728	0.927073	3.9912

framework [11] and stored in 2D cost matrix. This cost matrix is fed to allocation algorithm which gives an optimal mapping i.e., a suitable core for each application to run. The timing analysis and the power analysis are manipulated which is shown in the Fig. 2 and cost matrix is given in Table 1.

We have also compared the results for throughput and the power consumption by the NoC device for different topologies such as MESH, TORUS, and FAT-TREE. We did our experiments with different possibilities and observed the results for different frequencies. We have taken results for 8 number of cores whereas the frequency values 2.66, 3, 5, 6 and 7 GHz are taken as the frequency of the processing cores.

6 Results and Analysis

This implementation is quite flexible allowing the user to use different estimation parameters as per need. Here we have taken runtime dynamics as a parameter to estimate cost for running an application on a core. Table 2 contains different cost values when the input data, mentioned in Sect. 5, was fed to adaptive allocation algorithm allocating cores in the manner given in the table and when same data fed to the simple allocation algorithm (if applications are randomly mapped on core available at that time).

Table 2 Effective cost by different method

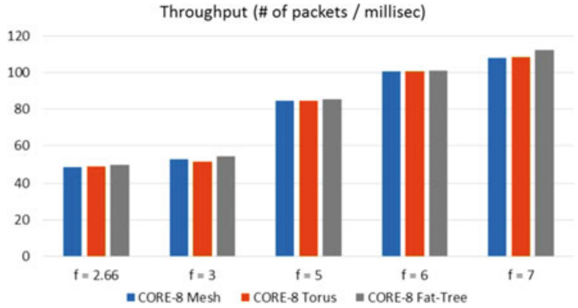
Mapping of processes by simple available core allocation method			Mapping of processes by adaptive allocation method		
Application	Allocated core	Cost	Application	Allocated core	Cost
FFT-1	Core 1	7.294882	FFT-1	Core 1	4.506933
FFT-2	Core 2		FFT-2	Core 2	
FFT-4	Core 3		FFT-4	Core 4	
FFT-8	Core 4		FFT-8	Core 3	

Cost value taken by simple allocation is almost two times the value obtained when adaptive allocation algorithm is used. The results will be better if one of the highest costs is along the diagonal in the cost matrix because in that case simple allocation will take that cost. In some cases, simple allocation from available cores may result in lesser cost than the optimal mapping. This anomaly can be due to the fact that simple allocation overlooks the timing and frequency constraints of an application. Thus, if this algorithm is used then we can use the process exploitation to get better results and at the same time better cores can be used minimizing the overall cost of running a complete application.

To evaluate the comparison results of different topologies accurately, we have taken three runs from the Sniper simulator for each of the possible combinations and then calculated the average of the collected three values. Figure 3 shows the comparison of the throughput of the overall system. For the lower value of frequency, we can observe the similar throughput of the system. But as the frequency of each processing core increases, the slight difference can be observed. We can find the overall performance of the system more or less to be same for these topologies.

Figure 4 shows the comparison of the throughput of the overall system for implemented NoC topologies. As expected, the throughput increases with the increase in the frequency of the processing cores. We have compared the throughput for frequencies of 2.66, 3, 5, 6 and 7 GHz. All the values were taken after a fixed number of cores that is 8. One by one we have calculated for all implemented topologies.

Fig. 3 Comparison for the throughput values



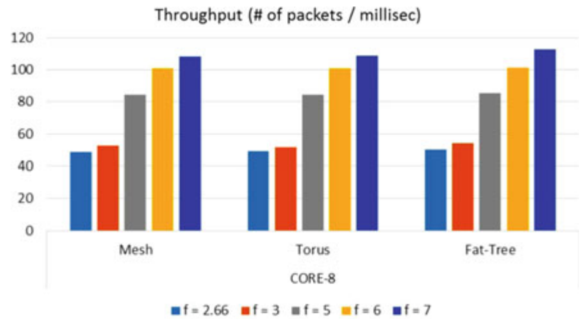


Fig. 4 Comparison for the throughput with the increase in the frequency

For the same configuration, we have taken the power consumption results also provided by the Sniper simulator in Figs. 5 & 6 [10].

With the above graph, we can conclude that power consumption in the Fat-Tree topology is towards the higher side in compared to the other topologies. Also, we can conclude the obvious point that with the increase in frequency, power consumption also increases.

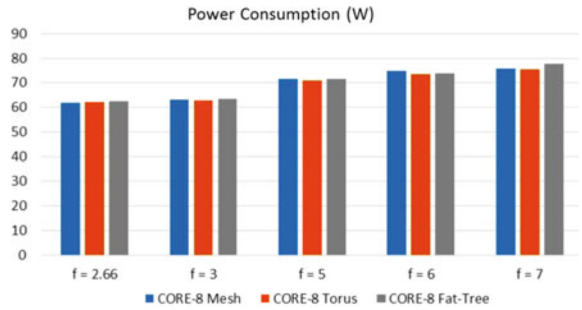


Fig. 5 The comparison graph for the power consumption of the over-all system

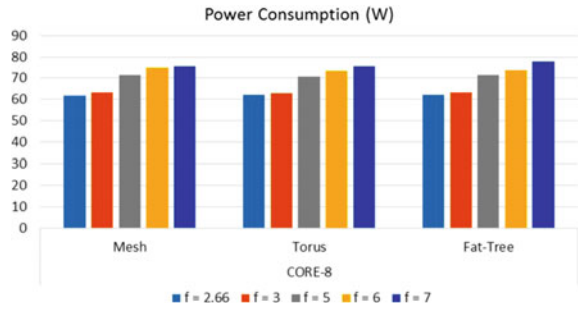


Fig. 6 Figure shows the increase in power consumption with the increase in the frequency of the processing core

7 Conclusions and Future Work

Results of proposed work show that algorithm results in an optimum mapping of applications with given cores which helps us in achieving our goal of getting the overall minimum cost. Thus, if an application is able to select better core from a pool of available cores then it can effectively reduce its running time and thus giving better results by consuming less power. For future work this algorithm can be extended for dynamic cores which can operate in more than one frequency configurations. This can also be used on a large scale once if it can be implemented on a chip level and thus can be used by wide variety of applications i.e. extending its scope from simple video coding to other applications which are affected by process variation and dark silicon.

For the comparison of different topologies, with the collected results we can conclude that Fat-Tree provides better throughput in comparison to other GRID based topologies such as Mesh and Torus [12]. The main reason for an increase in the throughput is that tree-based topology decreases the hop latency of the packets. With the decrease in the hop latency throughput increases. For future work we can implementing the topologies for more than 8 number of processing cores for different topologies such as Mesh, Fat-Tree etc. So, that we can compare the results for the higher number of processing cores.

References

1. Borkar S, Karnik T, Narendra S, Tschanz J, Keshavarzi A, De V (2003) Parameter variations and impact on circuits and microarchitecture. In: Proceedings 2003. Design automation conference (IEEE Cat. No.03CH37451), Anaheim, CA, pp 338–342
2. Raghunathan B, Turakhia Y, Garg S, Marculescu D (2013) Cherry-picking: exploiting process variations in dark-silicon homogeneous chip multi-processors. In: 2013 design, automation and test in Europe conference and exhibition (DATE). Grenoble, France, pp 39–44
3. Pagani S et al (2014) TSP: thermal safe power—efficient power budgeting for many-core systems in dark silicon. 2014 International conference on hardware/software codesign and system synthesis (CODES+ISSS). New Delhi, pp 1–10
4. Khdr H, Pagani S, Shafique MA, Henkel J (2018) Chapter four—dark silicon aware resource management for many-core systems. Adv Comput 110:127–170
5. Kanduri A, Rahmani MA, Liljeberg P, Hemani A, Jantsch A, Tenhunen H (2017) A perspective on dark silicon. Springer International Publishing, Cham, pp 3–20
6. Maiti S, Kapadia N, Pasricha S (2015) Process variation aware dynamic power management in multicore systems with extended range voltage/frequency scaling. In: 2015 IEEE 58th international Midwest symposium on circuits and systems (MWSCAS). Fort Collins, CO, pp. 1–4
7. Dighe S, Vangal SR, Aseron PA, Kumar S, Jacob T, Bowman KA, Howard J, Tschanz JW, Erraguntla V, Borkar N, De V, Borkar SY (2011) Within-die variation-aware dynamic-voltage-frequency-scaling with optimal core allocation and thread hopping for the 80-core TeraFLOPS processor. IEEE J Solid-State Circuits 46:184–193
8. Munkres JR (1957) Algorithms for the assignment and transportation problems. J Soc Ind Appl Math 5(1):32–38

9. Carlson TE, Heirman W, Eeckhout L (2011) Sniper: exploring the level of abstraction for scalable and accurate parallel multi-core simulation. In: SC '11: proceedings of 2011 international conference for high performance computing, networking, storage and analysis. Seatle, WA, pp 1–12
10. Akram A, Sawalha L (2016) A Comparison of x86 computer architecture simulators. In: Computer architecture and systems research laboratory (CASRL)
11. Li S, Ahn JH, Strong RD, Brockman JB, Tullsen DM, Jouppi NP (2009) McPAT: an integrated power, area, and timing modeling framework for multicore and manycore architectures. In: 2009 42nd annual IEEE/ACM international symposium on microarchitecture (MICRO). New York, NY, pp 469–480
12. Manivannan M, Pericàs M, Papaefstathiou V, Stenstrom P (2017) Runtime-assisted global cache management for task-based parallel programs. In: IEEE computer architecture letters, pp 1–1. 10.1109/LCA.2016.2606593

Genetic Algorithm for Feature Selection and Parameter Optimization to Enhance Learning on Framingham Heart Disease Dataset

Shiwani Gupta and R. R. Sedamkar

Abstract Classification algorithms as Support Vector Machine (SVM) and Neural Network (NN) have provided considerably good results in the diagnosis of Critical Care diseases. These Machine Learning Algorithms have hyperparameters whose values if chosen optimally can provide enhanced learning. Operating on entire set of features is computationally expensive and requires more number of instances. Hence, utilizing important features will reduce computation time. Both these objectives can simultaneously be obtained through nature-inspired algorithm as Genetic Algorithm (GA). In the proposed work, GA has been utilized for Feature Selection as well as tuning the Hyperparameters of SVM and NN. Optimal value of SVM Radial Basis Function (RBF) kernel parameters C and γ has been obtained. Similarly, the novelty lies in identifying optimal number of hidden layers, number of hidden nodes, learning rate, momentum, and optimizer for Multi Layer Perceptron (MLP) NN classifier. Results have been found better compared to utilizing Grid Search for the same. Further, when reduced set of features are used for learning; Sensitivity and Precision score have found to be promising. Sensitivity is of more importance when health care is talked about and F_1 score gives better accuracy since it is unbiased to data imbalance. Thus, we have dealt with Multiobjective Optimization problem utilizing metaheuristics with improvement in diagnostic performance.

Keywords Genetic Algorithm · Hyperparameter Optimization · Feature Selection

S. Gupta (✉) · R. R. Sedamkar
TCET, Mumbai, India
e-mail: shiwani.gupta@thakureducation.org

R. R. Sedamkar
e-mail: rr.sedamkar@thakureducation.org

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_2

1 Algorithms

1.1 Support Vector Machine

In order to save sufficient computation time and build models that generalize well, penalty parameter C and kernel function parameter γ need to be optimized. Though grid algorithm can identify these optimal values through exhaustive search, it cannot perform feature selection task. Support vector machine identifies the hyperplane $D(x) = wx + b$ with maximum margin for linear separability as shown in Fig. 1. For nonlinear separable problem, kernel trick is applied as shown in Fig. 2.

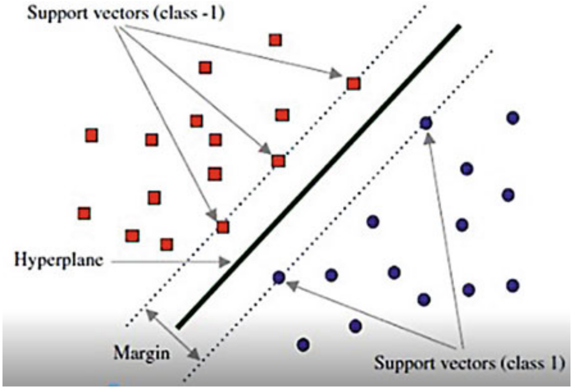


Fig. 1 Support vector machine [WordPress.com]

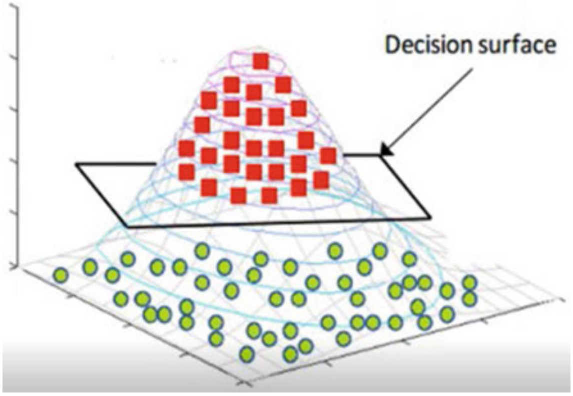


Fig. 2 Kernel trick [hackerearth.com]

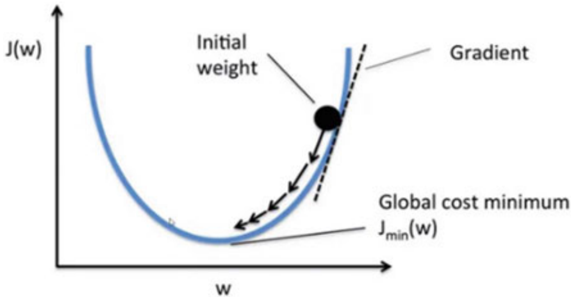


Fig. 3 Stochastic Gradient Descent [quora.com]

1.2 Neural Network

Neural network is modeled after the human brain and is capable of capturing complex nonlinear models. Activation functions operate in hidden nodes to transform net input to output. To avoid overfitting, bias is introduced. The activation function may be Logistic/Sigmoid, Rectified Linear Unit (ReLU), etc. For a classification problem, output is the probability of belonging to a certain class. The model is retrained w.r.t. the error backpropagated until convergence. Thus, we get optimal weights following Stochastic Gradient Descent (SGD) algorithm shown in Fig. 3. An extension to SGD is adaptive moment estimation (adam) optimization algorithm realizing the benefits of AdaGrad and RMSProp.

1.3 Genetic Algorithm

Genetic Algorithm is evolutionary algorithm inspired by natural selection, reproduction, and survival of fittest utilizing genetic operators—selection, crossover, and mutation. Encoding is termed as a chromosome, a single bit being called Allele. It utilizes population of solutions in each iteration. It randomly selects solution to produce other solution and evaluates each potential solution. Two parents are chosen, and genetic operators are applied and moved to the next generation until convergence.

Crossover is performed over random cutoff points. Mutation is bit flip with certain probability usually low. Elitism is to take the best fitness value to the next generation by replacing the worst one. Fitness value is the value of objective function. Selection is through Roulette wheel where size is proportional to individual's fitness or Tournament Selection where several tournaments among few chromosomes are selected at random. Winner of these tournaments is selected for crossover.

2 Literature Survey

Gupta and Sedamkar [1, 2] have proposed that simultaneous optimization of hyperparameters of NN or SVM RBF kernel and feature subset selection through GA can yield high diagnostic accuracy. Further, they claim in another paper that feature selection is essential for simpler, faster, robust, and more reliable ML models.

Huang and Wang [3] have simultaneously optimized parameters and feature subset without compromising the accuracy of SVM utilizing GA and compared performance with grid search. Steps taken for GA-based feature selection are:

- Scaling to avoid attributes in greater numeric range dominate those in smaller numeric ranges
- Evaluate the fitness of each chromosome
- Look for better solutions by selection, crossover, and mutation.

The data is partitioned into training and independent test sets via kfold cross-validation (CV). Fei and Min [4] search the best penalty and kernel function parameter, and finally, a less complex SVM model with fewer support vectors is proposed.

Bhatia et al. [5] have tried experimenting onto Cleveland multiclass using ‘one against one’ and binary class problem utilizing integer coded GA in order to avoid curse of dimensionality. The fitness value chosen is the classification accuracy achieved by feature subset with 25 generation, 50 population, 0.8 crossover probability, and 0.2 mutation probability. Similarly, Arrythmia dataset with 17 classes received sensitivity of 91.40%, accuracy 98.99%, and specificity 99.46% with genetic ensemble of classifiers [6].

ANN is one of the most promising computational intelligence techniques though the design requires the setting of the structure and tuning complex parameter [7]. Thus, GA has been used to select significant features and identify optimal number of hidden nodes along with Levenberg–Marquardt (LM) backpropagation algorithm, and the chromosome is shown in Fig. 4. The author further utilized GA to fine-tune weight of ANN. The chromosome is shown in Fig. 5. Thus Table 1 shows literature review of several feature selection techniques applicable to several critical care disease datasets inorder to enhance performance by utilising informative features.

For RBF N/W, GA solves subset selection problem. Lacerda et al. [11] utilized holdout kfold CV and bootstrap function to accomplish the same. Zhao et al. [14] have also analyzed that inappropriate parameter settings lead to lower classification accuracy. Better performance may be achieved by discarding noisy, irrelevant, and redundant features. The objective achieved was the highest classification accuracy, least number of selected features and least average processing time. Amin et al. [15]

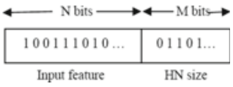


Fig. 4 Optimize features and hidden nodes [7]

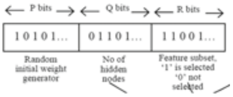

P bits
Q bits
R bits
10101...
01101...
11001...
Random initial weight generator
No of hidden nodes
Feature subset, ‘1’ is selected ‘0’ not selected

Fig. 5 Optimize initial weight, hidden nodes, feature subset [7]

Table 1 Review of literature on feature selection

Algorithm	Objective	Methodology	Dataset	Results
GA NN [7]	Identify significant features	Levenberg–Marquardt (LM) backpropagation	Cancer dataset	GA has been used to select 7 significant features out of 9 to achieve enhanced CV accuracy
GA NN [8]	Feature subset selection	Resilient BP, LM, Gradient descent with momentum. Fitness function is inverse of error	Breast Cancer dataset	99.43% best and 98.29% average correct classification
SVM GA [5]	Avoid curse of dimensionality	Integer coded GA The fitness value chosen is the classification accuracy achieved with 25 generation, 50 population, 0.8 crossover probability, and 0.2 mutation probability OvA for multiclass	Cleveland multiclass	Selecting 6 features out of 13.

GA ensemble [9]	Select optimal set of features	Fitness function as classification rate using Ensemble	Parkinson's disease	from 22 features to 7 by AdaBoost with 96.55% accuracy
				from 22 features to 10 by bagging with 98.28% accuracy

GA SVM [10]	Feature set selection	Fitness function is based on classification accuracy and no. of selected genes Roulette wheel selection	Cleveland heart disease	Accuracy enhanced by 4.64% onto dataset with 7 features selected out of 13
GA LoR [11]	Analyze data with high feature interaction and handle large-scale features and instances	GA wrapper with embedded regularization	Lung Cancer dataset	Train and test accuracy of 98.83% and 93.61%, respectively
GA SVM [12]	Evolutionary and clustering algorithm to enhance accuracy on reduced set	Kmeans has been used to remove noisy data and null values replaced by mean	Pima Indian Diabetes dataset	Enhanced accuracy by 2% and identified 5 critical attributes out of 8

(continued)

Table 1 (continued)

Algorithm	Objective	Methodology	Dataset	Results
GA SVM [13]	Optimize feature subset	Asymptotic behavior of SVM fused with GA	UCI datasets	Highest classification accuracy, least number of selected features and least average processing time

have utilized the global optimization advantage of Genetic Algorithm for initialization of NN weights with faster, stable, and accurate learning than backpropagation (BP). The initialization of NN weights in BP algorithm is a blind process, and there is slow convergence. LM optimization is the fastest BP algorithm though requires more memory. Testing was done on 20 chromosomes in 100 generations with 12, 10, and 2 neurons in input, hidden, and output layers, respectively. The fitness function is based on MSE.

Jabbar et al. [16] identified the value of hyperparameter k in kNN with crossover prob = 0.6 and mutation prob = 0.033 for 20 chromosomes in 20 generations each resulted in enhanced accuracy for several diseased datasets. Pima Indian Diabetes dataset lies under normal distribution; hence, feature selection utilizing GA SVM enhanced accuracy and identified critical attributes [17]. GA is used to select optimal set of features from 22 features of Parkinson's disease to 7 by AdaBoost and 10 by bagging ensemble utilizing fitness function as classification rate with 96.55% and 98.28% accuracy, respectively [9].

GA wrapper with embedded regularization achieved train and test accuracy of 98.83% and 93.61%, respectively, on lung cancer dataset [12]. Increase in the performance of NN by 10% through enhancing initial weights utilizing GA was obtained [18]. Optimal feature subset with min. MSE and max. R^2 chosen with 22, 5, and 1 neuron in input, hidden, and output layers, respectively. Future work states optimizing learning rate and momentum factor. Gokulnath and Shantharajah [10] utilized Roulette wheel selection and SVM for classification with single-point crossover followed by mutation to preserve genetic diversity and elitism. The data is preprocessed based on:

$$Z - \text{score normalisation}\, z = \frac{(x - \mu)}{\sigma} \text{ where } \mu = \text{mean and } \sigma = \text{s.d.} \tag{1}$$

The fitness function f is based on classification accuracy f_1 and no. of selected genes f_2 as shown in Eqs. 4–6 below. Comparison with other feature selection algorithms improved the classification accuracy by 4.64% onto Cleveland heart disease dataset.

Similarly, GA is able to optimize number of base classifiers in an ensemble with significantly lower classification error and produce a model that is robust to outliers [13, 19]. GA is also used to impute missing values when there is high rate of missing data with info gain as fitness function [20]. Table 2 shows literature review of several hyper parameter optimisation techniques applicable to several critical care disease datasets inorder to enhance performance.

Table 2 Review of the literature on parameter optimization

Algorithm	Objective	Methodology	Dataset	Results
GA NN [18]	To experiment highly accurate hybrid method for disease diagnosis	Enhancing initial weights utilizing GA. by min. MSE and max. R^2 chosen with 22, 5, and 1 neuron in input, hidden, and output layers, respectively	*Z*-Alizadehsani dataset	Increase in performance of NN by 10%
GA NN [7]	Optimal number of hidden nodes	Levenberg–Marquardt (LM) backpropagation algorithm	Cancer dataset	Enhanced CV accuracy
GA NN [8]	Finding optimal value of number of connections, i.e., hidden node size	Resilient BP, LM, gradient descent with momentum Fitness function is inverse of error	Breast cancer dataset	99.43% best and 98.29% average correct classification
GA NN [15]	Initialization of NN weights	LM optimization with 20 chromosomes in 100 generations and 12, 10, and 2 neurons in input, hidden, and output layers, respectively. The fitness function is based on MSE	Survey by the American Heart Association	7% more validation accuracy than train
SVM GA [5]	Selecting important and relevant features and discarding irrelevant and redundant ones	Integer coded GA Fitness value chosen is the classification accuracy with 25 generation, 50 population, 0.8 crossover, and 0.2 mutation probability	Cleveland multiclass	The maximum accuracy is obtained using ‘one against one’ multiclass SVM with RBF kernel having width 0.025 and penalty factor 150
SVM GA [4]	Less complex SVM model with fewer support vectors	Best penalty and kernel function parameter	Diabetic dataset	A less complex SVM model with fewer support vectors

(continued)

Table 2 (continued)

Algorithm	Objective	Methodology	Dataset	Results
GA ensemble [19]	Model robust to outliers	Optimize number of base classifiers	Statlog, Pima Indian Diabetes	5.86 av. out of 900 in an ensemble with significantly lower classification error Chooses diverse base learners
SVM GA [3]	Optimize parameters	Fitness function chosen is classification accuracy, number of selected features and feature cost	Statlog, Pima Indian Diabetes, Breast Cancer	Achieve higher av. AUC w.r.t. Grid search
GA kNN [16]	Global search in complex and large multimodal landscapes	kNN with GA	Hypothyroid data from corporate hospitals in A.P	Identified the best value of hyperparameter $k = 1$ in kNN with crossover prob = 0.6 and mutation prob = 0.033 for 20 chromosomes in 20 generation each resulted in enhanced accuracy for several diseased datasets
GA SVM [13]	Feature subset selection	Asymptotic behavior of SVM fused with GA	UCI datasets	The objective achieved was the highest classification accuracy, least number of selected features and least average processing time on

3 Proposed Architecture

Current work involves utilizing GA for parameter optimization of SVM and NN. Genetic algorithm is based on randomness. The methodology involved is as follows. The same is reflected in the flowchart in Fig. 6.

1. Set Number of iterations/generations (M), Number of individuals/chromosomes in initial population (N), crossover probability (p_c), mutation probability (p_m)
2. Create N population randomly of string size l
3. $m = 1, n = 1$
4. Select 2 parents through selection
5. Crossover the parents to get children at p_c
6. Mutate children at p_m
7. Calculate fitness of mutated children, save fitness value
8. Repeat 4–7 $n/2$ times (increment n)
9. Get a new generation of mutated children
10. Increment m
11. Repeat 4–10 M times
12. Chose the best fitness value from the last generation.

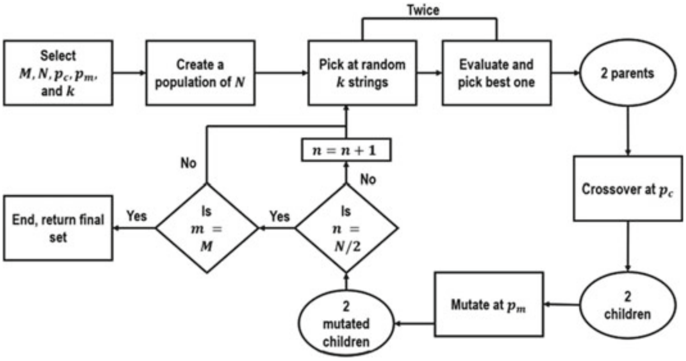


Fig. 6 Flowchart for applying GA

4 Results and Discussion

4.1 Dataset

According to yourtotalhealth website, WHO estimates 11.1 million deaths from cardiovascular artery disease (CAD) in 2020. Further, there are several datasets available on Kaggle.com on heart diseases. Cleveland and Stalog have less number of instances, whereas Hungarian and Switzerland have large missing values. Framingham Heart Disease dataset predicting ten year coronary heart disease (CHD) with 14,428 instances and 15 features are preferred for experimentation due to availability of more data and less percentage of missing values. Any disease dataset has been found to be imbalanced since number of instances in healthy class will be less compared to diseased class.

Feature engineering showed categorical features as gender, education level, currently smoking, taking BP medication and whether stroke, hypertension, diabetes are prevalent. These features were transformed by creating dummy features and then deleting redundant ones. Framingham heart disease dataset was containing unequal binary class distribution with 3596 and 644 instances, respectively, for patients suffering from TenYearCHD and healthy patients. Thus, dataset was balanced through synthetic minority oversampling technique (SMOTE) which enhanced all parameters, especially sensitivity results which are of more importance in medical diagnosis. Exploratory data analysis is done to reveal missing value in features BPMeds, education, glucose, and totchol. Since the missing percentage was less, kNN imputation with Manhattan distance for continuous and hamming for categorical was used for $k = 30$. EDA also demonstrates features in differing scale. Standard scaler is applied so that features in greater ranges do not overpower those in smaller ranges. Train:Test ratio kept is 60:40 as compared to 75:25. Table 3 reveals good sensitivity and F_1 score for NN and SVM classifiers.

Table 3 Comparison of base classifiers

Classifier	Accuracy (CV) (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F_1 score (%)
Decision tree	85	82	30	87	84
K nearest neighbor	85	71	49	88	79
MLP neural network	82	**84**	28	87	**85**
Support vector machine	83	**82**	32	87	**84**

$$\text{Accuracy} = (\text{TP} + \text{TN})/(P + N) \tag{2}$$

$$\text{Sensitivity or Recall} = \text{TP}/P \tag{3}$$

$$\text{Specificity} = \text{TN}/N \tag{4}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{5}$$

$$F_1\text{score} = 2 * \text{Precision} * \text{Recall}/(\text{Precision} + \text{Recall}) \tag{6}$$

where TP = True Positive, TN = True negative, FP = False Positive, FN = False Negative

$$P = \text{TP} + \text{FN and } N = \text{FP} + \text{TN}.$$

4.2 GA for Hyperparameter Tuning SVM

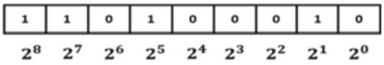
The experimentation is performed for initial population of 100 chromosomes iterated over 50 generations for convergence. The validation set is chosen of three folds. The parameters for GA are $p_c = 1$, $p_m = 0.2$, pop = 50, gen = 50, $k$fold = 3. A stack of random chromosomes is created each time a new set of solution belongs to the population. Precision is used to compute feature vector for binary chromosome. Then chromosome is decoded with respect to the range in which C and ɣ are allowable as shown in Eqs. 7–8 below:

$$\text{Precision} = (b - a)/(2^l - 1) \text{ where range}(a, b),\ \text{chromosome length} = l \tag{7}$$

$$\text{Decoding} = \Sigma(\text{bit} * 2^i) * \text{precision} + a \tag{8}$$

The initial chromosome chosen is binary comprising of 1st 15 bits for ɣ and last 15 for C regularization hyperparameter of SVM RBF kernel, respectively. The sample is shown in Fig. 7. Upper and Lower bound default ranges are $b = 1000$, $a = 10$ for C and $b = 0.99$, $a = 0.05$ for ɣ. This lower and upper bounds have been provided through knowledge from the published literature.

Fig. 7 Sample chromosome of 9 bits



Decoding the chromosome/genotype to a value of C or γ is called phenotype. Three contestants are chosen for tournament selection as per the knowledge gained from online sources. The basis is MSE with SVM classifier. Again this is the default number as stated in courses undergone. Diversity is created through unique random numbers.

We find that the accuracy for SVM w/o GA is **83%** and with is **84%**. Hyperparameter tuning with GA retrieved hyperparameter values as $C = 903.83$ and $\gamma = 0.055$. Default kernel parameters are $C = 1$ and $\gamma = 0.1$. Similar experimentation utilizing grid search gave 80% accuracy with SVM linear kernel and $C = 1$.

4.3 GA for Hyperparameter Tuning NN

Similar experimentation was performed with multilayer perceptron (MLP) classifier to optimize weights used in training NN. The data has been shuffled for generalization. The number of neurons range 6–10 and number of hidden layers 3–8 based on similar experimentations done by other researchers. The chromosome is now 32 bits with 1st 2 being combinatorial/discrete for number of hidden layer and number of neuron and next 30 being continuous for momentum γ and learning rate η, respectively. A population of 50 chromosomes is generated for 50 generations. NN improves training speed and accuracy. Training finds appropriate value for weights and bias, utilizing BPNN keeping track of mutated child per generation and updating the weights computing increment by Eqs. 9.

$$\text{Weight increment } \Delta w_{ij} = \left(\eta * \frac{\partial E}{\partial \omega_{i_j}}\right) \tag{9}$$

$$\Delta w_{ij} = \left(\eta * \frac{\partial E}{\partial w_{ij}}\right) + \left(\gamma * \Delta w_{ij}^{t-1}\right) \tag{10}$$

where η is the learning rate, γ is the momentum factor, Δw_{ij}^{t-1} is weight increment in the previous iteration, and $\frac{\partial E}{\partial \omega_{i_j}}$ is weight gradient.

The solver used is adam which gives better results than S.G.D. in terms of stability and speed. Upper and lower bound default ranges are $a = 0.01$, $b = 0.3$ for momentum and $a = 0.01$ and $b = 0.99$ for learning rate based on the literature reviewed. Two-point crossover has been chosen. The optimal value of hyperparameters obtained by GA is # of neurons = 6, # of hidden layers = 4, momentum = 0.086, learning rate = 0.060. Training accuracy achieved is **84%** with GA as compared to 82% w/o GA. Similar experimentation utilizing Grid Search gave 80% accuracy with MLP NN utilizing activation = tanh, alpha = 0.05, hidden_layer_sizes = (50, 50, 50), learning_rate being constant, and solver as S.G.D.

Table 4 Comparing performance with all and reduced features on SVM classifier

Classifier	Sensitivity (%)	F_1 score (%)
With all features	82	84
With reduced features	89	88

Table 5 Comparing performance with all and reduced features on MLP NN classifier

Classifier	Sensitivity (%)	F_1 score (%)
With all features	84	85
With reduced features	100	92

4.4 GA for Feature Selection

The feature selection task when performed through NN with binary chromosome of size = # of features, i.e., 15 reduces number of features to 9 selecting male, age, education, currentSmoker, BPMeds, sysBP, diaBP, heartrate, glucose as high risk factors, whereas feature selection task when performed through SVM returns 12 features, i.e., male, age, education, cigsperday, BPMeds, prevStroke, diab, sysBP, diaBP, BMI, heartrate, glucose as high risk factors. Initial level of Feature Engineering through Pandas profiling showed correlation in features SysBP and DiaBP with hypertension. Also, correlation in CurrentSmoke and CigPerDay features was found through Spearman and Pearson correlation. Results of feature selection through GA verify statistical findings. The experimentation is done for population of 80 chromosomes for 30 generations. Fivefold CV is performed for generalization based on courses undergone. Crossover is always performed whereas mutation is performed 20% of the time.

Results demonstrated in Tables 4 and 5 above show enhancement in performance w.r.t. sensitivity and $f1$ score when number of features are reduced utilizing GA with both SVM and NN.

5 Conclusion and Future Scope

Results claim the power of GA for parameter optimization of ML models and optimal feature selection utilizing MLP NN and SVM classifiers. The results demonstrate enhancement in accuracy when utilizing GA for hyperparameter tuning SVM and NN. Similarly, GA is able to identify informative features and thus reduce the dimensionality, thereby enhancing sensitivity and $f1$ score performance metric. Results of feature selection through GA are validated with results obtained through statistical analysis. Performance is enhanced further using feature engineering and exploratory data analysis in the form of removing data imbalance, imputing missing values, removing correlated features, scaling numerical features in differing scale and treating categorical data. Research findings claim that though grid search is used for hyperparameter tuning but cannot be used for feature selection and thus cannot fulfill the multi optimization objective.

Future work may test the power of GA for finding best set of classifiers in an ensemble, find hyperparameter values for ensemble learners, and impute large missing data as well. Current work can be extended utilizing NearMiss undersampling technique for balancing dataset and Multiple Imputation with Chained Equations (MICE) for imputing missing values where multiple columns have missing data as in the case of Hungarian and Switzerland heart disease datasets. Currently, the work is dealt with the same number of neuron per layer for MLPNN; hence, efforts shall be put in this direction in the future work.

References

1. Gupta S, Sedamkar RR (2019) Apply machine learning for healthcare to enhance performance and identify informative features. In: IEEE INDIACom; 6th international conference on computing for sustainable global development. BVICAM, New Delhi (INDIA), 13th–15th Mar 2019
2. Gupta S, Sedamkar RR (2019) Feature Selection to reduce dimensionality of heart disease dataset without compromising accuracy. In : 9th international journal of innovations in engineering and technology (IJIET)
3. Huang CL, Wang CJ (2006) A GA-based feature selection and parameters optimization for support vector machines. Elsevier Expert Syst Appl 31:231–240
4. Fei Y, Min H (2016) Simultaneous feature with support vector selection and parameters optimization using GA-Based SVM solve the binary classification. In: 2016 First IEEE international conference on computer communication and the internet
5. Bhatia S, Prakash P, Pillai GN (2008) SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features. WCECS, 22–24 Oct 2008, San Francisco, USA
6. Plawiak P (2017) Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals. In: Elsevier ScienceDirect swarm and evolutionary computation base data 2017
7. Ahmad F, Mat-Isa NA, Hussain Z, Boudville R, Osman MK (2010) Genetic algorithm—artificial neural network (GA-ANN) hybrid intelligence for cancer diagnosis. In: 2010 second international conference on computational intelligence, communication systems and networks IEEE Computer Society.
8. Ahmad F, Mat Isa NA, Halim M, Noor M, Hussain Z (2013) Intelligent breast cancer diagnosis using hybrid GA-ANN. In: 2013 fifth international conference on computational intelligence, communication systems and networks IEEE Computer Society.
9. Fayyazifar N, Samadiani N Parkinson's disease detection using ensemble techniques and genetic algorithm. IEEE Artif Intell Signal Process
10. Gokulnath CB, Shantharajah SP (2018) An optimized feature selection based on genetic approach and support vector machine for heart disease. Springer Science+Business Media, LLC, part of Springer Nature, 16 Mar 2018
11. Lacerda EGM, Carvalho ACPLF, Ludermir TB A study of cross validation and bootstrap as objective functions for genetic algorithms. In: IEEE proceedings of the VII Brazilian symposium on neural networks (SBRN'02).
12. Liu XY, Liang Y, Wang S, Yang ZY, Ye HS (2018) A hybrid genetic algorithm with wrapper-embedded approaches for feature selection. IEEE Access 2018

13. Oh DY, Gray JB (2013) GA-ensemble: a genetic algorithm for robust ensembles. Springer, Mar 2013
14. Zhao M, Fu C, Ji L, Tang K, Zhou M (2011) Feature selection and parameter optimization for support vector machines: a new approach based on genetic algorithm with feature chromosomes. Elsevier Expert Syst Appl 38:5197–5204
15. Amin SU, Agarwal K, Beg R (2013) Genetic neural network based data mining in prediction of heart disease using risk factors. In: IEEE conference on information and communication technologies (ICT 2013), pp 1227–1331
16. Jabbar MA, Deekshatulua BL, Chandra P (2013) Classification of heart disease using K-nearest neighbor and genetic algorithm. Elsevier ScienceDirect Procedia Technol 10:85–94
17. Santhanam T, Padmavathi MS (2015) Application of K-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. Elsevier ScienceDirect Procedia Comput Sci 47:76–83
18. Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA (2017) Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm, Elsevier ScienceDirect. Comput Methods Programs Biomed 141:19–26
19. Fletcher S, Verma B, Jan ZM, Zhang M (2018) The optimized selection of base-classifiers for ensemble classification using a multi-objective genetic algorithm. In: 2018 IEEE international joint conference on neural networks (IJCNN)
20. Shahzad W, Rehman Q, Ahmed E (2017) Missing data imputation using genetic algorithm for supervised learning. Int J Adv Comput Sci Appl 8(3)

Secured Crowdfunding Platform Using Blockchain

Megha Sahu, Ashish Gangaramani, and Asha Bharambe

Abstract Crowdfunding is a platform can be used to collect small amount from large number of people. In Traditional platform it is not easy to track the usage of the fund. Hence campaign creator can use money for their own need.This paper proposes a solution on how to prevent such fraud in crowdfunding platforms using blockchain and smart contracts. The main aim of this solution is to propose a solution that can reduce those effects. The important feature of Blockchain is that it maintains transparency among the nodes in the network. We are proposing a solution keeping this feature in mind to implement campaign as smart contracts designed for crowdfunding websites where campaign managers will need to get approval based for their requirements from backers. The proposed solution has been implemented using Ethereum and tested on Rinkeby Network.

Keywords Smart contract · Backer · Campaign · Campaign creator/manager · Rinkeby network · Metamask

1 Introduction

Crowdfunding is a method of collecting capital through the effort of friends, family, customers, and individual investors. Nowadays social media is our biggest tool of communication, crowdfunding and internet together can be great solution when you need some investment. When number of people sees potential in some idea or project, they can invest small amount and you can easily collect the target amount.

M. Sahu · A. Gangaramani (✉) · A. Bharambe
Vivekanand Education Society's Institute of Technology, Mumbai University, Mumbai, India
e-mail: ashishgangaramani@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_3

1.1 Benefits of Crowdfunding

You can easily reach to all kind of investors with more flexible investment options, there are many benefits of crowdfunding over traditional methods. Here we are presenting some.

Reach—Nowadays Internet is the easiest way to connect to people. Using crowdfunding systems you can easily reach to all kind of investors and ask them to share your campaign. Being online in there is a wide and global spectrum of investors who can take interest in created campaign.

Presentation—Crowdfunding requires you to provide all the essential details about your idea and your plan about when will you be ready with results, promised product and potential so that investor can get proper idea about the product that they are investing in.

PR and Marketing—Crowdfunding provides you a platform where you can easily present your idea, traditionally you visit number of investors before you find the one investor. You can easily share, promote your idea on social sites. After creation of a campaign, campaign can be advertised by sharing it on various online, as well as offline platforms, increasing chances of investment in business.

Validation of Concept—Such Platform gives an opportunity to present your concept to experts and get your concept verified, check if there is something missing or not feasible. Because many investors are taking look at a campaign, they may ask many important questions that can give an idea about proposed business and where it can stand in today's market.

Productivity—Instead of going to each and every investor campaign are filtered for particular type of campaign saving time of both investors and startups.

1.2 Problem Statement

Carpenter [1] have stated that current crowdsourcing platforms are at high risk by malicious campaign creators who may raise money in the name of a project and inappropriately misuse funds for their private use reason for the same, this is due to lack of transparency between investors and campaign managers. Our main aim is to create a platform which maintains transparency on how campaign manager will use the raised funds. Blockchain will help us to fore come the problem of transparency between investors and campaign manager.

Under the currently using Systems the amount that initiator gets is mostly not the actual amount collected. The reason is the middleman who takes the commission of promotions. Blockchain automates this process for campaign creator, no middleman is required hence no additional fees is required and also this is more secured as blockchain maintains the transparency between each node on the network. Campaign creator can manage their campaign on their own.

Cornell and Luzar [2] have mentioned the possible frauds on crowdfunding platform "Backer Fraud" when backer submits campaign to claim the refund later. "Broker/Portal Fraud" when campaign creators themselves involved in fraud "Backer Creator Fraud" when campaign manager invests in their own campaign.

1.3 Objective

- Our main objective is to create a web based solution for crowdfunding using blockchain. So that it can be easily accessed by the end user.
- Identify the possible frauds, vulnerabilities of currently implemented crowdfunding solutions.
- Provide more secure solutions to the user by preventing the above specified frauds to occur.

1.4 Scope

Although there are many problems and frauds related to crowdfunding But our scope is limited to few major section of any crowdfunded project.

- **Pre-empted fraud**: Before campaign is funded,it is shut down by platform itself and this action is generally initiated by users of that site.
- **Plagiarised-Fraud**: It generally takes place when campaign is using IP or assets that do not belong to them and is plagiarised, it may be reported by different organization processing these resources
- **Discern Fraud**: Main causes of these frauds are when contributors of site are not received or delayed with rewards or early access to products.

2 Review of Literature

2.1 Papers/Findings

The paper titled "Exploring Blockchain for Alternative Finance" stated that [3]: This paper was based on the variety of research on crowdfunding and its impact on society and economy. They created a work group to explore blockchain for alternative finance. Their main aim was to enable different target sectors to understand, deploy and support blockchain. The basis of their research were qualitative interviews with market and technology experts and SMEs, who were experienced in blockchain based products/services. Based on their outcomes, they complemented the practical perspectives from a research, legal and technological point of view.

They did research keeping in mind—technical aspect, partiality, legal aspect. They stated that blockchain can be a solution to various security issues like confidentiality, fraud and trust.

The paper titled Funding Community Projects with Smart Contracts on Blockchain [4] **stated that**: The main aim of this paper was to demonstrate free riding problem in civic crowdfunding i.e. people can still take advantage of projects without investing in it. This could be a reason for no investment in the campaign. Hence they suggested a way to implement incentive mechanism with cryptocurrency like Ether. They also stated that as blockchain is more reliable as it is distributed the interaction between the agent and the smart contracts so it guarantees to record all the transaction in the blockchain without any modification. Each node in blockchain is identified by public key, hence it is more prone to sybil attacks. This technology is new so they still need to deal with floating points, buffer over/under flows, re-entrant code etc.

The paper titled "Equity Crowdfunding Based on the Blockchain?—A Delphi Study [5]**" stated that**: The purpose of the study was to conclude that equity crowdfunding using blockchain has potential to close the gap faced by start-ups. They divided this study into two parts first expert based delphi survey and ten market driving forces and checking how they influence the equity crowdfunding if implemented with blockchain. They have used delphi technique which is used in building theory, in complex and interdisciplinary areas.

Figure 1 illustrates classic delphi techniques, which starts with operationalization of research question based on literature review and expert opinions that can be evaluated by further experts. Then, a survey is designed and executed with a pilot round beforehand. After conducting the first round, anonymous responses are processed and statistically aggregated for the next round. This controlled feedback allows experts to reassess their initial judgements. The whole process is repeated until a previously determined stopping criterion is reached. The result of study showed that out of 10 driving forces, 6 had positive influence on equity crowdfunding and hence experts agreed that blockchain has positive influence on equity crowdfunding.

The paper titled "Analysis and outlook of applications of blockchain technology to equity crowdfunding in China [6]**" stated that**: This paper was specifically for crowdfunding in china. They reviewed different papers and already implemented applications to conclude if blockchain can be used in equity crowdfunding or not. They wanted to allow anyone with a good start up idea to create campaign. They stated that implementing the solution with blockchain will provide low barrier entry

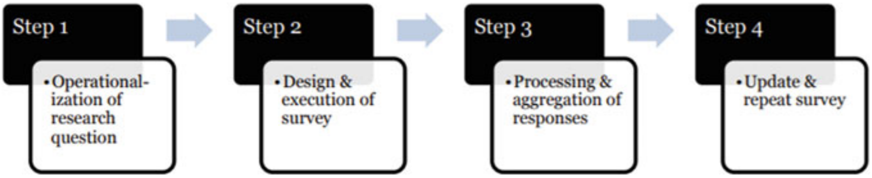

Step 1
• Operationalization of research question
Step 2
• Design & execution of survey
Step 3
• Processing & aggregation of responses
Step 4
• Update & repeat survey

Fig. 1 Delphi technique

points, low cost and high speed transaction as compared to traditional application. Their main aim was to maintain transparency and fight money laundering. As the reward to the investor, they offer equity or bond like shares. They have also mentioned some problem that is faced in implementing equity crowdfunding "Registration of shares—the location of investors can be drastically different", "Difficulty in confirmation of shareholder's rights". "Registration cost management—different location different cost". Blockchain for equity is still on paper, there are many technical and legal issues to be resolved.

Whirl White paper "Inspiring a Whirlwind of Good [7]**" stated that**: This is a whitepaper on an application called WHIRL "Pay-it-forward" which an crowdfunding project based on blockchain. The most interesting aspect in their application is that whenever someone creates a campaign, they will get karma points depending on the amount you contribute. A person can create campaign only when they have some karma points i.e. they have contributed to some campaign. As there can be lots of campaign this application only focuses on limited campaign at a time combined with karma points. And their unique point is that they use different dozen of cryptocurrencies hence this project can be used globally.

They are tackling two problems first "Many campaign closes unfunded" and secondly backers loses interest when they don't get their rewards. To solve first they have limited campaign at a time and only people long listed on different cryptocurrencies can back a project. And they give karma points in reward.

The paper titled "The Application of Blockchain Technology in Crowdfunding: Towards Financial Inclusion via Technology" [8] **stated that**: This paper is on crowdfunding in Malaysia, which is the first country in south Asia to implement it.They stated that "Crowdfunding is a practice of funding a project or venture by raising small amounts of money from a large number of people via the internet". According to paper crowdfunding can be a different financial source for start-ups, entrepreneur. Blockchain can bring crowdfunding to the next level by not just providing data security but also efficiency, affordability and more profitability. Building awareness is most important in crowdfunding. Creating road shows, conferences are necessary in empowering underprivileged groups. They tried to create engagement with private sectors and also created shariah—complaint crowdfunding to encourage crowdfunding in Muslim countries.

3 Proposed Solution

Weber and Friedrich [3] and Chandra et al. [5] have concluded that the blockchain has good influence on crowdfunding and can be beneficial for society and economy of the country if implemented properly. To attract more user to take initiative and invest in project, can incentivize the investors. This is a great aspect in crowdfunding. "Analysis and outlook of applications of blockchain technology to equity crowdfunding in China" [4] and "Inspiring a Whirlwind of Good" [7] have implemented different ways of providing incentives to the investor e.g. giving Karma points to investor.

But we found that giving ether as incentives which can be later on converted into their native currency, will be more beneficial for investor than giving karma points. Muneeza et al. [8] have stated how Malaysia has implemented crowdfunding using blockchain, how creating awareness about crowdfunding is important and that needs support of financial banks also.

An important aspect in crowdfunding is security and management of the money funded by the investors. In traditional system once investor transfers the money to campaign creator, they can use that money however they want. To overcome this we are proposing to use Blockchain to maintain transparency between investor and campaign creator. The investors will have all the right to decide if the campaign creator can withdraw the money that have invested or not after threshold reached.

We are using ethereum as it is public blockchain network that allows anyone to get connected to the network. No node in ethereum will have special rights like in other type of network. Ethereum uses proof of Stake. While PoW depends on computing power to mine blocks, PoS depends on a node's stake (typically the amount of currency a user holds) in the system. The more stake a user controls, the more authority they have over validation. Hence this is more suitable for crowdfunding as it gives more priority to higher investors.

3.1 Methodology

Figure 2 illustrates that the manager of the campaign will create a spending request on an Ethereum smart contract and contributors the investor, will approve the spending request.

Further carrying forward the spending request in the Fig. 3 green colour represents approval of contributors as YES and red colour as NO. If the result is more than 50% then will allow campaign creator to send request to the vendor. The ether will be transmitted to the verified vendor and vendor will provide the requirements to campaign creators. As we are directly transmitting ethers to verified vendor and not to the campaign creator. They cannot fake their requirements and use contributed amount for their personal use. Contributors in return will get some credits or offers as mentioned by campaign creator depending on their contribution in the campaign.

3.2 End User Diagram

Figure 4 illustrates all the users of the application i.e. campaign manager, investors and vendor. These have to be connected to the blockchain, to get connected they can register using application. The campaign manager can also be investor for some campaign and vice versa.

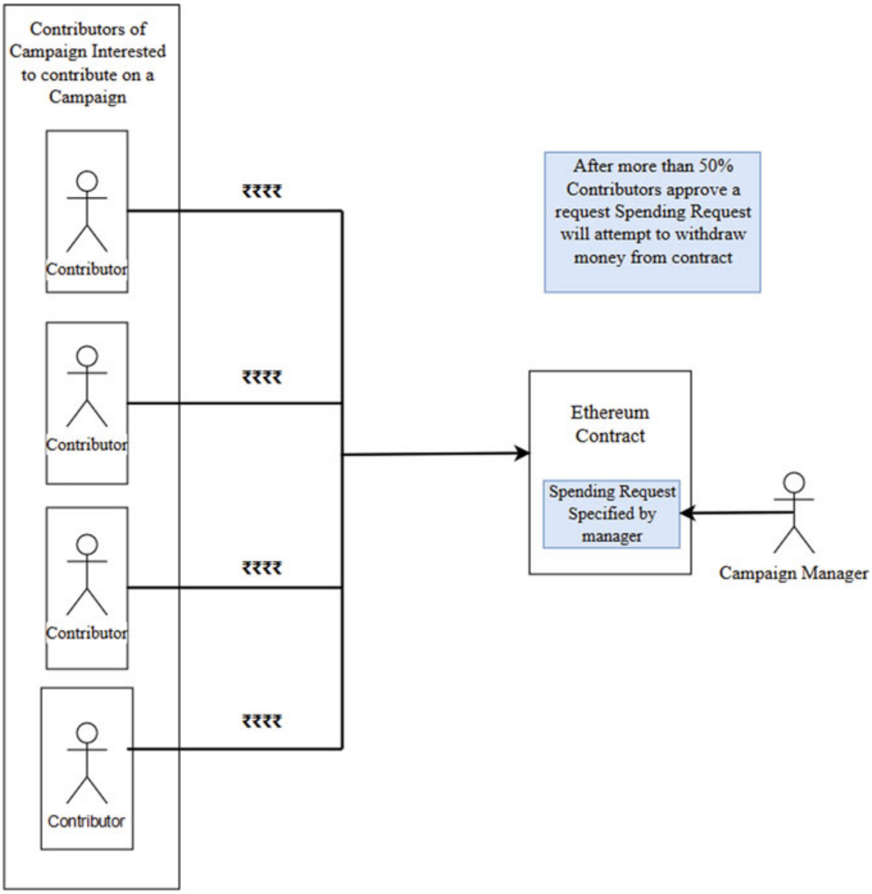

Contributors of Campaign Interested to contribute on a Campaign
Contributor
Contributor
Contributor
Contributor
₹₹₹₹
₹₹₹₹
₹₹₹₹
₹₹₹₹
After more than 50% Contributors approve a request Spending Request will attempt to withdraw money from contract
Ethereum Contract
Spending Request Specified by manager
Campaign Manager

Fig. 2 The solution architecture

3.3 *Use Case Diagram*

Figure 5 shows the use case diagram of the application which illustrates the involvement of different end user in activities. Example for a campaign manager to add the request there must be some investors.

4 Implementation and Results

A web based application is developed for our solution and to handle blockchain side of our application we have used solidity programming language, for end users to

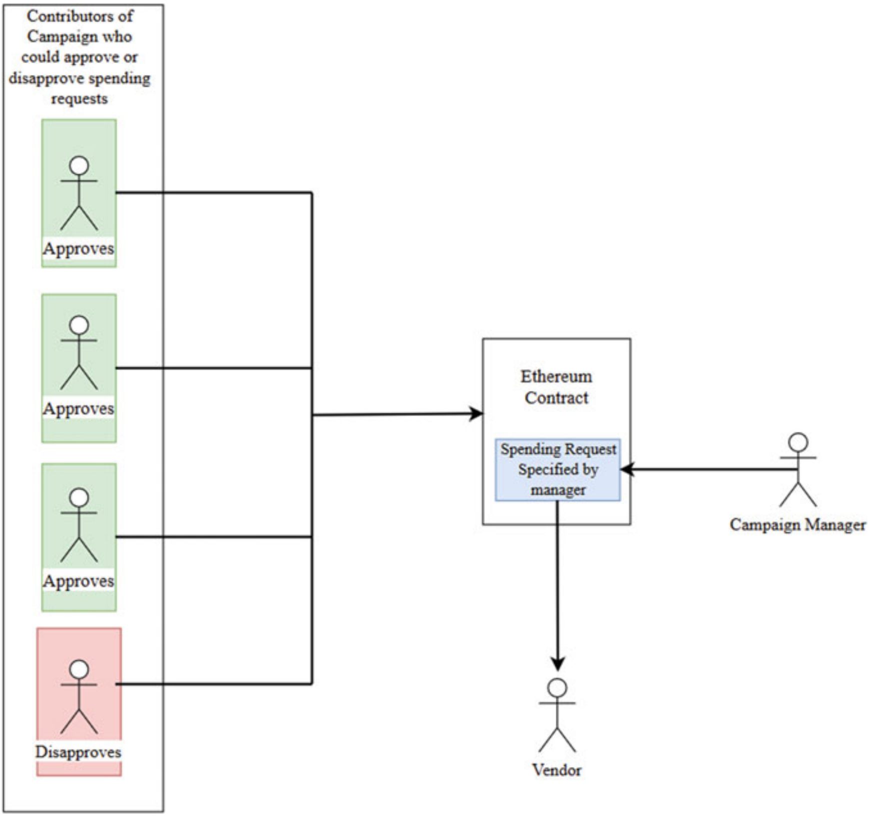

Contributors of Campaign who could approve or disapprove spending requests
Approves
Approves
Approves
Disapproves
Ethereum Contract
Spending Request Specified by manager
Campaign Manager
Vendor

Fig. 3 An approval module

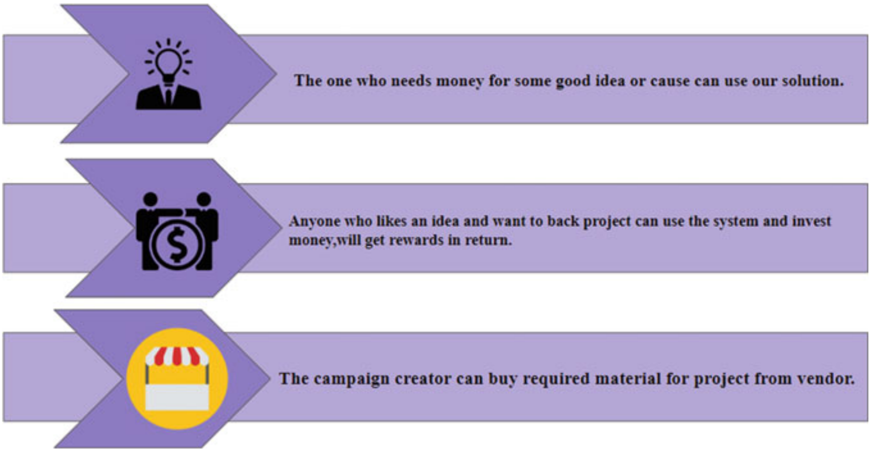

The one who needs money for some good idea or cause can use our solution.
Anyone who likes an idea and want to back project can use the system and invest money,will get rewards in return.
The campaign creator can buy required material for project from vendor.

Fig. 4 Users of the system

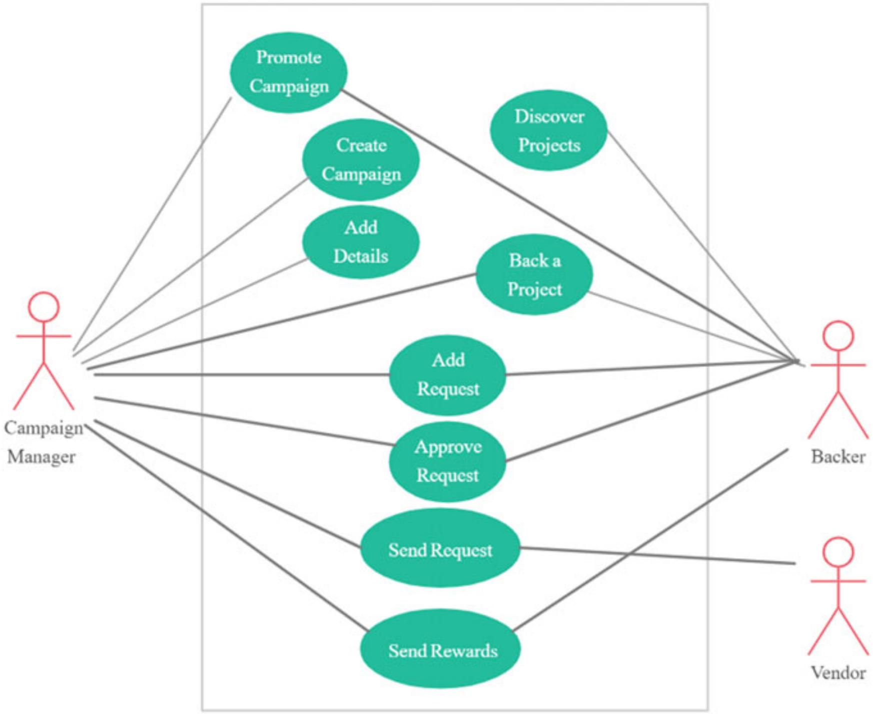


Fig. 5 Use case diagram

interact with our smart contract we have used React JS as our front end and to provide an interfacing between front end and block chain web.js is used as a translation layer.

4.1 Rinkeby

Rinkeby provides us with an Ethereum testnet. It is used by developers to test their written smart contracts. Currency on Rinkeby testnet is valueless. Rinkeby uses Proof of Work. We have used Rinkeby test network to make sure that all transactions are working in perfect manner when they will be deployed to an actual ethereum network, at that time there should not be a risk involved of users losing their cryptocurrency which possess an actual value.

4.2 MetaMask

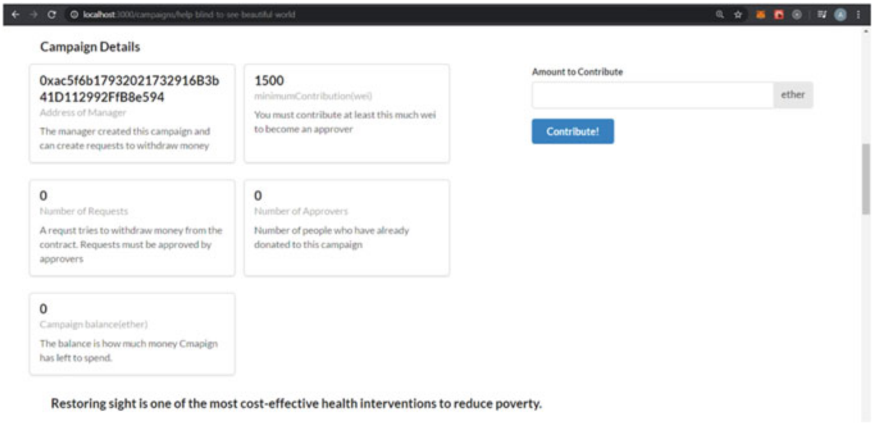
MetaMask is a browser extension which enables us to run decentralized applications (dApps) without actually registering on that network as an Ethereum node. It helps to connects to another ethereum node called Infura and runs smart contracts on that node. MetaMask also manages Ethereum wallet, which has Ethers of a user, and allow us to send to and receive from a decentralized application.

4.3 Pages of application

These are some prototype designs of our crowdfunding solution on blockchain, it contains a series of webpages that a normal end user will use to interact with an instance of a contract deployed on rinkeby test network.

4.3.1 Campaign Display Page

The new campaign has been added to the application and Fig. 6 displays that campaign, on this page users can see what the campaign is about and various details about campaigns like minimum contributions required, description about campaign, images of campaign, number of requests and number of approvers who have approved the campaign. Users can also contribute to that campaign on this page. To view all requests that the campaign manager has requested, the view requests button can be clicked.



localhost:3000/campaigns/help blind to see beautiful world

Campaign Details

0xac5f6b17932021732916B3b41D112992FfB8e594
Address of Manager
The manager created this campaign and can create requests to withdraw money

1500
minimumContribution(wei)
You must contribute at least this much wei to become an approver

0
Number of Requests
A requst tries to withdraw money from the contract. Requests must be approved by approvers

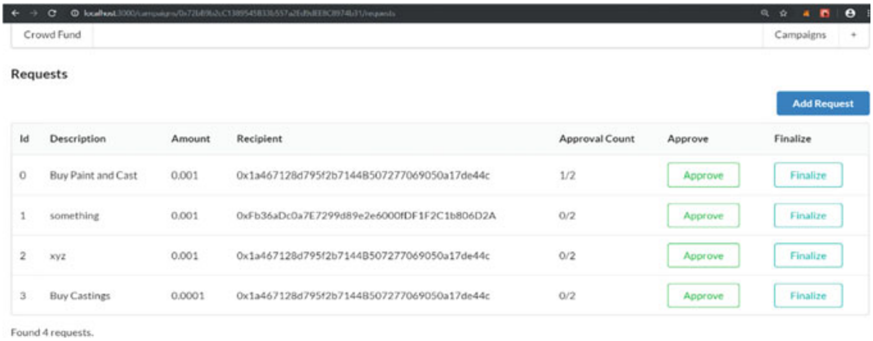
0
Number of Approvers
Number of people who have already donated to this campaign

0
Campaign balance(ether)
The balance is how much money Cmapign has left to spend.

Amount to Contribute
ether

Restoring sight is one of the most cost-effective health interventions to reduce poverty.

Fig. 6 View campaign page



Requests

Id	Description	Amount	Recipient	Approval Count	Approve	Finalize
0	Buy Paint and Cast	0.001	0x1a467128d795f2b7144B507277069050a17de44c	1/2	Approve	Finalize
1	something	0.001	0xFb36aDc0a7E7299d89e2e6000fDF1F2C1b806D2A	0/2	Approve	Finalize
2	xyz	0.001	0x1a467128d795f2b7144B507277069050a17de44c	0/2	Approve	Finalize
3	Buy Castings	0.0001	0x1a467128d795f2b7144B507277069050a17de44c	0/2	Approve	Finalize

Found 4 requests.

Fig. 7 View requests page

4.3.2 Requests Page

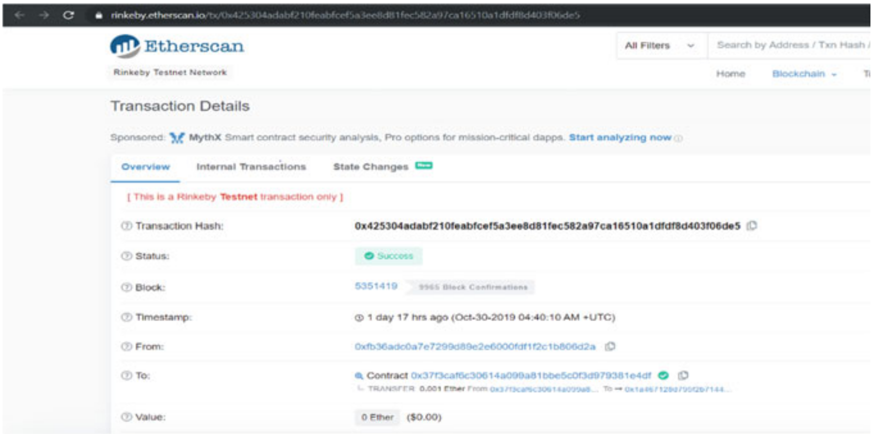
Figure 7 shows how users could view all the requests created by campaign manager, also they can approve the particular request but only if they have contributed to that campaign that request can only be finalized by campaign manager only and only if percentage of requests to campaign contributors is greater than 50%. Once creator finalizes the request ether will get transferred to the vendor’s account. Once the vendor provides requested material, can use ether or request to convert ether in their native currency.

4.4 Etherscan

Etherscan is a blockchain explorer for Ethereum and enables us to search the Ethereum blockchain for transactions, addresses, tokens, prices and other activities that take place on Ethereum or on Rinkeby test network. We are using Etherscan to confirm details of what is published on the Rinkeby test network.

Figure 8 shows transaction details on etherscan we can search details about any transaction by simply entering its hash no matter how old a transaction is we could verify that transaction on etherscan till we have its hash. Ether scan shows the following details:

- Status of a transaction whether it was successful or not.
- Block number of that transaction on Blockchain.
- Timestamp when it was mined, from whom this transaction was sent its Public address.
- To whom this transaction was intended to.
- Gas used for processing of this transaction.
- Nonce Position of the block in Blockchain.


rinkeby.etherscan.io/tx/0x425304adabf210feabfcef5a3ee8d81fec582a97ca16510a1dfdf8d403f06de5
Etherscan
All Filters
Search by Address / Txn Hash /
Home
Blockchain
Overview
Internal Transactions
State Changes
New

Rinkeby Testnet Network

Transaction Details

Sponsored: MythX Smart contract security analysis, Pro options for mission-critical dapps. Start analyzing now

[This is a Rinkeby Testnet transaction only]

Transaction Hash:	0x425304adabf210feabfcef5a3ee8d81fec582a97ca16510a1dfdf8d403f06de5
Status:	Success
Block:	5351419 9965 Block Confirmations
Timestamp:	1 day 17 hrs ago (Oct-30-2019 04:40:10 AM +UTC)
From:	0xfb36adc0a7e7299d89e2e6000fdf1f2c1b806d2a
To:	Contract 0x37f3caf6c30614a099a81bbe5c0f3d979381e4df TRANSFER 0.001 Ether From 0x37f3caf6c30614a099a8... To → 0x1a467129d795f2b7144...
Value:	0 Ether ($0.00)

Fig. 8 Details of a transaction on etherscan

All the above Transaction details ensure transparency as it's publicly available and these details further cannot be modified.

4.5 MongoDB

All of the campaign data like description and images cannot be stored on blockchain as it will increase size of a block making it expensive to mine, to counter counter this problem sensitive data like contribution amount, number of approvers, campaign balance are stored on blockchain and data like description of campaign, comments, images are stored in MongoDB which is a NoSQL database it's structured in documents unlike tables in SQL databases. MongoDB makes integration of applications easier and faster.

5 Conclusion

The Blockchain based Smart Contracts is the peer-to-peer network of thousands of distributed nodes which maintain a copy of the transactional data.In a blockchain each node/agent is identified by its public key, for testing environments Rinkeby network could be used to create demo users for a network, Ethereum and solidity programming has to be used for creation of smart contracts which will interact with our blockchain network.The application will be beneficial for society, people will get single platform to showcase their innovative ideas, people who needs money in some crisis can also use this app. Investor can trust this solution as campaign creator won't be able transfer the invested money for other causes, they can only use ether once the votes are more than 50%. Hence this methodology will help to prevent the frauds on crowdfunding platform.

References

1. https://www.digitaltrends.com/cool-tech/biggest-kickstarter-and-indiegogo-scams/. Last accessed Date 28 Nov 2019
2. https://www.crowdfundinsider.com/2014/03/34255-crowdfunding-fraud-big-threat/. Last accessed Date 28 Nov 2019
3. Weber C, Friedrich C Exploring blockchain for alternative finance. In: A publication of the ECN working group scoping paper
4. Zhu H, Zhou ZZ (2016) Analysis and outlook of applications of blockchain technology to equity crowdfunding in China. In: Zhu and Zhou financial innovation 2016
5. Chandra P, Ranjan A, Sawale J, Rajsekhar L, Singh G, Wadki H Funding community projects with smart contracts on blockchain
6. Heieck F, Fabian TEB, Lessmann S Crowdfunding based on the blockchain? A Delphi Study. Humboldt University Berlin, Chair of Information Systems
7. Whirl Inspiring a Whirlwind of Good. White paper.
8. Muneeza A, Arshad NA, Arifin AT (2018) The application of blockchain technology in crowdfunding: towards financial inclusion via technology paper. Int J Manage Appl Res 5(2)

Review of Wireless Sensor Network Security Schemes

Vaibhav D. Dabhade and A. S. Alvi

Abstract WSN is used in various applications such as hospital, environment monitoring, and experiments. It is also used in (military) [1, 2] battlefield for target tracking. If it is deployed in hostile environment in military application, then nodes can be captured and data can be updated. Wireless sensor network is equipped with low battery [3], storage, and computational power [4]. Hence, the main problem is the security of resource-constrained WSN. Various schemes are available to address the issues of WSN security. Proposed schemes are based on public and private encryption technique. Few schemes are based on biometric concept. While designing the security scheme researchers need to consider resource-constraint nature of wireless sensor network. Schemes are available but all are not applicable to WSN. This paper discusses various aspects of WSN to understand how schemes can be designed for WSN. Paper starts with application of WSN, summarizes security threats and analyzes various key management and security schemes.

Keywords Wireless sensor network security · Issues · Key management · Pairwise key generation

1 Introduction

A WSN is a group of spatially distributed autonomous devices, i.e., sensors. It can be used to monitor sound, temperature, vibration, pressure, motion, or pollutants [5]. Military application, such as battlefield surveillance [6, 7], is the main motivation for the work in WSNs. However, WSNs are now used in application areas such as industrial application, healthcare applications, environment, and habitat monitoring. In addition to these applications, nowadays, WSNs play an important role in traffic

V. D. Dabhade (✉)
Computer Engineering, MET's BKC, IOE, Nashik 422003, India
e-mail: vaibhavdabhade@rocketmail.com

A. S. Alvi
Information Technology, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_4

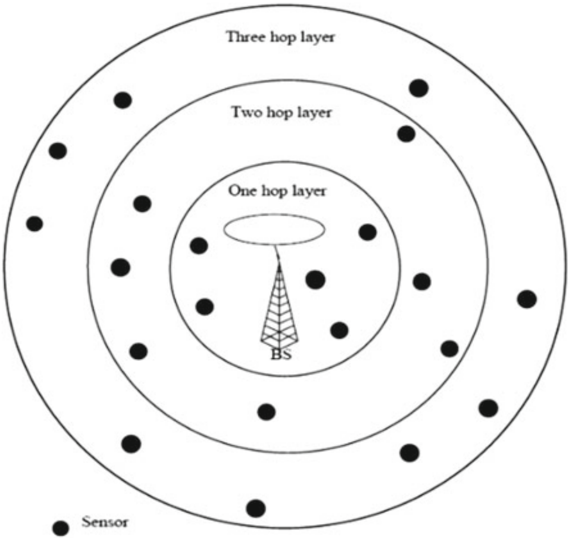
control and home applications [8]. WSNs are becoming popular nowadays because of its applications. We use various commercial applications based on WSNs. On the other hand, commercial use leads to some serious issues in WSNs, i.e., security. WSNs are equipped with low memory capacity, low processing power, and limited battery power. While designing and implementation of WSNs low development cost should be considered. Some outdoor WSN applications of WSNs, such as battlefield, traffic management, and monitoring, where network can be physically compromised, is the additional challenge for researchers. Security mechanisms which are developed and applicable for wired networks may not be useful or applicable for WSNs [6] because of various limitations. Following section covers the limitations of WSNs and security issues related to WSNs. WSNs collect information with in its range, process, and transfer it to other nodes. During this process, securing the data is very difficult. New schemes are required for WSNs as traditional schemes are not applicable [9]. Several challenges have to be addressed in WSN. First, WSNs should be economical, second, physical attack if sensor is accessible, and last is security problems [10]. Hence, traditional security schemes are not suitable, and new schemes are required for security. In next section, security issues in WSNs, schemes are discussed.

2 Related Work

For WSNs, various schemes are available, but it is very difficult to implement security schemes for power, computation, and memory constrained sensor nodes [11]. Few schemes provide high security solution but affect power consumption and memory utilization. Some WSN schemes are discussed here.

Key management scheme proposed in [12] has great security resilience and scheme enables key revocation but power consumption needs to be reviewed. The scheme proposed in [13] provides greater resilience against node capture if very few nodes are captured. One more scheme which is based on public cryptography, RSA, and ECC [14] protect gives high security but ECC is preferred over RSA. Session key establishment scheme [15] has good resilience. This scheme defends most of the attacks but communication cost is average which needs to be improved. Mixed asymmetric and symmetric cryptography [16] can be preferred over other schemes. It uses MAC to generate digital signature. As per the analysis conducted by author [16] mixed scheme does not offer resilience during initialization phase.

Fig. 1 Layered architecture



3 WSN Architecture

3.1 Layered Architecture

A layered architecture, as shown in Fig. 1, consists of a single powerful base station and nodes. It is used developing wireless backbones.

3.2 Clustered Architecture

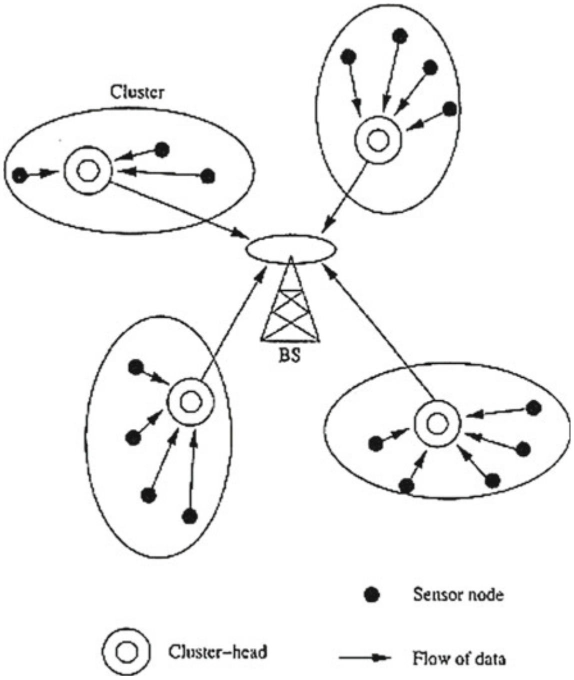
CA organizes the nodes into one or more clusters. As depicted in Fig. 2, each group.

Cluster is governed by a cluster head. Node transmits data to cluster head and then head forwards messages to a base station. This is suitable for data fusion and it is self-organizing.

While analyzing security schemes researchers need to understand WSN limitations. Limitations of WSNs are as follows.

Computational Power [8]: Symmetric and asymmetric cryptographic schemes are available for WSNs security but each schemes have their own advantages and limitation. Resilience of asymmetric schemes is good but it requires high computation and sensor nodes are equipped with very low computational power. Asymmetric cryptograpics schemes require more power than symmetric schemes. High computations power drains battery of sensor node. Hence, these issues need to be consider while proposing and implementing security mechanism for WSNs.

Fig. 2 Cluster architecture



4 Issues/Attacks

In wireless sensor networks, hackers/intruders can gain access to data collected by nodes. They can get access remotely to nodes. Once they get access to node, attackers read/modify data depending on type of attack. In passive type of attack, attacker collects data and analyzes characteristic of data flow. In active type of attacks, attackers modify data for his own benefit. Comparatively, the passive type of attacks are difficult to detect but can be protected by encrypting the contents. Encryption protects data but because of limited resources, care has to be taken while selecting and implementing encryption algorithms.

The term security is mainly comprised of authentication, confidentiality, and integrity [17]. In addition to this other,

- Authentication [18] allows a sensor node to verify/check the identity of the communicating node. It rejects messages coming from attackers or any other send if they fail to prove identity.
- Confidentiality is making information inaccessible to unauthorized user. Only authorize user can get access to data.
- Integrity checks that the data forwarded by sender is not altered. Message contents can be modified by attackers. If node compromised then data integrity is violated [14]. In addition to this, availability is also important. It allows authorize user to

access data from anywhere and anytime as per request. Below section focuses on key management, node capturing, and addition on new node in network.

- Availability: It deals with reliability in accessing resources in timely manner. For any authorized node data must be available.

To understand challenges, while proposing security schemes for WSN, various factors which may affect the performance of security scheme should be considered.

4.1 Key Management

Key management consists of (1) key generation (KG), (2) exchange (KE), (3) agreement, and (KA) (4) revocation (KR). In KG phase, new keys are generated. KE is required for sending keys among communicating nodes. In KA, node can create derived keys and finally key revocation which invalidates compromised keys.

Design requirement of key management scheme: Resilience, communication, and computational overhead affect performance of WNS. Only security scheme with highest resilience, low communication, and computation complexity should be considered for implementation. While designing of WSN security schemes, the following requirements need to address.

- To increase the resilience.
- To reduce the overall communication latency in the system.
- To minimize the average hop count in each routing path.
- To minimize the overall consumption of energy.

4.2 Node Capturing

In WSNs, most of the time user cannot deal with physical node capturing. Once it is captured network can be compromised. Compromised network portion may vary from security scheme to scheme.

4.3 Addition of Node

After configuration and initialization of WSN, new node addition is very challenging task. Authentication and communication with newly added node need attention. Few schemes protect network, if number of new nodes to be added in network increases but it affect computational overhead too.

5 Analysis of Security Schemes

Key management schemes are at center of the secure communications [19]. Configuring a secure sensor network is a non-trivial task due to resource-constrained sensors. Existing key management schemes for the WSN are not able to achieve security and efficiency requirement.

This section deals with key management and establishment schemes in detail. When security comes into the picture, key management plays crucial role in WSNs. Dynamic nature of wireless sensor networks increases challenges during designing of schemes.

5.1 Key Management Scheme

In [12], key distribution consists of three phases,

1. Pre-distribution of keys,
2. Shared-key discovery,
3. Establishment of path-key.

First phase of this scheme [12] consists of few offline steps, after key pre-distribution phase every node has to discover shared key with its neighbors. During network initialization the shared-key discovery phase occurs. In path-key establishment, phase path-key is assigned to pairs of nodes within communication range. If common key in two sensors identified (as keys are selected from key-pool randomly), sensors can use it as their pairwise key directly. Otherwise, a procedure for path-key establishment is initiated which generates a path-key. Results discussed in [12], this scheme addresses security issues but power consumption, computational speed need to be analyzed for effective implementation.

5.2 Random Key Pre-distribution Schemes

Chan [13] proposed an improved scheme based on [12]. This scheme is also known as q-composite scheme. Any two neighboring sensor nodes search a common key from their key rings to establish a secure link [13]. Proposes a modification to the base scheme in which q common keys ($q \geq 2$) are needed. When the amount of key overlap increases, the network resilience against node capture increases. If the amount overlap of required key increases, it becomes very difficult for an attacker to break a link with a given key set.

This scheme ensures high level of resilience against node capture if captured nodes are very few in numbers. This scheme may reveal large fraction of network if compromised nodes are in large numbers. If q increased it becomes difficult to obtain small amounts of initial information from the network if adversary able to capture small number of initial node.

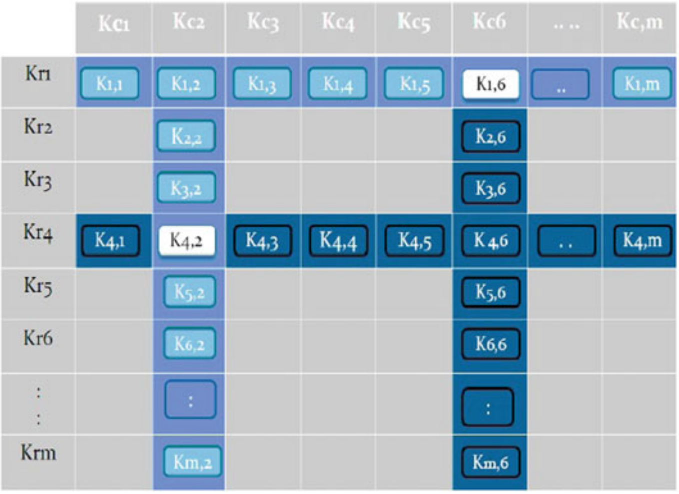
5.3 *Improved Pairwise Key Establishment Scheme for WSN*

In [12, 13] and preloaded keys are used by the communication pairwise keys between sensors. Once few nodes are compromised, the adversary may get access to other or all nodes. In [12, 13] and [20], high resilience and full network connectivity cannot be achieved.

To overcome these problems, [9] proposed enhanced pairwise key establishment scheme (called as pairwise key scheme), to achieve security and efficiency requirement of WSNs. In this scheme, two types of keys, set-up key, and pairwise key are used. Cheng and Agrawal [9] ensure on unique pairwise key generation for WSN. If adversary captured a node and exposed pairwise keys, then he cannot use captured keys to find and calculate keys of other nodes because pairwise key calculated by pair of node are not repeated and can be used by respective pair only.

In this scheme, one matrix k is prepared. KDC randomly selects column and row value to pick keys and preload in each sensor. Then, neighbors broadcasts messages to find shared keys and generate pairwise keys. Figure 3 shows common keys share by two sensor nodes. Keys are generated using the following formula.

$$\mathrm{PK}_{a-b} = \mathrm{nonce}_a \oplus k_{(a,b)} \oplus \mathrm{nonce}_b. \tag{1}$$



	Kc1	Kc2	Kc3	Kc4	Kc5	Kc6	.. ..	Kc,m
Kr1	K1,1	K1,2	K1,3	K1,4	K1,5	K1,6	..	K1,m
Kr2		K2,2				K2,6		
Kr3		K3,2				K3,6		
Kr4	K4,1	K4,2	K4,3	K4,4	K4,5	K4,6	..	K4,m
Kr5		K5,2				K5,6		
Kr6		K6,2				K6,6		
: :		:				:		
Krm		Km,2				Km,6		

Fig. 3 Common keys share by two sensor nodes

Results of test conducted by [9] shows that, this technique provides highest resilience, full network connectivity, and low communication overhead.

5.4 Scheme Based on RSA and ECC

Faleh Alfaleh and other researchers have studied and compared two popular cryptographic schemes and found it useful as compared to other cryptosystems [14]. RSA and ECC are analyzed in [14]. RSA is public-key encryption. In RSA, two keys are used for encryption and decryption process. One key is used to enciphering on sending side and another key is used to decrypt ciphertext on receivers end. RSA formula to create ciphertext is:

$$C = M^e \bmod n \tag{2}$$

RSA formula to recover original message is:

$$M = C^d \bmod n \tag{3}$$

where

M Plaint text,
C Ciphertext,
e Encryption key,
d Decryption key, and
n Some number (product of two prime numbers).

Among all asymmetric algorithms RSA is secure and reliable algorithm. In addition to this, elliptic curve cryptography (ECC) algorithm can also be used. No doubt that RSA is more secure that ECC but considering resource-constrained nature of WSNs and when various evaluation criteria are applied, it is clear that ECC is suitable for WSNs. The computation cost, less memory, low bandwidth, and running time of cryptographic operations must be managed to increase the efficiency of WSN.

5.5 Session Key Establishment Scheme

In WSNs, gateway is used to access information collected by sensor nodes. For secure communication, session key needs to be exchanged between communicating nodes. Alotaibi [15] proposed key agreement scheme. User uses biometric feature to login in the system and creating a secure connection with the node via gateway node. It restricts the password guessing attack and forgery attack. Proposed work is verified using BAN-logic. In BAN-logic, verification of the transmitted message takes place while communication. An authenticated user and sensor agree upon shared session key. Agreement of key is done when scheme is executed and user is verified. Results given [15] shows that, this scheme defends most of attacks and communication cost is average which needs to be improved.

5.6 Mixed Public and Secret-Key Cryptography

Most of the researchers recommend symmetric cryptographic methods for making wireless sensor networks secure. In WSNs security, role of key management is very important [21]. It has to be analyzed, verified, how the given security scheme affects security when criteria such as low battery consumption, low bandwidth, and high computational speed is considered. As per literature survey, it has been observed that symmetric schemes are preferred but they are not optimal at every time. Asymmetric schemes are comparatively secure but reduce performance of nodes which affects whole network. As proposed in [16], symmetric and asymmetric methods can be used by for WSNs security. In this scheme, MAC is used to generate digital signature. Evaluation shows that the proposed scheme gives maximum connectivity, only if the neighbor discovery phase is long enough. It also provides good resilience after initialization phase.

6 Analysis of Schemes

See Table 1.

7 Conclusion

In this paper, it is discussed basic of WSNs, its advantages, application, challenges, and few methods to implement WSNs effectively. While developing system for any field/area to address problem, resource-constrained nature of WSNs needs to be considered. Most challenging task is to design application/system with low power, communication limitation, and computational capability. Limited battery and processing capability restrict researchers in making highly secure network. To achieve higher level of security, then such application may consume more power. Offering security in WSNs is a difficult task. Various schemes have been proposed in literature to address problem like security, effective utilization of power, speedy computation/processing with easy implementation. To make WSNs secure, key management scheme must be simple and flexible. Most of the key management schemes are based on the key pre-distribution and pairwise key concept. After

Table 1 Analysis of WSNs schemes

Scheme	Resilience	Remark
KM Scheme by Eschenauer and Gligor [12]	High	Scheme enables key revocation but power consumption needs to be reviewed
q-composite scheme [13]	High	Resilience against node capture is high when the captured nodes number is small
Improved pairwise key establishment scheme [9]	High	Communication overhead can be reduced to improve power consumption
RSA and ECC [14]	High	ECC better that RSA. The computation cost, less memory, low bandwidth must be managed to increase the efficiency of WSN
Session key establishment scheme [15]	Good	Results shows that, this scheme defends most of attacks and communication cost is average which needs to be improved
Mixed public and secret-key cryptography [16]	Good	In this scheme, MAC is used to generate digital signature
Energy-efficient key management protocols [22]	High	Scheme is efficient only for low new node addition and authentication rate

reviewing schemes, it is clear that, each scheme has its weakness and cannot simultaneously achieve both security and efficiency requirements of WSNs. Considering current advances, schemes which will deliver desirable outcome are required.

References

1. Sridhar Raja D, Vijayan T, Kalaiselvi B (2018) Advances and recent trends in wireless sensor network. Int J Pure Appl Math 119(7)
2. Heinzelman W, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocols for wireless microsensor networks. In: Proceedings of Hawaaian international conference on systems science, Jan 2000
3. Gholizadeh I, Amiri E, Javidan R (2018) An efficient key distribution mechanism for large scale hierarchical wireless sensor networks. IEEE Sens J
4. Bashaa MH, Al-Alak SM, Idrees AK (2019) Secret key generation in wireless sensor network using public key encryption. In: Proceedings of the international conference on information and communication technology, 15–16 Apr 2019
5. Faulkner XY, Okamoto ME (2008) Secure wireless sensor networks. In: Availability, reliability and security, 2008. ARES 08. Third international conference, 4–7 Mar 2008
6. Chen X, Makki K, Yen K, Pissinou N (2009) Sensor network security: a survey. Commun Surv Tutor IEEE
7. Tahir H, Shah SAA (2008) Wireless sensor networks—a security perspective. IEEE
8. Xie H, Yan Z (2019) Data collection for security measurement in wireless sensor networks: a survey. IEEE Internet Things J
9. Cheng Y, Agrawal DP (2007) Improved pairwise key establishment for wireless sensor networks. IEEE
10. Zia T, Zomaya A (2006) Security issues in wireless sensor networks. In: ICSNC, International conference on systems and networks communication (ICSNC'06), pp 40, 2006.

11. Healy M, Newe T, Lewis E (2009) Security for wireless sensor networks: a review. 80–85. https://doi.org/10.1109/SAS.2009.4801782
12. Eschenauer L, Gligor VD (2002) A key-management scheme for distributed sensor networks. In: Proceedings of the 9th ACM conference on computer and communications security, Nov 2002
13. Chan H, Perrig A, Song D (2003) Random key pre-distribution schemes for sensor networks. In: IEEE symposium on security and privacy. Berkeley, California, 1–14 May 2003
14. Alfaleh F, Alfehaid H, Alanzy M, Elkhediri S (2019) Wireless sensor networks security: case study. In: Published in 2nd international conference on computer applications and information security (ICCAIS), May 2019
15. Alotaibi M (2018) An enhanced symmetric cryptosystem and biometric-based anonymous user authentication and session key establishment scheme for WSN. In: Published in IEEE Access, Nov 2018
16. Griotti M, Gandino F, Rebaudengo M (2017) Mixed public and secret-key cryptography for wireless sensor networks. In: Published in tenth international conference on mobile computing and ubiquitous network (ICMU)
17. Xu C, Ge Y (2009) The public key encryption to improve the security on wireless sensor networks. In: 2009 Second international conference on information and computing science, Manchester, pp 11–14. https://doi.org/10.1109/ICIC.2009.10
18. Cionca V, Newe T, Dadarlat V (2009) Setting up secure wireless sensor networks. In: It intelligent computer communication and processing, 2009. ICCP 2009. IEEE 5th international conference, 27–29 Aug 2009
19. Perrig A, Stankovic J, Wagner D (2004) Security in wireless sensor networks. Commun ACM 47(6)
20. Tasci SM, Bayramoglu E, Levi A (2008) Simple and flexible random key predistribution schemes for wireless sensor networks using deployment knowledge. In: International conference on information security and assurance
21. Cheng Y, Agrawal D (2006) Energy efficient session key establishment in wireless sensor networks. 136–142
22. Doerr L, Heigl M, Fiala D, Schramm M (2019) Comparison of energy-efficient key management protocols for wireless sensor networks. In: Proceedings of the 2019 international electronics communication conference, July 2019

Development of Agriculture Field Using Machine Learning

Rupali A. Meshram and A. S. Alvi

Abstract Indian farmers are behind as compared to other countries just not because of economic condition, but it has many reasons like they are lacking in the latest technologies, unaware about soil analysis, plant diseases, water table, quality of seeds and most important is a traditional way of farming. Indian farmers are not aware of modern way of farming. Various machine learning techniques are developed to improve farming techniques. The farmers can improve fruits quality and crop production with the help of machine learning. In this paper, we review agriculture problems that solved by using machine learning and also provide common steps that used to identify the objects from image dataset. In a nutshell, smart farming is the need of today's farmer.

Keywords Machine learning · Deep learning · Big data · Deep convolutional neural networks (CNNs) · Support vector machine (SVM)

1 Introduction

Farmers are facing various crop problems like diseases on plant, fruits ripeness, diseases on flowers, etc. Machine learning techniques are used to solve agriculture problems. Machine learning is an imminent field of computer science which can be applied to the farming sector quite effectively. Machine learning learns from past experiences and is able to build a model which would most likely be able to comprehend future instances. Deep learning is a subfield of machine learning. Deep learning is based on neural network models and work well on large sample sizes rather than small sample sizes. Deep learning provides high accuracy, existing commonly used image processing technique which is use for classification or prediction. Transfer learning is used with the deep learning. Deep transfer learning is used for object recognition and identification.

R. A. Meshram (✉) · A. S. Alvi
Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India
e-mail: rupalimeshram235@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_5

2 Literature Survey

Various researchers provide the new way of farming by introducing machine learning techniques in agriculture field. New researchers are needed to know where to use these techniques to improve crop production. Authors [1] have been proposed crop selection method (CSM) which use to solve crop selection problem, and maximize net yield rate of crop.

According to authors [2] there are three SOM-based models, namely, supervised Kohonen networks (SKN), counter-propagation artificial networks (CP-ANN) and XY-fusion (XY-F) which use to predict within field variation in wheat yield, based on online multi-layer soil data and satellite imagery crop growth characteristics. The SKN model can be used to predict wheat yield and to classify field area into different yield potential zones.

Novel active learning method proposed by authors [3] that can recognize crop and weed species by using differences in their spectral reflectance. Best results for the active learning were achieved by using SOM and MOG-based one-class classifiers while mediocre results were obtained using auto-encoder network and SVM-based one-class classifier.

The smartphone-based system that uses the classification model learnt to do real-time prediction of the state of health of a farmer's garden [4]. Remote sever can determine the state of disease of a plant through the uploaded image.

Authors [5] have employed SVM using different kernel functions including Cauchy kernel, Invmult Kernel and Laplacian Kernel. Grid search and N-fold cross-validation techniques which extract relevant features related to image of tomato leaf and to detect and identify type of disease that infects tomato plant.

For an automatic crop disease recognition method, the probabilistic neural networks (PNNs) classifier used by authors [6], which combined the statistical features of leaf images and meteorological data.

Principle component analysis (PCA) and support vector machine (SVM) were classifying infected/uninfected tomato fruits according to its external surface; it uses feature fusion method with color and texture features [7]. The proposed system, grading tomato fruits based on surface defects.

To identify the phenological stage, classification of rice crop with multi-temporal co-polar X-band SAR images, authors [8] were used support vector machines (SVM) with linear and nonlinear kernel, k-nearest neighbors (kNN) and decision trees (DT).

Support vector machine (SVM) and a random forest (RF) method used by authors [9] for assessing the feasibility of in-season winter wheat mapping and investigating potential classification. It is improvement by using Synthetic Aperture Radar(SAR) images, optical images, and the integration of both types of data in urban agricultural regions with complex planting structures in Southern China. Authors [9] used convolutional neural network (CNN) to predict the type of previously unseen images of plant seedling.

Table 1 gives detail about methodology used on particular problem of agriculture field with future scope provided by researchers.

3 Summary and Discussion

Table 1 provides brief literature about yield prediction, disease detection, weed detection, crop quality, species recognition, water management and soil management, by referring Table 1, new researchers get the knowledge about which algorithm or methodology used by previous researchers and it also provides future scope to particular problem. It is observed that the deep convolutional neural networks (CNNs) and probabilistic neural networks (PNNs) classifier used to identify the plant diseases. The deep convolutional neural networks (CNNs) applied on various plants like tomato, rice and banana to identify their diseases. Also, it is used to predict types of images of plant seedling, counting fruits, plant identification and exacting plant images. Whenever we need to solve a problem involving with plant images and object detection from the image in that case we can apply deep convolutional neural networks (CNNs). The support vector machine (SVM) is used with other techniques for soil estimation, healthy and diseased plant leaf, ripeness of fruits, plant diseases and stages of crops or fruits. Fuzzy rule-based classification approach (FRBCS) also identifies ripeness of fruits. The support vector machine (SVM) applied on various plants like tomato, rice, grapes and wheat. A new time series model based on long short-term memory (LSTM) used to predict water table depth, especially in areas where hydrogeological data are difficult to obtain. Extreme learning machines (ELM), multivariate regression splines (MARS), M5 tree and support vector regression (SVR) for groundwater contamination risk mapping, in the complex aquifer system. Smart irrigation decision support system (SIDSS) estimates the weekly irrigations needs of a plantation. Groundwater prediction is also has a more scope to develop new algorithm to improve crop productions. CNN gives the better performance as compared to other detection techniques of machine learning. CNN required huge dataset to train model.

Performance of various models proposed by researcher (studied in literature) is also important for further study. (1) Accuracy of CSM method depends on predicted value of influenced parameters. (2) SKN network for the prediction of wheat yield with a correct classification reached 91.3% for both cross-validation and independent validation. (3) The image processing techniques with DSS using multiclass SVM gives accuracy up to 96.66% for grape plant disease classification. (4) Classification accuracy 94.29% achieved by fuzzy rule-based learning algorithm. (5) CNN and fine-tuning give 96.3%. (6) SVMs provide accuracy of 90.2% for plant disease detection. (7) Deep simulated learning for counting fruits provides 91%. (8) PNN classifier correct rate was 91.08%. (9) CNNs-based rice disease detection model achieves an accuracy of 95.48%. Researchers can extend this mention algorithm for getting more accuracy in achieved result.

Table 1 Survey on agriculture problem

Methodology	Purpose	Future scope
Multidisciplinary model based on (IoT), sensors, cloud computing, mobile computing, big data analysis [10]	Increase in agricultural production and for cost control of agro-products	Focusing on interfacing different soil nutrient sensors with beagle black bone. Use various data mining algorithms for agricultural big data analysis for getting the preferred outcome
Deep neural network (DNN) [11]	A cost-effective and high-performance option for field-level and in-season crop-type classification for corn/soybean dominated corn belt landscape (case study in Champaign County, Illinois)	Texture features extracted from high temporal-spatial resolution fusion data. To extending this approach and including other types of crop types and natural vegetation types. Reducing expected uncertainties in the classification approach. To use multi-sensor fusion data, such as the STARFM algorithm
convolution neural networks(CNN) [12]	Classification method for high-resolution agricultural remote sensing images	If the multi-temporal characteristics of crop remote sensing images can be added to the CNN training, the prediction may get good results
A novel pipeline [13]	Accurately extracts both detected object regions and dense semantic segmentation for extracting both stalk counts and stalk width. With novel pipeline, they demonstrated accurate measurement of multiple plants attributes	To integrate more accurate positioning to merge multiple views of the stalks into more accurate measurements of stalk count and stalk width
Least squares support vector machines and cubist method compare over the two multivariate methods: principal component regression and partial least squares regression [14]	Estimating soil total nitrogen (TN), organic carbon (OC) and moisture content (MC). LS-SVM best for MC and OC and cubist for TN estimation of soil	This machine learning techniques can be used in field spectroscopy for offline and online prediction of the soil parameters studied in fields with similar soil type and variability

(continued)

Table 1 (continued)

Methodology	Purpose	Future scope
A new time series model based on long short-term memory (LSTM) [15]	Predicting water table depth, especially in areas where hydrogeological data are difficult to obtain	Deeper, wider and more robust LSTM-based model, in order to provide more accurate water table depth prediction worldwide. The proposed model also can be combined with other methods, like PCA and wavelet transform. Also, it can be applied to other time series prediction tasks, such as soil water change and streamflow prediction
Smart irrigation decision support system (SIDSS) [16]	Estimates the weekly irrigations need of a plantation, on the basis of both soil measurements and climatic variables	To extend and evaluate the system in plantations different than citrus and analyze the performance under several conditions and regions. To improve the accuracy of the system the past rainfall information may be used
Data mining techniques, optimization model with genetic algorithm [17]	Optimal water allocation relative to maximizing irrigation uniformity and minimizing yield reduction	Address other objectives, such as economical return and improved yield
Fuzzy rule-based classification approach (FRBCS) [18]	To estimate the ripeness of tomatoes based on color	The proposed approach can also be applied for other climacteric crops such as mango and bell pepper. External features other than colors such as shape, size, and texture can be involved for classification of tomato. This will help in better quality evaluation

(continued)

Table 1 (continued)

Methodology	Purpose	Future scope
Decision support systems (DSS) using multiclass SVM [19]	To perform classification between healthy and diseased leaf of grapes plants	The accuracy of the system can be further improved by improving the training ratio
Deep convolutional networks [20]	Classify and detect plant diseases from leaf images	Developing a complete system that contains the server-side components with a trained model and smart mobile devices application. It contains features like displaying identified diseases in fruits, vegetables, and other plants, based on leaf images captured by the mobile phone camera The usage of the model by training it for plant disease recognition on wider land areas, combining aerial photos of orchards and vineyards captured by drones and convolution neural networks for object detection
Convolutional neural network models [21]	Plant disease detection and diagnosis using simple leaves images of healthy and diseased plants	The development of an automated pesticide prescription system that would require a confirmation by the automated disease diagnosis system to allow the purchase of appropriate pesticides by the farmers. That would drastically limit the uncontrolled acquisition of pesticides that leads to their overuse and misuse, with the consequent catastrophic effects on the environment
Deep convolutional network [22]	Plant identification from leaf vein patterns	Deep learning in agriculture, in particular weeds detection and identification, and seeds viability tests

(continued)

Table 1 (continued)

Methodology	Purpose	Future scope
An automated multiclass classification approach uses principal components analysis (PCA) in addition to support vector machines (SVMs) and linear discriminant analysis (LDA) algorithms [23]	Tomato ripeness measurement and evaluation via investigating and classifying the different maturity/ripeness stages	To apply the proposed approach on different crops, other than tomatoes, in order to automate the whole process of harvesting and detect damages to save crops To use non-destructive/non-invasive detection technologies of food quality/maturity such as hyperspectral imaging systems, colorimetric, near-infrared spectroscopy, and non-invasive smart sensing technologies
Deep convolutional neural networks (CNNs) [24]	Effectively classify 10 common rice diseases through images recognition	To apply other deep architectures and other training algorithms, such as the restricted Boltzmann machine which achieves a better performance on object recognition. Model can extend for fault diganosis. Deeper analysis of the training method with and without labeled samples. The model result could be further extended to the distributed state estimation problems for sensor networks and nonlinear time-varying systems
Deep convolutional neural network [25]	Counts fruits efficiently even if fruits are under shadow, occluded by foliage, branches, or if there is some degree of overlap among fruits (tomato)	To count fruits in all stages; add green fruits to the synthetic dataset. To develop a mobile application based on the proposed algorithm which can be used directly by farmers for yield estimation and cultivation practices
Convolution neural networks [26]	To identify and classify banana diseases	To test more banana and plants diseases with the proposed model
Partial least square regression (PLSR), ν support vector regression (ν-SVR), and Gaussian process regression (GPR) methods [27]	For wheat leaf rust disease detection as well as evaluating the training sample size and influence of disease symptoms effects on methods predictions	PLSR, ν-SVR, and GPR need to be tested on various sensors and different varieties of wheat in order to be used in the field

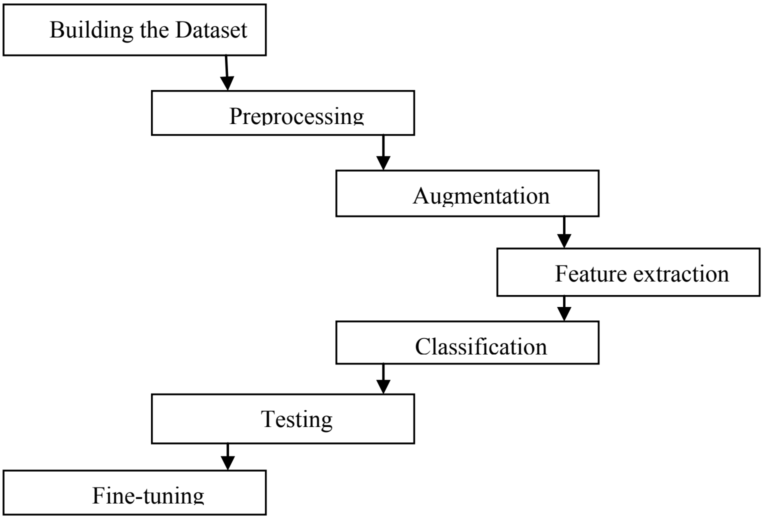


Fig. 1 Steps to process the data

4 Processing of Dataset

Datasets can be collected in many ways (online/offline). Data preprocessing/preparation/cleaning is the process of detecting and removing corrupt or inaccurate records from a dataset. Image data augmentation is a technique that can be used to artificially expand the size of a training dataset by creating modified versions of images in the dataset. We can improve the performance of the model by augmenting the data we already have (Fig. 1).

Feature extraction used to extract relevant feature/weight. Classification is the process of predicting some identified data. There are a lot of classification algorithms are available like decision tree, Naive Bayes, artificial neural networks, *k*-nearest neighbor (KNN), etc. Fine-tuning is a process to take a network model that has already been trained for a given task, and make it perform a second similar task. Fine-tuning machine learning predictive model is a crucial step to improve accuracy of the forecasted results.

5 Conclusion

The aim of this article is to motivate more researchers to analyze and experiment with machine learning. The overall benefits of survey are encouraging for its further use toward smarter, more sustainable farming and increasing crop production. This article performed a review of machine learning in agriculture. Researcher were identified and analyzed, examining the problem they addressed, tools/techniques, the proposed solution and future scope. It provides various methodologies used on different plant/fruits so researchers can identify other plants/fruits for their research for existing problems in agriculture field. Augmentation and fine-tuning are important steps to improve the performance.

References

1. Kumar R, Singh MP, Kumar P, Singh JP (2015) Crop selection method to maximize crop yield rate using machine learning technique. In: International conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India, pp 138–145 6–8 May 2015
2. Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM (2016) Wheat yield prediction using machine learning and advanced sensing Techniques. Comput Electron Agric 121:57–65
3. Pantazi X-E et al (2016) Active learning system for weed species recognition based on hyperspectral sensing. Biosys Eng. https://doi.org/10.1016/j.biosystemseng.2016.01.014
4. Owomugisha G, Mwebaze E (2016) Machine learning for plant disease incidence and severity measurements from leaf images. In: 2016 15th IEEE international conference on machine learning and applications, 978-1-5090-6167-9/16 $31.00©2016, IEEE. https://doi.org/10.1109/ICMLA.2016.126, pp 158–163
5. Mokhtar U, Alit MAS, Hassenian AE, Hefny H (2015) Tomato leaves diseases detection approach based on support vector machines. IEEE, pp 246–250
6. Shi Y, Wang XF, Zhang SW, Zhang CL (2015) PNN based crop disease recognition with leaf image features and meteorological data. Int J Agric Biol Eng 8(4):60–68
7. Semary NA, Tharwat A, Elhariri E, Hassanien AE (2014) Fruit-based tomato grading system using features fusion and support vector machine. In: Filev D et al (eds) Intelligent systems 2014, advances in intelligent systems and computing 323. https://doi.org/10.1007/978-3-319-11310-4_35, pp 401–410
8. Küçük Ç, Taşkın G, Erten E (2016) Paddy-rice phenology classification based on machine-learning methods using multitemporal Co-polar X-Band SAR images. IEEE J Sel Topics Appl Earth Observ Remote Sens: 1–11
9. Ashqar BAM, Abu-Nasser BS, Abu-Naser SS (2019) Plant seedlings classification using deep learning. Int J Acad Inf Syst Res IJAISR 3(1):7–14. ISSN: 2000-002X
10. Channe H et al (2015) Multidisciplinary model for smart agriculture using internet-of-things (IoT), sensors, cloud-computing, mobile-computing and big-data analysis. Int J Comput Technol Appl 6(3):374–382. ISSN: 2229-6093.
11. Cai Y et al (2018) A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens Environ 210:35–47
12. Chunjing Y, Yueyao Z, Yaxuan Z, Liu H (2017) Application of convolutional neural network in classification of high resolution agricultural remote sensing images. In: The international archives of the photogrammetry, remote sensing and spatial information sciences, vol XLII-2/W7, 2017 ISPRS Geospatial Week 2017, Wuhan, China, 18–22 Sept 2017
13. Baweja HSB, Parhar T, Mirbod O, Nuske S (2018) StalkNet: a deep learning pipeline for high-throughput measurement of plant stalk count and stalkwidth. In: Hutter M, Siegwart R (eds) Field and service robotics, springer proceedings in advanced robotics 5. Springer International Publishing AG 2018. https://doi.org/10.1007/978-3-319-67361-5_18271-284
14. Morellos A et al (2016) Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosys Eng. https://doi.org/10.1016/j.biosystemseng.2016.04.018

15. Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018) Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J Hydrol. https://doi.org/10.1016/j.jhydrol.2018.04.065
16. Navarro-Hellín H et al (2016) A decision support system for managing irrigation in agriculture. Comput Electron Agric 124:121–131. https://doi.org/10.1016/j.compag.2016.04.003,0168-1699/_2016ElsevierB.V
17. Hassan-Esfahani L et al (2015) Assessment of optimal irrigation water allocation for pressurized irrigation system using water balance approach, learning machines, and remotely sensed data. Agric Water Manag 153:42–50. https://doi.org/10.1016/j.agwat.2015.02.0050378-3774/%C2%A92015ElsevierB.V
18. Goela N, Sehgal P (2015) Fuzzy classification of pre-harvest tomatoes for ripeness estimation—an approach based on automatic rule learning using decision tree. Appl Soft Comput 36:45–56
19. Waghmare H, Kokare R, Dandawate Y (2016) Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system. In: 3rd international conference on signal processing and integrated networks (SPIN), pp 513–518
20. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D (2016) Deep neural networks based recognition of plant diseases by leaf image classification. In: Computational intelligence and neuroscience, vol 3289801, p 11. Hindawi Publishing Corporation. https://dx.doi.org/10.1155/2016/3289801
21. Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. Comput Electron Agric 145:311–318
22. Grinblat GL, Uzal LC, Larese MG, Granitto PM (2016) Deep learning for plant identification using vein morphological patterns. Comput Electron Agric 127:418–424
23. El-Bendary N et al (2014) Using machine learning techniques for evaluating tomato ripeness. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2014.09.057
24. Lu Y et al (2017) Identification of rice diseases using deep convolutional neural networks. Neurocomputing 267:378–384. https://doi.org/10.1016/j.neucom.2017.06.0230925-2312/%C2%A92017ElsevierB.V
25. Rahnemoonfar M, Sheppard C (2017) Deep count: fruit counting based on deep simulated learning . MDPI Sensors 17:905. https://doi.org/10.3390/s17040905pp.1-12
26. Amara J, Bouaziz B, Algergawy A (2017) A deep learning-based approach for banana leaf diseases classification. In: Mitschang B et al. (Hrsg.): BTW 2017—workshopband, lecture notes in informatics (LNI), Gesellschaft für Informatik, Bonn 2017, pp 79–88
27. Ashourloo D, Aghighi H, Matkan AA, Mobasheri MR, Rad AM (2016) An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. IEEE J Sel Topics Appl Earth Observ Remote Sens: 1–8

Security Issues in Cloud Computing

Akram H. Shaikh and B. B. Meshram

Abstract Cloud computing services like software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) are delivery models which provides software, application platform, and infrastructure recourses to consumer respectively. This paper presents service models, deployment models, and essential characteristics of cloud computing and cloud security and privacy issues. This paper also describes detailed literature on cloud security issues and layerwise classification of each issue. The various vulnerabilities, attacks, and defense mechanisms are provided for securing cloud environment

Keywords Cloud computing cloud services · Vulnerabilities · Attack · Defense mechanisms

1 Introduction

During the last decade, the Internet has known a great invasion that has totally changed the world. This rapid evolution has made access to the Internet an easy task with the advent of different devices like cellular phones, laptops, tablets, and others that connected to the virtual world anywhere and at any time. In other hand, multi-tenant architecture, virtualization, remote access, the ubiquity of broadband networks, and the establishment of universal standards of interoperability between software. All of these elements had been a good step for the emergence of cloud computing in the world of distributed systems as a successor of cloud computing. However, with all of these benefits, there are still a number of barriers, that effect the

Akram H. Shaikh (✉)
Research Scholar, Department of Computer Engineering, Veermata Jijabai Technological Institute, Matunga, Mumbai, India
e-mail: ahshaikh_p18@ce.vjti.ac.in

B. B. Meshram
Professor, Department of Computer Engineering, Veermata Jijabai Technological Institute, Matunga, Mumbai, India
e-mail: bbmeshram@ce.vjti.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_6

widespread adoption of cloud computing, and the most important one is security. The cloud computing system is based on the trust, that makes security and confidentiality the major issues. These and other security issues whether on the technical or on the legal plan decrease the confidence of cloud computing and its adoption. This paper covers an overview about security issues in cloud computing.

The paper is organized below.

- Section 2 describe cloud computing model.
- Section 3 describes cloud security issues and problems.
- Section 4 identifies the vulnerabilities threat and solutions for cloud security and last section concludes the results.

2 Cloud Computing Model

According to the National Institute of Standards and Technology (NIST), cloud computing is defined as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and it is composed of three service models, deployed in four deployment models, and offers five characteristics [1] as shown in Fig. 1.

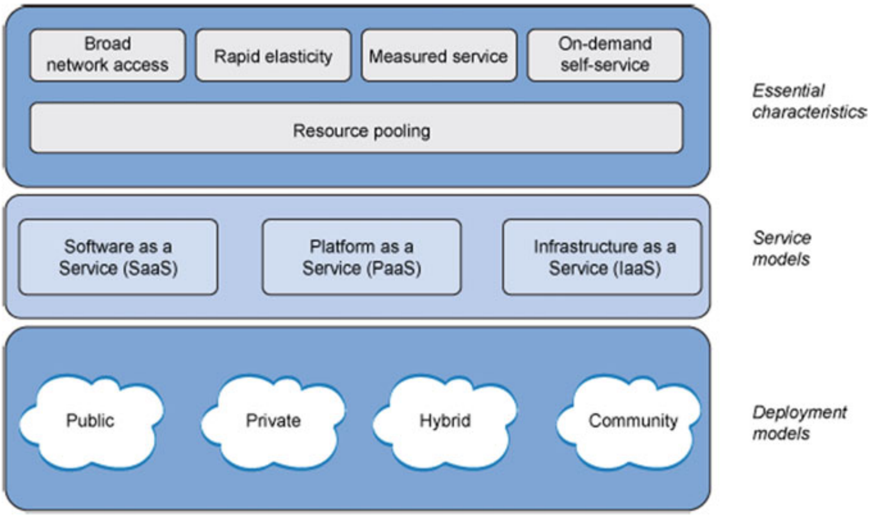


Fig. 1 Cloud computing services

2.1 Service Models

1. Software as a service (SaaS): The capability provided to the client is to utilize the supplier's applications running on a cloud foundation and open from different customer devices through a thin customer interface, for example, Internet browser. In other words, in this model, a total application is offered to the client as a service on request. A good example of this would be using a Web browser to view email as provided by Microsoft, Yahoo, or Google [1].
2. Platform as a service (PaaS): Here, the consumer can deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations [1].
3. Infrastructure as a service (IaaS): The consumer can provision processing, storage, networks, and other fundamental computing resources and he is able to deploy and run arbitrary software, which can include operating Systems and applications. The consumer does not manage or control the underlying cloud infrastructure, but has control over operating systems; storage, deployed applications, and possibly limited control of select networking components [1].

2.2 Deployment Models

1. Public: The cloud infrastructure is available to the public or a large industry group, and it is owned by an organization selling cloud services.
2. Private: The cloud infrastructure operated solely for an organization. It may be managed by the organization.
3. Community: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns. It is managed by organizations or a third party and may exist on premise or off premise.
4. Hybrid: The cloud infrastructure is a composition of two or more

2.3 Essential Characteristics

1. On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically.
2. Broad network access: All services in cloud computing are available through the network and accessed by different devices such as desktop computer, mobile phones, smart phones, and tablet devices…
3. Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and

virtual resources dynamically assigned and reassigned according to consumer demand.

4. Rapid elasticity: Resources can be elastically provisioned and released. The user is free to purchase additional resources and opportunities in any quantity and at any time.
5. Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering capability (pay-per-use basis) at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts).
6. Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering capability (pay-per-use basis) at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts).

3 Cloud Security Issues

3.1 Software as a Service (SaaS) Security Issues

SaaS provides application services on demand such as email, conferencing software, and business applications such as ERP, CRM, and SCM [2]. SaaS users have less control over security among the three fundamental delivery models in the cloud. The adoption of SaaS applications may raise some security concerns.

1. **Application security**: These applications are typically delivered via the Internet through a Web browser. However, flaws in Web applications may create vulnerabilities for the SaaS applications. Attackers have been using the Web to compromise user’s computers and perform malicious activities such as steal sensitive data [2]. The Open Web Application Security Project (OWASP) has identified the ten most critical Web applications security threats [3]. There are more security issues, but it is a good start for securing web applications.
2. **Multi-tenancy**: SaaS applications can be grouped into maturity models that are determined by the following c characteristics: scalability, configurability via metadata, and multi-tenancy [2]. In multi-tenancy, a single instance serves all customers. This approach enables more efficient use of the resources but scalability is limited. Since data from multiple tenants are likely to be stored in the same database, the risk of data leakage between these tenants is high. Security policies are needed to ensure that customer’s data are kept separate from other customers [4].
3. **Data security**: Information security is a typical worry for any innovation, yet it turns into a significant test when SaaS clients need to depend on their suppliers for legitimate security [4]. In SaaS, authoritative information is regularly prepared in plaintext and put away in the cloud. The SaaS supplier is the one liable for the security of the information while is being handled and put away [2].

4. **Accessibility**: Getting to applications over the Web by means of Internet browser makes access from any system device, including open PC and cell phones. In any case, it likewise opens the support of extra security dangers. The Cloud Security Alliance [5] has discharged a report that portrays the present condition of versatile registering and the top dangers here, for example, data taking versatile malware, unreliable systems (Wi-Fi), and vulnerabilities found in the network OS and authority applications, uncertain commercial centers, and nearness-based hacking.

3.2 Platform as a Service (PaaS) Security Issues

PaaS facilitates deployment of cloud-based applications without the cost of buying and maintaining the underlying hardware and software layers [4]. Same as SaaS, PaaS also brings data security issues and other challenges that are described as follows.

1. **Third-party relationships**: Moreover, PaaS does not only offer conventional programming languages, but also does it offer third-party Web services mechanism such as mashups. Mashups merge more than one source element into a single integrated unit. Thus, PaaS models also succeed to security issues related to mashups such as data and network security [4]. Also, PaaS users have to depend on both the security of Web-hosted development tools and third-party services.
2. **Underlying infrastructure security**: In PaaS, developers do not usually have access to the underlying layers, so providers are responsible for securing the underlying infrastructure as well as the applications services [4]. Even when developers are in control of the security of their applications, they do not have the assurance that the development environment tools provided by a PaaS provider are secure.

3.3 Infrastructure as a Service (IaaS) Security Issues

IaaS provides a group of resources such as servers, storage, networks, and other computing resources in the form of virtualized systems, which are accessed through the Internet [1]. Users are entitled to run any software with full control and management on the resources allocated to them. IaaS providers must undertake a substantial effort to secure their systems in order to minimize these threats that result from creation, communication, monitoring, modification, and mobility. Here is some of the security issues associated to IaaS.

1. **Virtualization**: Virtualization enables clients to make duplicate, share, relocate, and move back virtual machines, which may enable them to run an assortment of utilizations [1]. In any case, it likewise presents new open doors for assailants in view of the additional layer that must be verified [4]. Virtual machine security becomes as significant as physical machine security, and any defect in it is

possible that one may influence the other. Virtualized situations are defenseless against a wide range of assaults for ordinary frameworks; be that as it may, security is a more noteworthy test as virtualization includes more purposes of section and more interconnection multifaceted nature. In contrast to physical servers, VMs have two limits: physical and virtual [6].

2. **Virtual machine monitor**: The virtual machine monitor (VMM) or hypervisor is liable for virtual machines detachment; hence, if the VMM is undermined, its virtual machines may conceivably be undermined too. The VMM is a low-level programming that controls and screens its virtual machines, so as any conventional programming it involves security defects [1].
3. **Virtual networks**: System segments are shared by various inhabitants because of asset pooling. As referenced previously, sharing assets enables assailants to dispatch cross-occupant assaults. The most secure path is to snare each VM with its host by utilizing devoted physical channels. Nonetheless, most hypervisors utilize virtual systems to interface VMs to convey all the more legitimately and effectively. For example, most virtualization stages, for example, Xen, give two different ways to arrange virtual systems: spanned and steered, yet these procedures increment the likelihood to play out certain assaults, for example, sniffing and parodying virtual system [4].

4 Vulnerabilities Threats and Solutions for Cloud Security

This section deals with table threats vulnerabilities and solutions.

4.1 Threats: Account or Service Hijacking

A record theft can be performed by various ways, for example, social building and powerless accreditations. In the event that an aggressor accesses a client's qualification, he can perform noxious exercises, for example, get to delicate information, control information, and divert any exchange [7]. Due to insecure interfaces and APIs vulnerability in SaaS, PaaS, IaaS account or service, hijacking attack is made, the possible solutions for which are 1. Identity and access management guidance [8]. 2. Dynamic credential [9] as detailed as shown in Table 1.

4.1.1 Vulnerabilities: Insecure interfaces and APIs

Cloud providers present services that can be used through APIs. (The Security of the cloud depends upon the security of interfaces) [7]. Some problems are:

1. Powerless accreditations
2. Insufficient authorization checks

Table 1 Layerwise threats vulnerabilities and solutions

S. No.	Delivery model	Threats	Vulnerabilities	Solution or defense mechanism
1	SaaS PaaS IaaS	Account or service hijacking	Insecure interfaces and APIs	1. Identity and access management guidance [8] 2. Dynamic credential [9]
2	SaaS PaaS IaaS	Data leakage	1. Data-related vulnerabilities 2. Vulnerabilities in virtual machines 3. Vulnerabilities in virtual machine images 4. Vulnerabilities in virtual networks	FRS techniques [10] Digital Signatures [11] Encryption [12] Homomorphic encryption [12]
3	SaaS PaaS IaaS	Denial of service	1. Insecure interfaces and APIs, 2. unlimited allocation of resources	Cloud providers can force policies to offer limited computational resources 1. Filtering tree Trace back and filter system
4	SaaS	Customer data manipulation	Insecure interfaces and APIs	Web application scanners [13]
5	IaaS	VM escape	Vulnerabilities in hypervisors	Hypersafe [14] Trusted cloud computing platform (TCCP) [15]
6	IaaS	Malicious VM creation	Vulnerabilities in virtual machine images	Mirage [16]
7	IaaS	Insecure VM migration	Vulnerabilities in virtual machines	PALM [17] VNSS [18]
8	IaaS	Sniffing/spoofing virtual networks	Vulnerabilities in virtual networks	Virtual network framework based on Xen network modes: “bridged” and “routed” [19]

3. Insufficient input data validation.

4.1.2 Countermeasures for Account or Service Hijacking

Below are the solutions of account or service hijacking.

1. **Identity and access management guidance**: Cloud Security Alliance (CSA) is a non-benefit association that advances the utilization of best practices so

as to give security in cloud situations. CSA has given an identity and access management guidance [8] which gives a rundown of prescribed best rehearsed to guarantee personalities and secure access the executives. This report incorporates centralized directory, access management, identity management, role-based access control, user access certifications, privileged user and access management, separation of duties, and identity and access reporting.

2. **Dynamic credentials**: Xiao and Gong [9] present a algorithm to make dynamic accreditations for mobile cloud computing frameworks. The dynamic qualification changes its worth once a client changes its area or when he has traded a specific number of information parcels.

4.2 Threats: Data Leakage

Data leakage happens when the data gets into the wrong hands while it is being transferred, stored, and audited or processed [7]. Due to data-related vulnerabilities in virtual machines, vulnerabilities in virtual machine images, Vulnerabilities in virtual networks on SaaS, PaaS, IaaS data leakage attack is made, the possible solutions for which are FRS techniques [10], digital signatures [11], encryption [12], homomorphic encryption [12].

4.2.1 Vulnerabilities

Here is some vulnerability in data leakage

1. **Data-related vulnerabilities**

Here, in this type of vulnerabilities, data can be collocated with unknown owners, incomplete data cannot be completely removed, data backup done by untrusted user data deduplication: a technique that stores only a copy of redundant data which may be not secured.

2. **Vulnerabilities in virtual machines**

   1. Possible covert channels in the collocation of VMs
   2. Unrestricted allocation and deallocation of resources with VMs [4]
   3. Uncontrolled migration---VMs can be migrated from one server to another server due to fault tolerance, load balance, or hardware maintenance
   4. Uncontrolled snapshots---VMs can be copied in order to provide flexibility, which may lead to data leakage
   5. Uncontrolled rollback could lead to reset vulnerabilities.

4.2.2 Countermeasures for: Data Leakage

There are mainly three solutions for Data leakage attack which are listed below.

1. **Fragmentation-redundancy-scattering (FRS) technique** [10]: This method means to give interruption resilience and, in result, secure storage. This method comprises of first separating delicate information into irrelevant parts, so any piece does not have any noteworthy data independent from anyone else. At that point, pieces are dissipated in a repetitive style crosswise over various locales of the disseminated framework.
2. **Digital signatures**: Somani et al. [11] propose to verify information utilizing computerized signature with RSA calculation while information is being moved over the Internet. They guaranteed that RSA is the most conspicuous calculation, and it very well may be utilized to secure information in cloud situations.
3. **Homomorphic encryption**: The three fundamental activities for cloud information are move, store, and procedure. Encryption strategies can be utilized to verify information while it is being moved all through the cloud or put away in the supplier's premises. Cloud suppliers need to unscramble figure information so as to process it, which raises protection concerns. In [12], they propose a technique dependent on the utilization of completely homomorphic encryption to the security of mists. Completely homomorphic encryption permits performing subjective calculation on figure writings without being unscrambled. Current homomorphic encryption plans bolster predetermined number of homomorphic tasks, for example, expansion and duplication.

4.3 Threat: Denial of Service

It is possible that a malicious user will take all the possible resources. Thus, the system cannot satisfy any request from other legitimate users due to resources being unavailable. Due to the vulnerabilities in insecure interface and APIs and unlimited resource allocation, denial-of-service attack is made the possible solution for which is filtering tree and trace back filtering system.

4.3.1 Vulnerabilities

Here are some vulnerabilities of denial service.

1. **Insecure interfaces and APIs**

Cloud providers offer services that can be accessed through APIs. The security of the cloud depends upon the security of these interfaces [14]. Some problems are weak credential, and insufficient authorization checks insufficient input data validation

2. **Unlimited allocation of resources**

Inaccurate modeling of resource usage can lead to overbooking or over-provisioning.

4.3.2 Countermeasures for Threat: Denial of Service (Denial-Of-Service Prevention Mechanisms)

There are two possible solutions which are filtering tree and trace back filtering system which are listed below

1. **A filtering tree** that goes about as an assistance intermediary inside a service-oriented architecture (SOA) model is displayed. The creators inspect the uncertainties in institutionalized cloud APIs and how these can be misused in provisioning, the board and checking of administrations. They propose adding a mark reference component to each SOAP solicitation to guarantee that it originates from a genuine source. Twofold marks are created utilizing hashed qualities of each SOAP envelope, for example, the quantity of youngster or header components. The customer IP address is kept in the message header alongside a riddle that is put away as a component of the WSDL record. The proposed framework needs to examine every bundle exclusively, which can prompt a bottleneck in DDoS circumstances.
2. **A trace back and filter system is proposed** to protect the cloud from DDoS attacks. SOA trace back is used by adding a tag to SOA packets to record the route taken. This system cannot identify the source of attack, because the tag is only added to the packet once it is relatively close to the server. The tests used in the paper do not consider spoofed IP addresses or the fact that an attacker is likely to make use of zombie machines.

4.4 Threat: Customer Data Manipulation

Due to data-related vulnerabilities in insecure interfaces and APIs on SaaS customer data manipulation attack is made, the possible solutions for which are Web application scanners [13]. Users attack Web applications by manipulating data sent from their application component to the server's application. For example, SQL injection, command injection, insecure direct object references, and cross-site scripting

4.4.1 Vulnerabilities

Vulnerabilities of customer data manipulation are insecure interfaces and APIs.

1. **Insecure interfaces and APIs**

Cloud providers present services that can be used through APIs

4.4.2 Countermeasures for Threat: Customer Data Manipulation

Here is the possible solution for customer data manipulation.

1. **Web application scanners**: Web applications can be an easy target because they are exposed to the public including potential attackers. A Web application scanner [13] is a program which scans Web applications through the Web front-end in order to identify security vulnerabilities.

4.5 Threat: VM Escape

It is designed to exploit the hyper-visor to take control of underlying infrastructure [1]. Due to Data related vulnerabilities in hyper visor VM escape attack is made, the possible solutions for which are Hyper Safe [14] and TCCP9 Trusted Cloud Computing Platform) [15].

4.5.1 Vulnerabilities in Hypervisors

1. Complex hypervisor code [1]
2. Flexible configuration of VMs or hypervisors to meet organization needs can be exploited.

4.5.2 Countermeasures for Threat: VM Escape

1. **Hypersafe** [14]: It is a methodology that gives hypervisor control-stream trustworthiness. Hyper Safe Security will probably ensure two types of Procedure, first is Non-acceptable memory lock-down which uses compose shielded memory pages from being adjusted and second is Limited pointed ordering that changes in control information from pointer records. So as to assess the adequacy of this methodology, they have led four kinds of assaults, for example, adjust the hypervisor code, execute the infused code, alter the page table, and alter from an arrival table. They inferred that hypersafe effectively anticipated every one of these assaults, and that the presentation overhead is low.
2. **Trusted cloud computing platform**: TCCP [15] enables suppliers to offer shut box execution conditions and enables clients to decide whether nature is secure before propelling their VMs. The TCCP includes two basic components: a trusted virtual machine monitors (TVMM) and a trusted coordinator (TC). The TC deals with a lot of confided in hubs that run TVMMs, and it is kept up yet a confided in outsider. The TC partakes during the time spent propelling or moving a VM, which confirms that a VM is running in a confided in stage.

4.6 Threat: Malicious Virtual Machine Creation

An attacker who creates a legitimate account can make a VM image containing malicious code such as a Trojan horse and store it in the provider storehouse [1]. Due to virtual machine image vulnerability on IaaS malicious virtual machine creation attack is made, the possible solutions for which is mirage [16].

4.6.1 Vulnerabilities in Virtual Machine Images

1. Uncontrolled placement of VM images in public repositories [6]
2. VM images are not able to be patched since they are dormant artifacts.

4.6.2 Countermeasures for Malicious Virtual Machine Creation

1. **Mirage**: In [16], the authors suggest a virtual machine image management system in a cloud computing atmosphere. This technique includes security features: access control framework, image filters, a provenance tracking, and repository maintenance services.

4.7 Threats: Insecure Virtual Machine Migration

Live migration of virtual machines exposes the contents of the VM state files to the network. An attacker can do the following actions: 1. Access data illegally during migration [4]. 2. Transfer a VM to an untrusted host [4] Create and migrate several VM causing disruptions or DoS. Due to vulnerabilities in virtual machine on IaaS insecure virtual machine migration attack is made, the possible solutions for which are protection aegis for live migration of VMs [17] and VNSS [18].

4.7.1 Vulnerabilities in Virtual Machines

Uncontrolled snapshots---VMs can be copied in order to provide flexibility, which may lead to data leakage.

4.7.2 Countermeasures for Threat: Insecure Virtual Machine Migration

Here are the possible solutions for insecure virtual machine migration.

1. **Protection aegis for live migration of VMs (PALM):** Zhang et al. [17] proposes a secure live migration framework that jams uprightness and security insurance

during and after movement. The model of the framework was executed dependent on Xen and GNU Linux, and the consequences of the assessment demonstrated that this plan just includes slight personal time and relocation time because of encryption and unscrambling.

2. **VNSS**: Xiaopeng et al. [18] propose a security framework that customizes security policies for each virtual machine, and it provides continuous protection thorough virtual machine live migration. They implemented a prototype system based on Xen hypervisors using stateful firewall technologies and user space tools such as ip tables, xm commands program, and conntrack-tools.

4.8 Threat: Sniffing/Spoofing Virtual Networks

A malicious VM can listen to the virtual network or even use ARP spoofing to redirect packets from/to other VMs. Due to vulnerabilities in virtual networks on IaaS sniffing/spoofing virtual networks attack is made, the possible solutions for which is virtual network security [19].

4.8.1 Vulnerabilities in Virtual Networks

Sharing of virtual bridges by several virtual machines [19].

4.8.2 Countermeasures for Threats: Sniffing/Spoofing Virtual Networks

Here is the list of possible solutions for sniffing/spoofing virtual networks

1. **Virtual network security**: Wu and et al. [19] present a virtual network framework that secures the communication among virtual machines. This system depends on Xen which offers two design modes for virtual systems: "spanned and directed." The virtual system model is made out of three layers: steering layers, firewall, and shared systems, which can counteract VMs from sniffing and caricaturing. An assessment of this methodology was not performed when this production was distributed.

5 Conclusion

The cloud services are provided to the customers; however, the customer is not satisfied due the vulnerabilities and threats in distributed cloud environment. The SaaS is prone to the attacks such as denial of service and customer data manipulation, and this paper detailed the solution mechanisms for SAAS such as trace back filters system and Web application scanners, respectively. The IaaS is prone to the attacks such as VM escape and malicious VM creation which has the solution mechanisms such as hypersafe, TCCP, and mirage, respectively. The attacker can also do attacks on PaaS such as account or service hijacking and data leakage having solution mechanisms such as identity and access management guidance, dynamic credential and FRS techniques, digital signatures, homomorphic encryption, respectively. Though there are attack protection mechanisms provided by many people, there is no silver bullet to protect the cloud from the ever-changing technology and intelligent attacker and poor coders

References

1. Mazhar A, Khan SU, Vasilakos AV (2015) Security in cloud computing: Opportunities and challenges Elsevier Trans. Inf Sci 305:357–383
2. Ju J, Wang Y, Fu J, Wu J, Lin Z (2010) Research on key technology in SaaS. In: International conference on intelligent computing and cognitive informatics (ICICCI), Hangzhou, China. IEEE Computer Society, Washington, DC, USA, pp 384–387
3. OWASP (2010) The ten most critical web application security risks. Available https://www.owasp.org/index.php/Category OWASP top ten project
4. Hashizume K, Rosado DG, Fernández-Medina E, Fernandez EB (2013) An analysis of security issues for cloud computing. J Internet Serv Appl 4:5
5. Cloud Security Alliance (2012) Security guidance for critical areas of mobile computing. Available https://downloads.cloudsecurityalliance.org/initiatives/mobile/Mobile_Guidance_v1.pdf
6. Morsy MA, Grundy J, Müller I (2010) An analysis of the cloud computing security problem. In: Proceedings of APSEC 2010 cloud workshop. APSEC, Sydney, Australia
7. Cloud Security Alliance (2010) Top threats to cloud computing V1.0. Available https://cloudsecurityalliance.org/research/top-threats
8. Cloud Security Alliance (2012) SecaaS implementation guidance, category 1: identity and Access management. Available https://downloads.cloudsecurityalliance.org/initiatives/secaas/SecaaS_Cat_1_IAM_Implementation_Guidane.pdf
9. Xiao S, Gong W (2010) Mobility can help: protect user identity with dynamic credential. In: Eleventh International conference on mobile data management (MDM). IEEE Computer Society, Washington, DC, USA, pp 378–380
10. Wylie J, Bakkaloglu M, Pandurangan V, Bigrigg M, Oguz S, Tew K, Williams C, Ganger G, Khosla P (2001) Selecting the right data distribution scheme for a survivable storage system. CMU-CS-01-120, Pittsburgh, PA
11. Somani U, Lakhani K, Mundra M (2010) Implementing digital signature with RSA encryption algorithm to enhance the data security of cloud in cloud computing. In: 1st international conference on parallel distributed and grid computing (PDGC). IEEE Computer Society Washington, DC, USA, pp 211–216
12. Tang J, Cui Y, Li Q, Ren K, Liu J, Buyya R (2016) Ensuring security and privacy preservation for cloud data services. ACM Comput Surv 49(1):39
13. Fong E, Okun V (2007) Web application scanners: definitions and functions. In: Proceedings of the 40th annual Hawaii Inter.national conference on system sciences. IEEE Computer Society, Washington, DC, USA
14. Wang Z, Jiang X (2010) HyperSafe: a lightweight approach to provide lifetime hypervisor control-flow integrity. In: Proceedings of the IEEE symposium on security and privacy. IEEE Computer Society, Washington, DC, USA, pp 380–395

15. Santos N, Gummadi KP, Rodrigues R (2009) Towards trusted cloud computing. In: Proceedings of the 2009 conference on hot topics in cloud computing. San Diego, California. USENIX Association Berkeley, CA, USA
16. Wei J, Zhang X, Ammons G, Bala V, Ning P (2009) Managing security of virtual machine images in a cloud environment. In: Proceedings of the 2009 ACM workshop on cloud computing security. ACM New York, NY, USA, pp 91–96
17. Zhang F, Huang Y, Wang H, Chen H, Zang B (2008) PALM: security preserving VM live migration for systems with VMM-enforced protection. In: Trusted infrastructure technologies conference, 2008. APTC'08, Third Asia-Pacific. IEEE Computer Society, Washington, DC, USA, pp 9–18
18. Xiaopeng G, Sumei W, Xianqin C (2010) VNSS: a network security sandbox for virtual computing environment. In: IEEE youth conference on information computing and telecommunications (YC-ICT). IEEE Computer Society, Washington DC, USA, pp 395–398
19. Wu H, Ding Y, Winer C ,Yao L (2010) Network security for virtual machine in cloud computing. In: 5th International conference on computer sciences and convergence information technology (ICCIT). IEEE Computer Society Washington, DC, USA, pp 18–25

Categorization of Plant Leaf Using CNN

Suhasini Parvatikar and Deepa Parasar

Abstract As plants are main parts of environment, there are variety of trees in this environment, and it is unable to recognize different kinds of plant leaf. Thus, this paper studies about plant leaf detection using CNN method. This method will help us to classify leaf, which would indirectly help us to classify different kinds of plants. Combination of deep neural networks forms a convolutional neural network (CNN), which is used for recognition of plant leaf. In neural networks, convolutional neural network is mainly used for recognition of images, classifying different types of images, face recognition, etc. CNN image takes image as an input, processes that image, and classifies it under different type of category.

Keywords Neural network · CNN · Classification

1 Introduction

There are billions of plants in the environment. Imagination of life on earth without plants is not possible; they are our source of oxygen and also balance our ecosystem. It is very difficult for a scientist, botanist, or expert to recognize a leaf of a plant from such a large volume of plants. For this, we require some effective classification algorithms. Machine and deep learning technique could help to solve problems by recognizing the leaf of a particular plant as deep learning technique performs feature extraction process automatically.

We could use deep learning method for simplifying our process. Thus, as we are taking image as an input, we propose a method called convolutional neural network (CNN) for identifying leaf of a plant.

S. Parvatikar (✉)
Mumbai University, Kharghar, Navi Mumbai, India
e-mail: er.suha@gmail.com

D. Parasar
Amity University, Panvel, India
e-mail: dparasar@mum.amity.edu

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_7

2 Literature Survey

Liu et al. [1] designed a system for the classification of the plant. The following steps are carried out: the images are preprocessed before giving it to classifier. Using the classification algorithm support vector machine (SVM), type of leaf is classified. Next, the parameters of the SVM design are optimized.

Padao [2], authors here, used naïve Bayes classification algorithm for identifying plant leaf. Texture and shape of leaf are given as input to the classifier. The output of naïve Bayes algorithm shows high accuracy of the model.

Kumar et al. [3], is a survey paper. The researcher contrasts different classification strategies such as k-nearest neighbor classifier, probabilistic neural network, genetic algorithm, support vector machine, and principal component analysis. The aim of this survey is to provide knowledge of different techniques of classification which can be used to classify the leaf of a plant.

Kadir et al. [4], designed a system using principal component analysis (PCA) which is the technique for increasing variable space dimensionality by describing uncorrelated variable space. The goal of authors is to improve grade of leaf identification system. Shape, color, and texture of leaf are given as input to system. Using the identification system, the PCA algorithm was used to transform the features into orthogonal features, and the results were then given to the classifier using probabilistic neural network (PNN). PCA's output increases the accuracy on datasets of the leaf identification system.

Buttrey and Karo [5] designed a system using a hybrid (composite) classifier. The authors have used two classifiers by using k-nearest neighbor (kNN) and classification trees. The inputs are given first to classifier to extract features and then classify them using k-NN which reduced processing load associated with k-NN, and the result shown were better than individual algorithm.

3 Proposed Methodology

The proposed system uses CNN to classify the leaf. This application uses the specific features of each leaf which are computed and then further used for classification of them (Fig. 1).

4 Convolution

A CNN can be implemented [6] as a feedforward neural network, wherein only a few weights are active (in color) and the rest of the weights (in black) are zero. The working of CNN and information flow is as shown in Fig. 2.

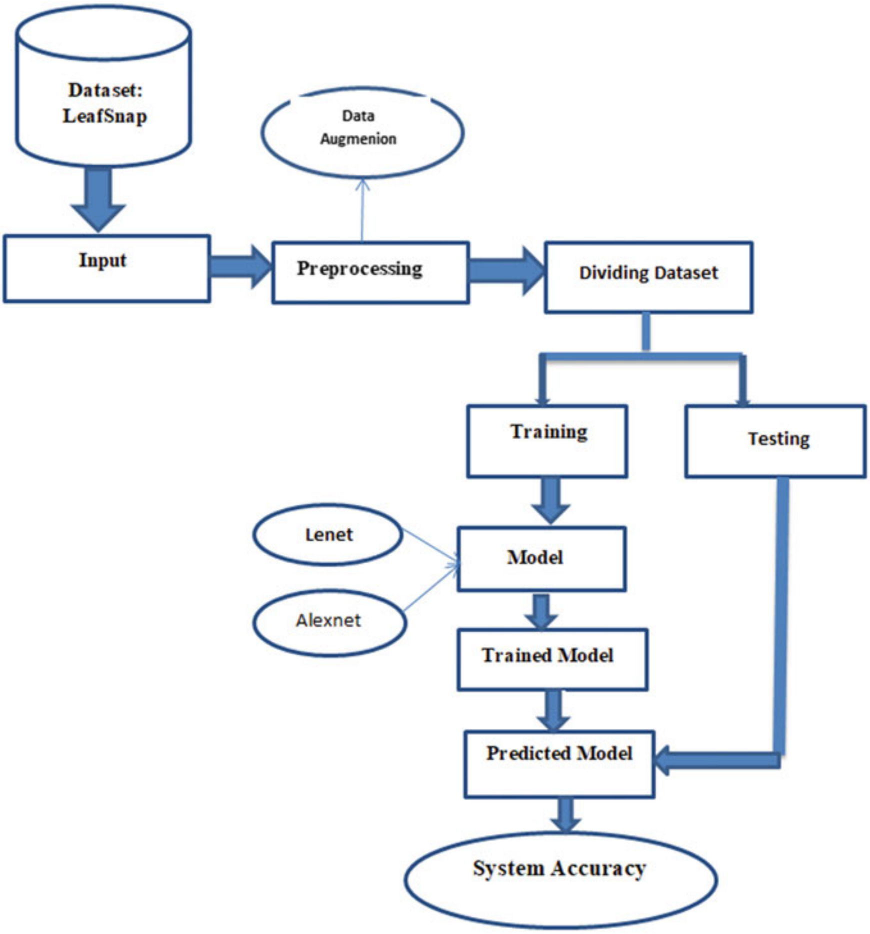


Fig. 1 Flow diagram of proposed system

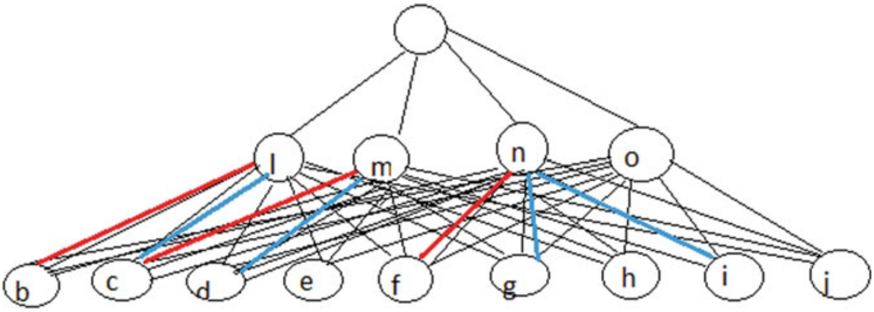


Fig. 2 Fully connected network

5 Convolutional Neural Network

The demand for CNN has recently increased as CNNs remove the need for manual extraction of features---the features are learned directly from CNN. Convolutional neural networks work as neural network in which each neuron receives some input. The only difference is that the input provided to a CNN is in the form of image or a time series data on which each neuron performs operations and generates an output.

Input Layer ≫ Hidden Layers ≫ Output.

Here, input layer takes the input, and output gives the desired output that means hidden layer is doing the feature extractions work as shown in Fig. 3 [7].

The steps in CNN can be summarized [8] as:

a. Input image is given to convolutional layer.
b. Filters/feature maps are applied to the image, which gives us a convolutional layer.
c. Apply rectifier function.
d. Perform pooling to reduce dimensionality size.
e. We then flatten our pooled feature map before inserting into an artificial neural network, i.e., fully connected network.
f. Output the class using an activation function and classify images.

Step 1: Convolution: The input image is multiplied with every element by filter matrix to produce a output to reduce the size of input image. Only main features that are important are considered in this step.

Eg: 7 × 7 image matrix convolution is multiplied by 3 × 3 filter matrix which gives output called "Feature Map" as shown in Fig. 4 [9].

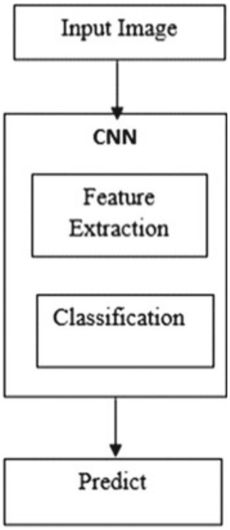
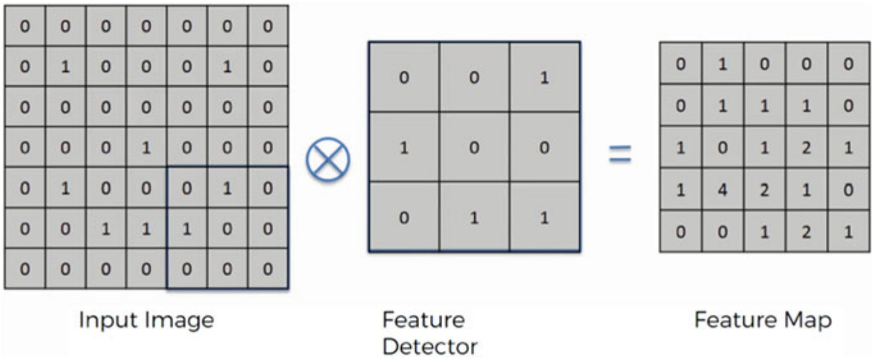


Fig. 3 Working process of CNN



$$\begin{bmatrix} 0&0&0&0&0&0&0\\ 0&1&0&0&0&1&0\\ 0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0\\ 0&1&0&0&0&1&0\\ 0&0&1&1&1&0&0\\ 0&0&0&0&0&0&0 \end{bmatrix} \otimes \begin{bmatrix} 0&0&1\\ 1&0&0\\ 0&1&1 \end{bmatrix} = \begin{bmatrix} 0&1&0&0&0\\ 0&1&1&1&0\\ 1&0&1&2&1\\ 1&4&2&1&0\\ 0&0&1&2&1 \end{bmatrix}$$

Fig. 4 Convolution process

Step 2: Apply rectifier function: In NN, we use activation function to transform summed weighed input into output. RELU is a nonlinear activation function which allows complex relationship in data to be learned**.**

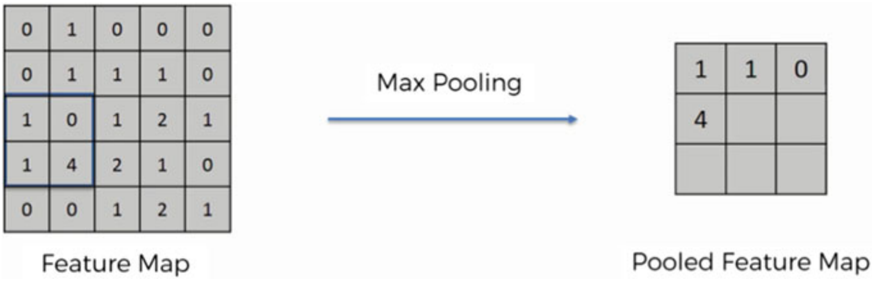
This function returns 0 for negative input and x for positive value. It returns that value back. So, it can be given as:

$$f(x) = \max(0,\ x) \tag{1}$$

Step 3: Pooling: Its function is to gradually reduce dimension size which will help in network computation. The pooling layer operates independently on each feature map. It takes a function map of each matrix of 2 × 2 and selects the value present in that box. Then, 2 × 2 matrix is passed through the entire feature map choosing the value in each pass from left to right. These values then form a pooled map of features a new matrix.

Different categories [9] of pooling:

- Max pooling: Largest value from 2 × 2 matrix is picked
- Average pooling: Mean value from 2 × 2 matrix is picked (Fig. 5).



$$\begin{bmatrix} 0&1&0&0&0\\ 0&1&1&1&0\\ 1&0&1&2&1\\ 1&4&2&1&0\\ 0&0&1&2&1 \end{bmatrix} \xrightarrow{\text{Max Pooling}} \begin{bmatrix} 1&1&0\\ 4&&\\ && \end{bmatrix}$$

Fig. 5 Pooling

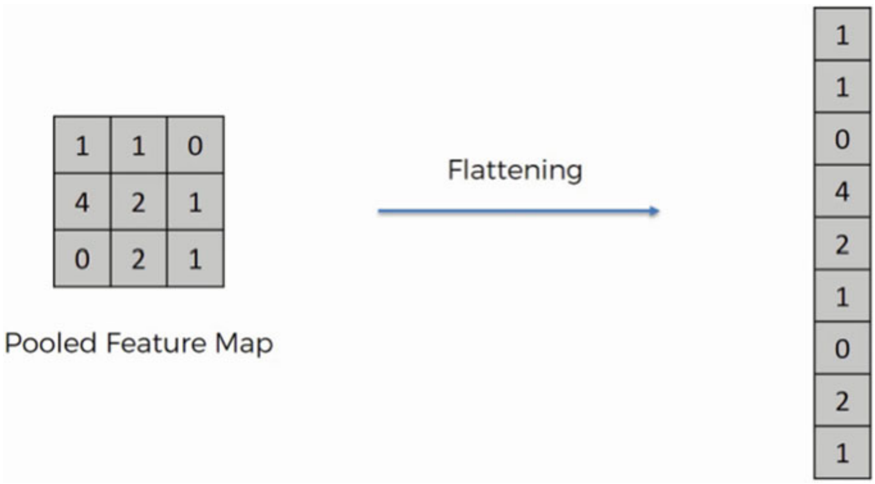


Fig. 6 Flattening

Step 4: Flattening: Flattening [10] transforms the entire matrix of previous step into a single column. It is then given as input to neural network (Fig. 6).

Step 5: Fully connected network: Fully connected layers [10] are an essential component of convolutional neural networks (CNNs), which have been proven very successful in recognizing and classifying images for computer vision. The data will be transmitted through the network, and the prediction error will be measured. It is done by the using the function as shown in Eq. (2):

$$f_j(z)_j = \frac{\mathrm{e}^{z_j}}{\sum_{k=1} \mathrm{e}^{z_k}}. \tag{2}$$

The error is then propagated back to boost the prediction through the process. It is important that output produced by output layer must be scaled between 0 and 1 which represents probability of each class.

6 Dataset

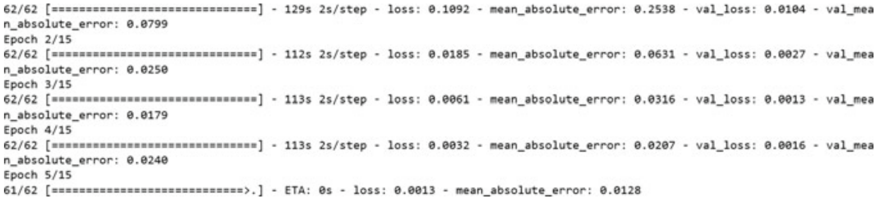
The LeafSnap dataset consists of images of leaves taken from different sources: A total of 23,147 laboratory images consist of high-quality images taken of pressed leaves. 7719 images consist of images captured by mobile devices. The dataset consists of 185 species of plants collected from northeastern US.

7 Results

The system trains the neural network using the training set and evaluates its performance on the test set (Fig. 7).

Our project is partially done, and now currently system only classifies using ALEXNET model.

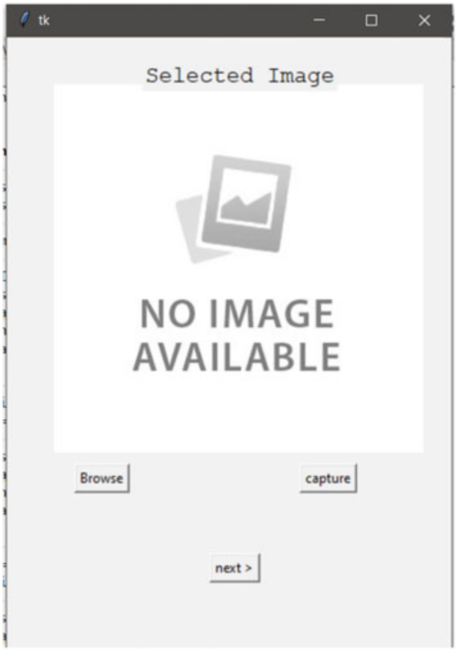
The following steps are carried out in classifying leaf of a plant (Figs. 8, 9, 10, and 11).

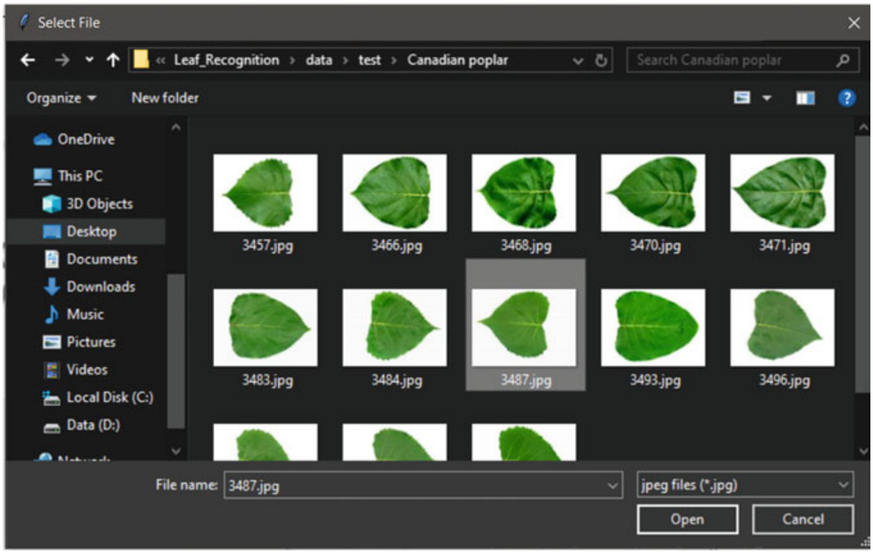


```
62/62 [==============================] - 129s 2s/step - loss: 0.1092 - mean_absolute_error: 0.2538 - val_loss: 0.0104 - val_mea
n_absolute_error: 0.0799
Epoch 2/15
62/62 [==============================] - 112s 2s/step - loss: 0.0185 - mean_absolute_error: 0.0631 - val_loss: 0.0027 - val_mea
n_absolute_error: 0.0250
Epoch 3/15
62/62 [==============================] - 113s 2s/step - loss: 0.0061 - mean_absolute_error: 0.0316 - val_loss: 0.0013 - val_mea
n_absolute_error: 0.0179
Epoch 4/15
62/62 [==============================] - 113s 2s/step - loss: 0.0032 - mean_absolute_error: 0.0207 - val_loss: 0.0016 - val_mea
n_absolute_error: 0.0240
Epoch 5/15
61/62 [============================>.] - ETA: 0s - loss: 0.0013 - mean_absolute_error: 0.0128
```

Fig. 7 Accuracy rate

Fig. 8 Click browse or capture to capture leaf image



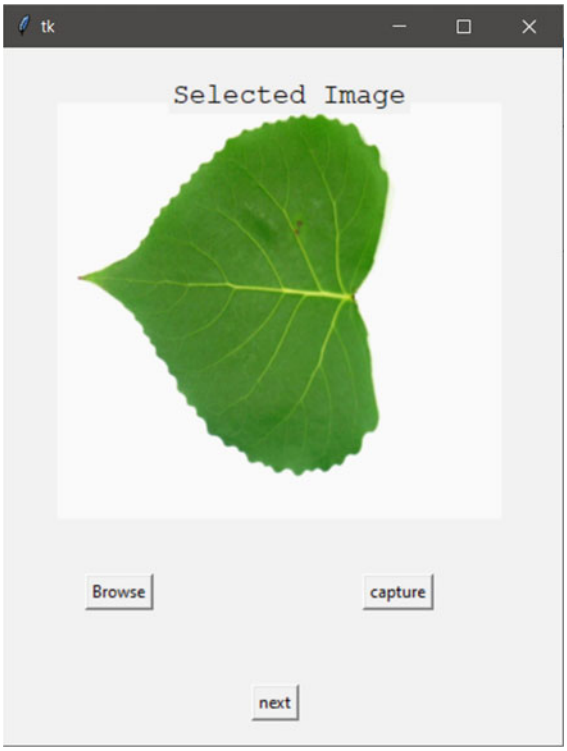


File name: 3487.jpg

jpeg files (*.jpg)

Fig. 9 Select image to make prediction

Fig. 10 Click next button



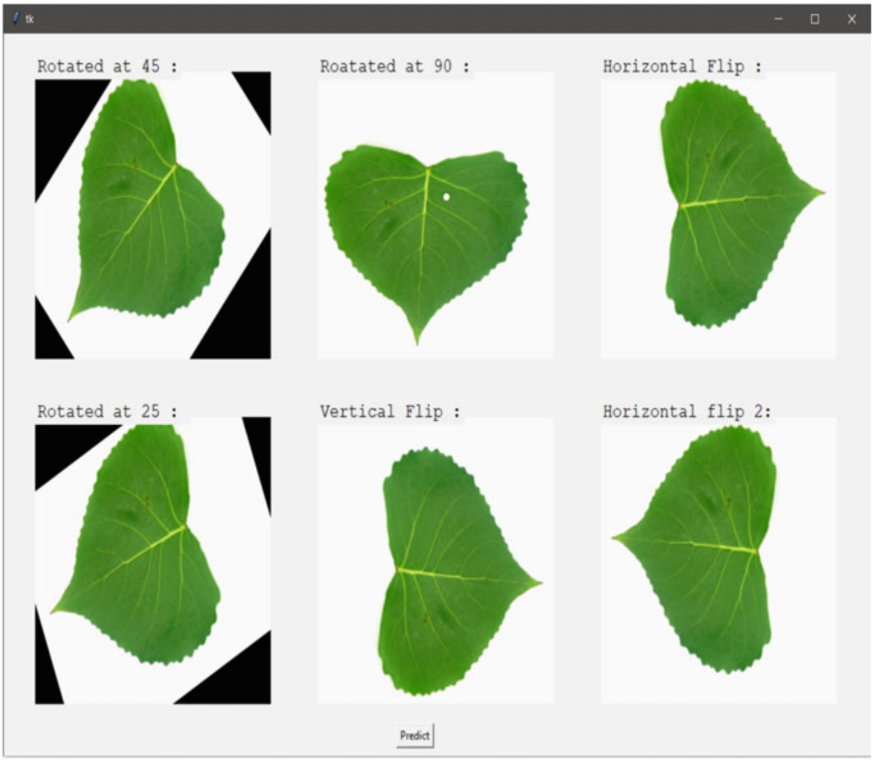
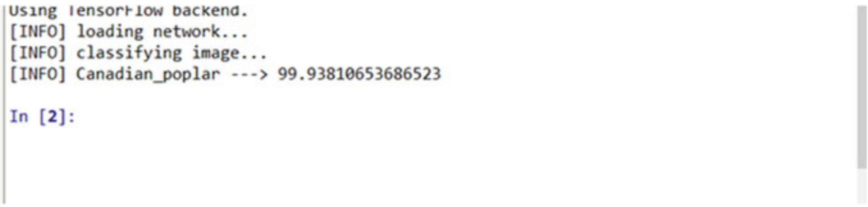


Fig. 11 Data augmentation preview

8 Testing Results

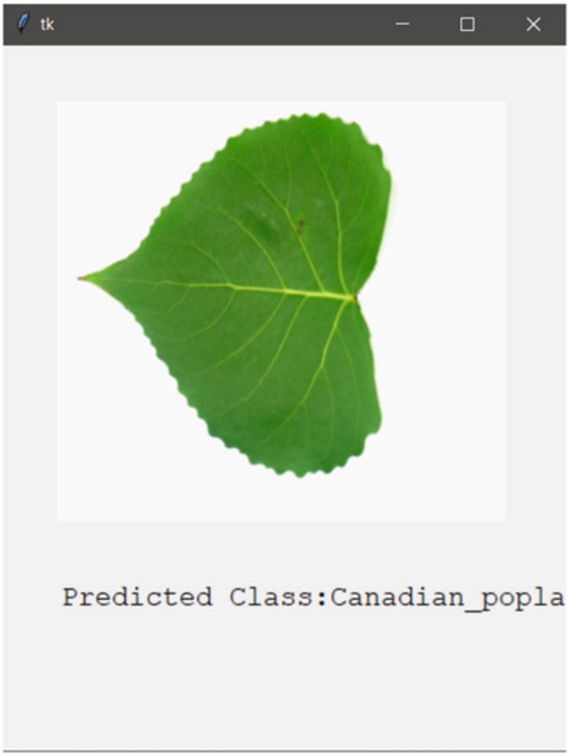
See Figs. 12 and 13.



```
Using TensorFlow backend.
[INFO] loading network...
[INFO] classifying image...
[INFO] Canadian_poplar ---> 99.93810653686523

In [2]:
```

Fig. 12 Executing file



```
Predicted Class:Canadian_popla
```

Fig. 13 Predicted class of leaf

9 Conclusion

In this paper, we studied and are trying to implement a new method to classify leaves using the CNN model using two architectures of CNN like AlexNET, LeNet, etc., with little difference in architecture, and finally, the output needs to classify leaf type using architecture with greater accuracy rate.

CNN is mostly used for input images. Thus, in fully connected network, it accepts weights from each node, while in convolutional layer each neuron is only connected to a nearby neurons only which helps to extract only relevant information.

Reference

1. Liu W, Li Z, Lin J, Liang D (2017) The PSO-SVM-based method of the recognition of plant leaves. In: 2017 IEEE international symposium on parallel and distributed processing with applications and 2017 IEEE international conference on ubiquitous computing and communications (ISPA/IUCC). Haikou, China, 1350–1355

2. Padao FRF, Maravillas EA (2015) Using Naïve Bayesian method for plant leaf classification based on shape and texture features. In: 8th IEEE international conference humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM). The Institute of Electrical and Electronics Engineers Inc. (IEEE)—Philippine Section, Water Front Hotel, Cebu, Philippines, 9–12 Dec 2015
3. Kumar M, Kamble M, Pawar S, Patil P, Bonde N (2013) Survey on techniques for plant leaf classification. Int J Modern Eng Res IJMER 1(2):538–544. ISSN: 2249-6645
4. Kadir A, Nugroho LE, Susanto A, Santosa PI (2012) Performance improvement of leaf identification system using principal component analysis. Int J Adv Sci Technol 44
5. Buttrey SE, Karo C (2002) Using k-nearest-neighbor classification in the leaves of a tree. Comput Stat Data Anal 40:27–37. www.elsevier.com/locate/csda
6. Mitesh Khapra nptel videos https://www.youtube.com/watch?v=PmZp5VtMwLE
7. Jeon W-S, Rhee S-Y (2017) Plant Leaf Recognition Using a Convolution Neural Network. Int J Fuzzy Logic Intell Syst 17(1):26–34
8. https://www.superdatascience.com/convolutional-neural-networks-cnn-softmax-cross-entropy/
9. https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148
10. https://heartbeat.fritz.ai/a-beginners-guide-to-convolutional-neural-networks-cnn-cf26c5ee17ed

Harvest Treasure: Prediction of Best Crop Yield

Gresha S. Bhatia, Simran Bhagwandasani, Rahul Bhatia, Urjita Bedekar, and Pranit Naik

Abstract Agriculture is one of the most critical and essential occupations practiced in our country. It is an economic sector that plays an essential role in the overall development of the country. Thus, the modernization of agriculture is significant and thus will lead the farmers of our country toward profit. Earlier, the sowing of crops was performed by considering the farmer's knowledge in a particular field and about a specific crop. However, as the weather conditions change very rapidly, farmers cultivate more and more crops that do not give an expected yield, thereby reducing their profits. Being this as the current situation, many of them do not have enough knowledge about the new crops and are not entirely aware of the benefits they get while farming them. Also, farm productivity can be increased by understanding and forecasting crop performance in a variety of environmental conditions. The proposed system applies machine learning and prediction algorithms to identify the pattern among data and then process it as per input conditions. This in turn will propose the best feasible crops according to given environmental conditions. Thus, this system will only require the land area of the user, and it will suggest a number of profitable crops providing a choice directly to the farmer about which crop to cultivate. As past year production is also taken into account, the prediction will be more precise.

Keywords Machine learning · Yield prediction · Indian agriculture

1 Introduction

Agriculture in India has a full-size history. More than 60% of the land in the country is used for agriculture to cater to the needs of 1.2 billion people. India is ranked second worldwide in farm output. Agriculture and other sectors like forestry and fisheries accounted for 16.6% of GDP in 2009 [1, 2]. The production of crops relies on different factors like climatic and geographical. Accurate statistics about the character of an ancient yield of the crop is a critical modeling input, which is useful

G. S. Bhatia (✉) · S. Bhagwandasani · R. Bhatia · U. Bedekar · P. Naik
Vivekanand Education Society's Institute of Technology, Mumbai, India
e-mail: gresha.bhatia@ves.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_8

to farmers and authorities organizations for decision making regarding the crop to plant. The project has been to extract expertise from these statistics mining that may be able to bridge the understanding of the facts to the crop yield estimation. This task is aimed at statistics mining strategies and follows them to the various variables consisting inside the database to set up if significant relationships may be discovered and the usage of fuzzy common sense to discover the circumstance of crops on a diverse situation of climatic conditions.

2 Literature Survey

A literature survey is the most crucial part which escorted us to our research. After having a look at the previous papers aiming the same objectives which we have and based on an intensive study, we plan to improve the limitations, challenges faced, and various drawbacks analyzed from those papers. After referring to the work done by Teeda et al. [3], we have understood that climatic analysis based on location, rainfall, soil type, yield capacity, and irrigation techniques is a fundamental prerequisite that needs to be carefully implemented to plan better farming structures for a satisfactory yield. Farming risks have significantly reduced with the advent of technologies. Empirical approaches like regression, artificial neural network, fuzzy logic, and institution approach of statistics have been employed to predict the weather parameters. For the prediction of rainfall, clustering and classification methods of data mining have been used. The work done by Zingade et al. [4] is spectacularly presented. They have considered significant parameters like soil, water, and temperature. For the soil measure, they have considered nitrogen, phosphorus, and potassium, along with the pH value of the soil. Taking zinc and sulfur nitrate into consideration, along with these, can help to improvise the system. Moreover, we have chosen to perform classification using a random forest classifier which overfits the linear regression used above, since the parameters are quite significant in number. After the analysis on ‘The Farmer’s Handbook’ published by Desai Fruits, [5], we have an understanding of choosing the sufficient parameters required for selecting the type of crop suitable for cultivation. Project by Priya, published in the International Journal of Engineering Science and Research Technology, has performed the prediction of crop yield focusing only on Tamil Nadu district [6]. We intend to build a system that satisfies the majority of places of possible farming in India. The system proposed by Meeradevi and Mundada [7] uses wireless sensor network for precision in detecting parameters for prediction of crop yield. The system uses sensors for estimating parameters such as pH or humidity. The sensed data is further used for analysis. Main advantage of this system is that it is portable and can be used for real-time prediction. The system proposed by Kumar et al. [8] estimates crop yield for sugarcane crop using least squares support vector machine but gives an average accuracy of 90% approximately which can be still improved. However, the system is domain independent and can be used across various disciplines.

3 Challenges in Existing Systems

1. **IoT-based plants**: For constant analysis of the fields and crop growth, many projects have obtained to install IoT devices for data gathering, which are way costlier and require regular maintenance [9].
2. **Web interface**: Many projects have their UI based on Python which is hosted on the web server, our mobile application is very much user-friendly, and we intend to bring the information available in the local languages which are region-specific [10].
3. **Parameters under consideration**: The existing projects have a limited number of parameters considered for training. We intend to consider the majority of parameters which are accountable for the plants' growth.

4 Proposed Problem Statement

Accurate yield estimation is an essential parameter to be considered for farmers to sow the crops accordingly. Machine learning systems are being abundantly used nowadays in order to build decision support tools for the contemporary farming system to improve yield production while reducing operational cost and environmental impact to a great extent. The capability of machine learning algorithm to access tons of data and mold them into relatable format helps in predicting the future results when a particular test case is provided to the trained system; thereby, it would provide a great support in predicting the estimated yield of a particular crop provided with sowing area, time, and climatic conditions of that particular crop. Our system proposes to use an algorithm that will work accurately, even when the number of parameters under consideration is more. This system detects the location of the user. From the location, the nutrients of the soil, such as nitrogen, phosphorus, and potassium, are obtained. We have also considered parameters like solar radiation and precipitation, which many other systems fail to consider. The processing part also takes into consideration two more datasets, i.e., one obtained from the weather department, forecasting the weather expected in the current year, and one from National Centers for Environmental Prediction (NCEP) [11]. This data is related to the demands of various crops obtained from various government Web sites. The proposed system applies various algorithms to identify the pattern among data and then process it as per input conditions, which will propose the best suitable crops according to the given environmental conditions. Thus, this system will require the area of the field, and it will suggest several profitable crops providing a choice directly to the farmer about which crop to cultivate. Since historical data about several parameters like solar radiation, minimum and maximum temperature, and precipitation are taken into account, the output is expected to be much more accurate.

5 Proposed System

Since other systems developed are using data mining techniques to predict the optimal crop yield, our system aims to provide an accurate yet effective result based on more number of parameters dependent on climatic conditions to provide maximum profit, reduced risks, and more than expected yield to the farmers. The proposed solution is represented in Fig. 1 and is described as follows:

Input: The prediction of the crop is dependent on various factors such as weather and past crop production to predict the crop correctly. The location of the user is taken as an input, which will help us to acquire information about weather conditions like humidity and temperature.

Data acquisition: The system mines the weather conditions like humidity, minimum and maximum temperature, and amount of solar radiations depending on location in the respective area from the other datasets. Similarly, crop yield is extracted from historical data. This would give an excellent idea about the demands of crops in the location.

Output: The random forest algorithm creates a forest with several trees. The most suitable crop is predicted by the system using the random forest algorithm, and the user is provided with the best-suited crop for a given set of climatic conditions.

The block diagram represents the basic working of the system, wherein an accumulated dataset is trained based on random forest algorithm and fed to the model. Simultaneously, the farmer inputs the area of the land required for sowing, and the location and weather data are automatically detected using weather APIs and fed

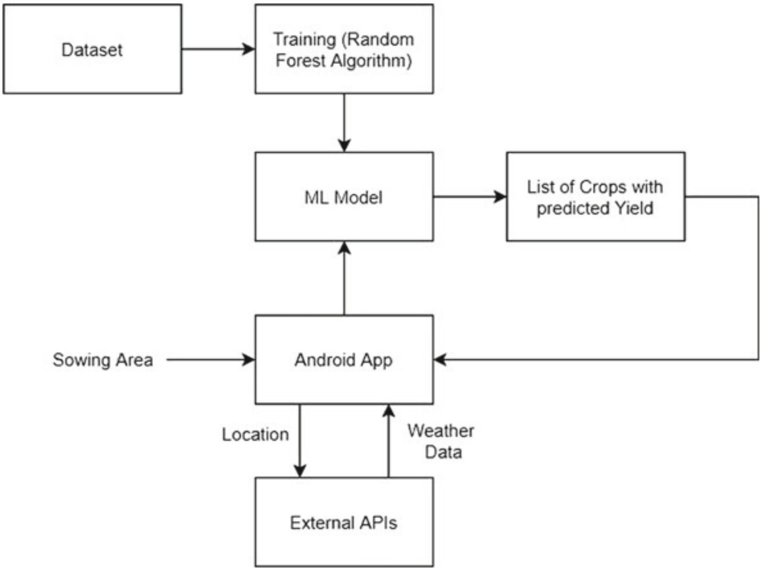


Fig. 1 Block diagram of harvest treasure

to the Android app which is then connected to the model which results in the final output of list of crops with the yield predicted for the fed sowing area, and the result is displayed on the Android app.

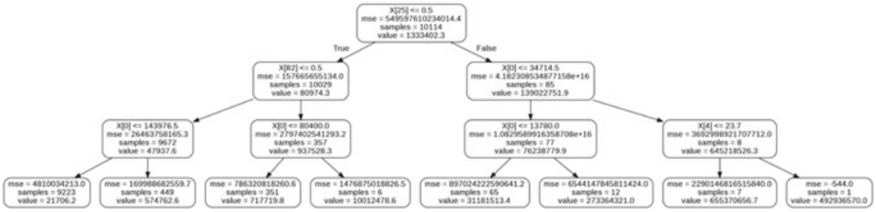
The sowing area will be entered by the user as input. Our classification algorithm will be analyzing climatic conditions such as rainfall based on the location and sowing time. The algorithm will then suggest the best crop to plant for getting the maximum yield, and this can be used by farmers to acquire maximum profit as observed in Fig. 2.

The farmer provides location access to the Android application and also inputs such as the sowing area. The Android application then suggests most suitable crops along with their expected yield (Fig. 3).

6 Implementation

To predict the best crop yield, we have considered various parameters such as area, production, rainfall, elevation, max. temperature, min. temperature, precipitation, relative humidity, and solar. Since the data for the majority of parameters was in the form of float and string, we have performed encoding for the data. Using one hot encoding, we have converted categorical parameters to their corresponding numeric value.

After this, we have divided our dataset into training and testing parts with keeping the test size of 30 per cent from the entire dataset. On the training data, we have then performed various regression algorithms such as linear regression, linear regression with L2 regularization, polynomial regression, gradient boosted regression, nearest neighbor regression, support vector machine (radial basis function kernel and linear kernel), and few more. We are analyzing the accuracy of all these regression models, and the random forest has a maximum accuracy ranging between 92 and 96%. To train the random forest model, we performed five-level splits and have used 200 n_estimators.



Here is a sample of a small tree with step size of three which is randomly selected from the 200 trees.

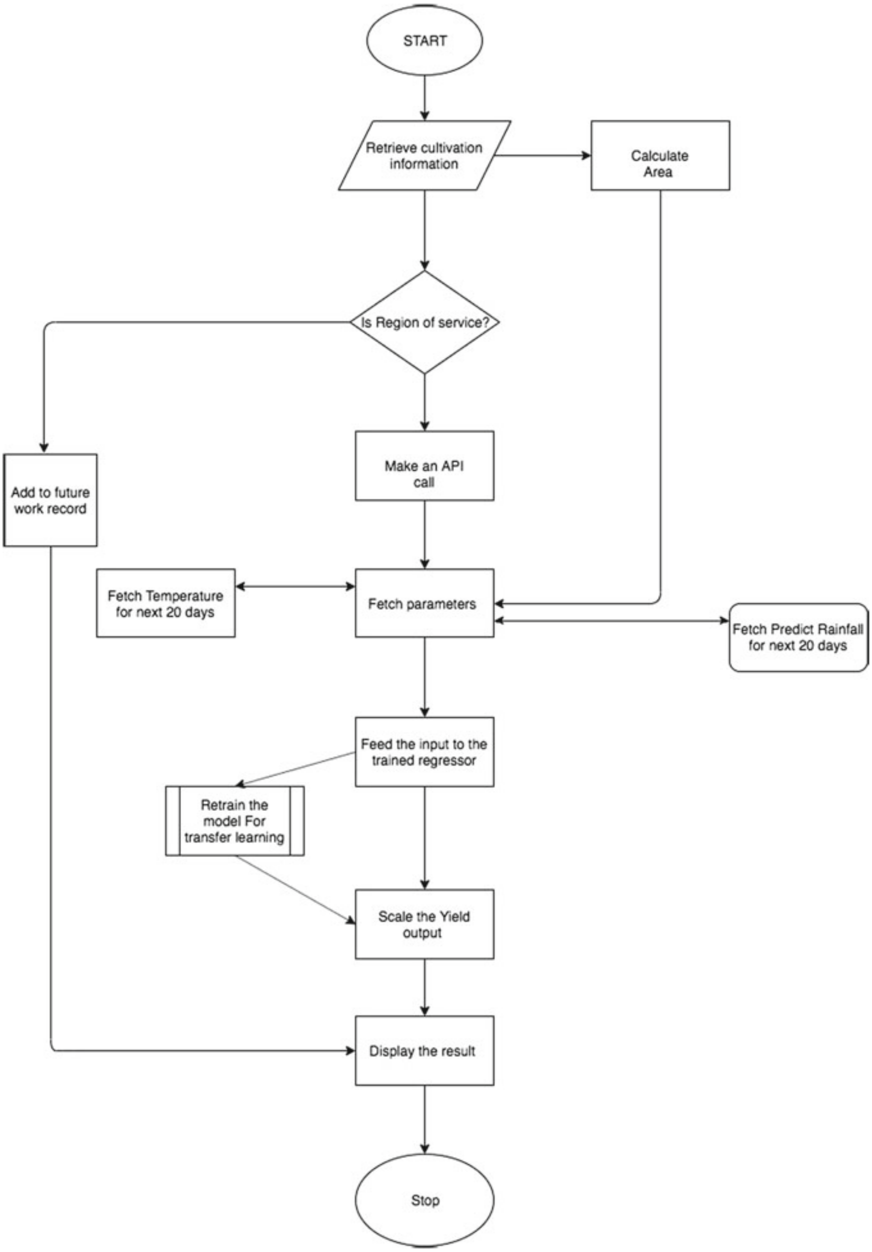

START
Retrieve cultivation information
Calculate Area
Is Region of service?
Add to future work record
Make an API call
Fetch Temperature for next 20 days
Fetch parameters
Fetch Predict Rainfall for next 20 days
Feed the input to the trained regressor
Retrain the model For transfer learning
Scale the Yield output
Display the result
Stop

Fig. 2 Flowchart of harvest treasure

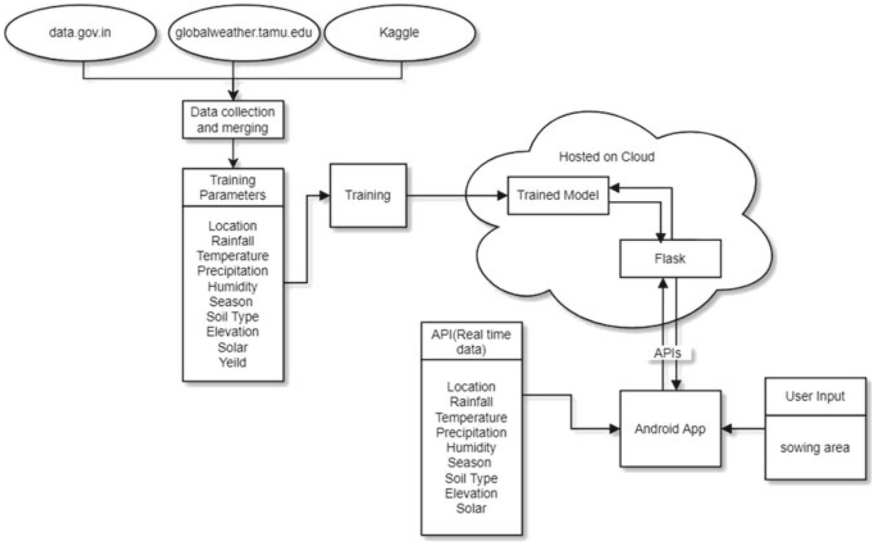


Fig. 3 Methodology employed

7 Comparative Analysis of Various Algorithms Tested

Accuracy: The term accuracy is defined as the degree to which the result of a specification, calculation, or measurement conforms to the correct value or a standard [12].

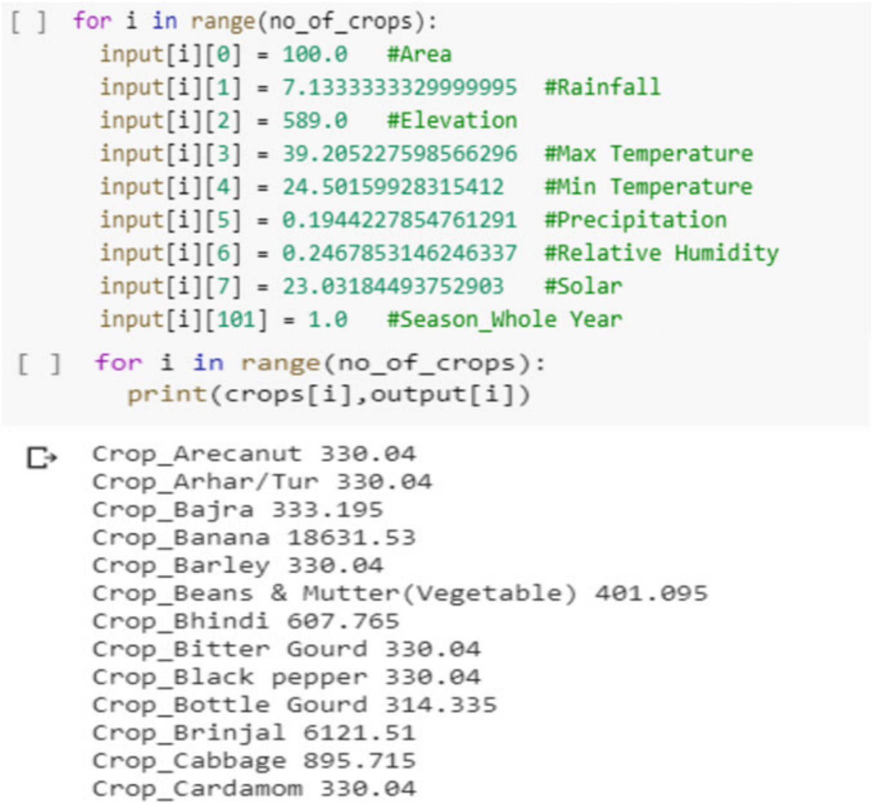
Error rate: The term error rate refers to the frequency of errors occurring, defined as "the ratio of total number of data units in error to the total number of data units transmitted" [13] (Table 1).

Table 1 Comparison of evaluation parameters for various algorithms tested

Algorithm	Accuracy	Error rate
Hidden layers: 2	0.0079	0.9921
Nodes in each layer: 104, 256,		
128.1		
Activation: 'relu'		
Linear regression	0.21	0.79
Linear regression with L2 regularization	0.16	0.84
Polynomial regression	0.45	0.55
Random forest regression	0.95	0.05
Gradient boosted regression	−0.017	1.823

8 Outputs Obtained

The parameters that we have selected to evaluate our proposed system include accuracy and error rate. Since the system is being evaluated on parameters like location, rainfall, season, and climate, the quality of being precise is about 95.4% for the system developed, having an error rate of 5.12% in a dataset of 200 trees.



```
[ ]  for i in range(no_of_crops):
       input[i][0] = 100.0   #Area
       input[i][1] = 7.1333333329999995  #Rainfall
       input[i][2] = 589.0   #Elevation
       input[i][3] = 39.205227598566296  #Max Temperature
       input[i][4] = 24.50159928315412   #Min Temperature
       input[i][5] = 0.1944227854761291  #Precipitation
       input[i][6] = 0.2467853146246337  #Relative Humidity
       input[i][7] = 23.03184493752903   #Solar
       input[i][101] = 1.0   #Season_Whole Year

[ ]  for i in range(no_of_crops):
       print(crops[i],output[i])
```

```
Crop_Arecanut 330.04
Crop_Arhar/Tur 330.04
Crop_Bajra 333.195
Crop_Banana 18631.53
Crop_Barley 330.04
Crop_Beans & Mutter(Vegetable) 401.095
Crop_Bhindi 607.765
Crop_Bitter Gourd 330.04
Crop_Black pepper 330.04
Crop_Bottle Gourd 314.335
Crop_Brinjal 6121.51
Crop_Cabbage 895.715
Crop_Cardamom 330.04
```

9 Conclusion and Future Work

The proposed system takes into consideration data related to weather and past year production and suggests the most suitable crops that can be cultivated in environmental conditions of various parts of India. As the system lists out the best-possible crops, this would help the farmer in deciding which crop is appropriate to cultivate. Since this system takes into consideration past production, it will help the farmer get insight into the demand for crops cultivated at the stipulated time in the market. The main aim of this research paper is to provide a methodology, so that it can perform descriptive analytics on crop yield production in an effective manner. The future scope of the project is to collect soil nutrients for every particular piece of land and combine the trained system and data about soil nutrients to get a cumulative output based on all these factors. The system can be expanded to provide regional language support, and the aim is to expand the system to all the remote places of India.

References

1. https://www.expo2015.org/magazine/en/economy/agriculture-remains-central-to-the-world-economy.html
2. https://en.wikipedia.org/wiki/Agriculture_in_India
3. Teeda K, Vallabhaneni N, Sridevi T (2018) Analysis of weather attributes to predict crops for the season using data mining. Int J Pure Appl Mathe 119(12):12515–12521
4. Zingade DS, Omkar B, Nilesh M, Shubham G, Chandan M (2018) Machine learning based crop prediction system using multi-linear regression. 3(2)
5. https://www.manage.gov.in/publications/farmerbook.pdf
6. Priya P, Muthaiah U, Balamurugan M (n.d.) Predicting yield of the crop using machine learning algorithm. Int J Eng Sci Res Technol 7(4):1–7
7. Meeradevi, Monica RM Automated control system for crop yield prediction using machine learning approach. Int J Appl Eng Res 14(2). ISSN 0973–4562
8. Kumar A, Kumar N, Vishal V (2018) Efficient crop yield prediction using various machine learning algorithms. Int Res J Eng Technol 05(06)
9. https://www.iofficecorp.com/blog/cost-of-iot-sensors
10. https://www.quora.com/How-many-Indian-farmers-use-the-internet
11. Data Collection: https://globalweather.tamu.edu/#pubs
12. https://www.researchgate.net/post/How_can_I_calculate_the_accuracy
13. https://classeval.wordpress.com/introduction/basic-evaluation-measures/

A Fuzzy Expert System for Malaria Disease Detection

Swati Jayade, D. T. Ingole, and Manik D. Ingole

Abstract It is found that the malaria disease is a prime and major cause to the human health. The pernicious trappings of malaria stooge to the physical body cannot be making light of it. In this research work, a fuzzy-based expert system for the total handling of malaria disease had granted for providing judgment support platform to the specialist and healthcare researchers in the same endemic province. The proposed and implemented system consists of major components which include the cognitive content, the fuzzification, the inference engine and de-fuzzification for decision making. The fuzzy inference engine developed during this work is that the root sum square. This method is the depiction of inference that was designed and developed to infer the info from the fuzzy-based rules used in this algorithm. Triangular fuzzy membership function was accustomed that shows the degree of attendance of every input specification, and therefore, the de-fuzzification technique employed during this research is that the centre of gravity. This fuzzy-based expert system had been developed to help and to support clinical perception, diagnosis and therefore the expert's proficiency. For validation and empharical analysis, the data of thirty patients with malaria defection was used. The results that were calculated are within the range of that was predefined and predicted by the territory proficient.

Keywords Malaria · Fuzzy logic · Cognitive content · Fuzzy expert system

S. Jayade (✉)
Department of Electronics & Telecommunication, Government Polytechnic, Washim, India
e-mail: swatijayade22@gmail.com

D. T. Ingole · M. D. Ingole
Department of Electronics & Telecommunication, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India
e-mail: dtingole@gmail.com

M. D. Ingole
e-mail: manik.ingole2031964@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_9

1 Introduction

The ruinous accouterments's of malaria parasites to the human being as well as other animal cannot be undervalue. Malaria is a harmful disease that caused principally by plasmodium falciparum. P. falciparum and three alternative malaria bedbugs that taint human (p. vivax, p. ovale, and p. malariae) are impart by many type of species of Anopheles mosquitoes widely in the tropics [1–2]. Malaria disease is that the familiar explanation for mortality within the tropics. Today's nature is one beside expanding approach to intelligent systems. In new time, artificial intelligent methods have undoubtedly been utilized in medical applications and research exercise are potent on doctor systems as reciprocal solution to standard technique for locating solution to medical problems [3, 4]. The evolution of data technology has opened unparalleled convenience in health healthcare delivery system because the demand for intelligent and knowledge-based systems has heightened as contemporary medical practices now a day become highly knowledge-intensive. The diagnosing of regional diseases embroils several steps of hysteria and imprecision [5]. The assignment of health problem diagnosis and handling it is compound due to the various functionality involved. Its difficulty level is high due to tons of imprecision and uncertainties. It is not possible for patients to illustrate indeed how they feel and sense, medical practitionars and nurses cannot understand exactly what they identify or observe, and laboratories finding are dotted with some possible errors caused may be by the carelessness of lab workers or improper functioning of the equipments in lab. Medical workers many not squarely identify how diseases transform the traditional working of the body [4]. Out of these ramifications in practice make long-established quantitative path of study disproportionate. Computer workstation tools help to arrange, stock up and reclaim relevant medicinal information desired by the practitioner in handle with each severe case and suggesting appropriate diagnosis, prognosis, remedial decisions and decision-making procedure [6]. A specialist framework is a savvy PC code that uses the information storehouse of at least one specialists and conclusion methods for issue taking care of [7]. Human specialists take care of issues by utilizing a union of genuine information and thinking capacity. In a specialist framework, proposals two imperative are contained in two separate however related segments: an information base and an induction motor. The information base get ready explicit realities and rules about the subject and the derivation motor set up the thinking capacity that empowers the master framework to shape result. Master frameworks additionally give advantageous apparatuses as UI and elucidation offices. UIs, similarly as with any application, empower individuals to shape doubt, furnish data and connect with the framework. The application and use of fuzzy-based perception to medical review and diagnosis of diseases are reviewed in these literatures [6, 8, 9]. This experimentation work represents a fuzzy expert system for the administration of malaria disease. It was developed by considering clinical perception, medical diagnosis and the domain experts. The main aim of this work is to provide a decision backing stack to scientists working on and other medicinal healthcare practitioners. This system will help medical practitioners for diagnosing and decision making in case of malaria where there is a shortfall of doctors in the benefit of society.

2 Literature Review

The symptomatic choices taken by restorative specialists rely on recognition, experience, aptitude, information, capacity and impression of the therapeutic researcher. As the unpredictability of framework builds, it is difficult to pursue a specific way of finding with no error. Fuzzy-based rationale introduces incredible thinking strategies that can deal with vulnerabilities and imprecision. Numerous frameworks have transformed into noteworthy in the extension and continuance of the human services zone. As of late, much research exertion has been fixed in structuring knowledge frameworks. A few reciprocal works have indicated that tropical sicknesses garbage a significant general wellbeing contest in the tropics [10, 11]. Be that as it may, the deliberate endeavor is constantly been made to control the acceleration and transmission of tropical ailments inside and between neighborhoods. In the work vehicle tried out by [12], it was accounted for that month to month jungle fever rate rates and vector densities were utilized for reconnaissance and adaptable tuning of the ecological administration procedures; which brought about an elevated level of work. Inside 3–5 years, intestinal sickness related casualty, grimness and commonness rates were diminished by 70–95%. In an ongoing report, it was inferred that the jungle fever control program that underline natural administration were profoundly viable in diminishing grimness and mortality [4].

In another examination did by [6], the financial results of jungle fever control arranging were featured. The utilized of master frameworks in therapeutic applications is powerful this can be found in [7, 13]. A decent number of master frameworks have been a development on tropical ailments. In [6], a specialist framework is created on tropical ailments to help paramedical staff during exercise and in the understanding of numerous basic illnesses exhibited at their facilities. In another examination completed by [7], a specialist framework on endemic tropical ailments was developed. The objective of the framework was to be a wellspring of training for specialists, and for therapeutic understudies, on the analysis of a portion of the essential tropical endemic illnesses. In [14], they planned a paper to sponsorship clinical judgment making. The paper centers around the separation between three sorts of clinical choice help devices: for data the board, centering thought, and patient-explicit assessment. There are numerous master frameworks for restorative determination and remedies of some tropical tainting dependent on choices are presented in [15, 9–5, 7]. In [13], authors have explained the seriousness of setting preference for health research. In their paper, the point was to organize inquire about in dismissed tropical illness than prominent tropical maladies. An expository strategy that depends on the prioritization system of the worldwide discussion of wellbeing research was utilized for information examination.

In [10], scientists stress the mindfulness that examination is basic in the battle against tropical illness. In any case, the tight assets empty can support just a small amount of the promising exploration space. Thus, prioritization is fundamental for wellbeing research and observable exercise has gone into creating a powerful prioritization framework [10, 4, 6]. Tropical sickness look into special program for research and guidance in tropical diseases, was made to address the requirement for an examination into underestimated tropical malady that speaks to significant general medical issues in advancing nations [7]. In [9], researchers broke down and assessed decision support systems and expert systems which are utilized as help for clinical choices just as potential outcomes to redesign the osmosis and empower the more prominent compromise of these frameworks in the present arrangement of the human services. These creators went to the accompanying results: instruments that are utilized in the clinical choice frameworks are classified as per specialized accessibility, and in that way, their utilization is tight and particularly their absorption with new frameworks for quality medicinal services. Along these lines, the advancement of new frameworks for clinical choice help is huge. Creators in [16] built up a structure for the use of information innovation to the administration of tropical infections. The point of the framework was to sponsorship restorative master in the difficult and burdensome errand of diagnosing and giving treatment to tropical illnesses. The framework gave a structure that will help the therapeutic workforce in provincial regions, where there are a shortage of specialists, during the time spent commitment to essential human services to the individuals. In a similar perspective, [2] built up a fuzzy-based master framework approach utilizing various specialists for gainful follow-up of endemic infections.

In the design of LEPDIAG, the significant highlights that were been itemized are a various master condition, a homeostatic capable containing the model of safe return, a presentation evaluator that can relate the watched signs and indication with those foresee by the homeostatic master and a prognostic master which improves the administration agenda for the patients. In [9], they investigate the issues of undertreatment and uncertainty condition of acute asthma cases in ED. A novel approach, known as the fuzzy logic principle, is employed to determine the severity of acute asthma. The fuzzy set theory, known as Fuzzy Rule-based Expert System for Asthma Severity (FRESAS) determination is embedded into the expert system (ES) to assess the severity of asthma among patients in ED. A medical expert system for managing tropical diseases was suggested by [17]. The proposed Medical Expert Solution (MES) system was to assist medical doctors to diagnose syndrome related to a given tropical disease, suggests the likely ailment, and advances possible treatment based on the MES diagnosis.

The objective of the research was to apply the concept of fuzzy logic technology to determine the degree of harshness on tropical diseases. The root sum square of drawing belief was employed to ascertain the data from the rules developed.

3 Fuzzy Logic

As stated by authors in [8], fuzzy-based rationale is settled as a lot of numerical standards for information portrayal dependent on degrees of investment as opposed to on fresh enrollment of traditional paired rationale. This amazing asset to accessory imprecision and vulnerability was at first acquainted by [8] with improved agreement, durability and ease answers for genuine issues. Fuzzy-based sets have been applied in numerous fields in which inner conflict assumes a key job. Restorative determination is a praiseworthy case of equivocalness and vulnerability. Fuzzy-based set hypothesis is an input to the interest for thoughts and approaches for dealing with non-measurable vulnerability. A fuzzy-based set is a set with fuzzy edge. Characterized fuzzy-based sets or classes for every factor permits transitional evaluations of participation in them, which implies each set could have components that live mostly to it; the level of having a place is called enrollment functions extending from 0 to 1. If X is the universe of talk and its components are de-noted as X, interestingly with fresh set, at that point the fuzzy-based set An of X has attributes work related to it. The fuzzy-based set is communicated by a participation work, defined as seek after:

μ_A: $X \rightarrow$ [0,1]

$\mu_A(X) = 1$ if X is totally in A.

$\mu_A(X) = 0$ if x is not in A.

$0 < \mu_A\ (X) < 1$ if x is partly added in A.

$\mu_A(X)$ represent to which the integer value X included in the fuzzy set A. The datavalue 0 coincide to the infinite non-set membership and the value 1 belongs to the total infinite membership. Therefore, a fuzzy set membership function $\mu_A(X)$ shows the association-degree to some value element X of the domain of discourse X under consideration. This maps each value of X to a data-membership scale between 0 and 1 in various functions like triangular, trapezoidal, curvaceous and Gaussian. Triangular fuzzy-membership function that is extensively used is selected in this study of research. Triangular fuzzy-membership value function computes by the equations as below:

$$\mu_A(X) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{c-a} & \text{if } x \in [a, c] \\ \frac{b-x}{c-b} & \text{if } x \in [b, c] \\ 0 & \text{if } x \geq c \end{cases}$$

where a, b, and c have been defined by experts doctors.

3.1 Fuzzy Expert System

Many fuzzy expert systems have been developed but the efficiency is depending basically upon the decision of the domain scientists on different problems related

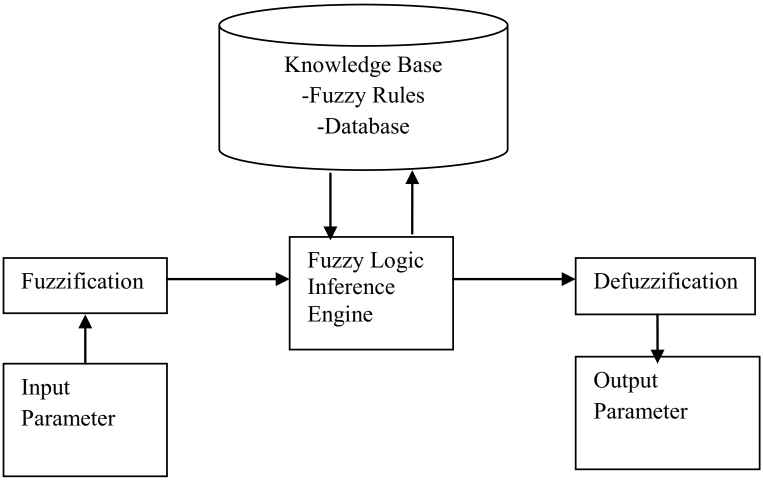


Fig. 1 Architecture of proposed system

to the work under consideration. This system is evaluated with the help of medical practitioners and specialist. The developed system, christened has an structural design conferred in following Fig. 1. The development of system involves many components like fuzzification, deduction (inference) engine and de-fuzzification. It is a rule-based system that uses fuzzy logic rather that Boolean logic. The system is designed and developed solely based on the knowledge-base, fuzzy logic rules and dataset.

Procedure for Fuzzy Logic-Based Malaria Diagnosis

The algorithm designed for the fuzzy rules-based diagnostic process of malaria:

1. First give m signs and observed symptoms of patient as an input the classification.
   where $m =$ numeral of signs and syndrome.
2. Look for the knowledge-base for the disease d, which has the syndrome identified.
3. Then find the weighing factors (wf) (it is the degree of intensity of symptoms associated)
   $wf = 1, 2, 3, 4$
   Where 1 = Mild, 2 = Moderate, 3 = Severe, 4 = Very Severe.
4. Use fuzzy logic rules on the input.
5. Correlate the fuzzy data as inputs into their particular weighing factors to find out their degree of relationship.
6. Calculate the fuzzy rule base evaluating.
7. Find out the firing strength of the rules R.
8. Compute the quantity of truth R, of each rules by evaluating the nonzero minimum data value.
9. Identify the level/depth of the disease.
10. Declare the fuzzy system investigation for decision making.

The Knowledge-Base of the System

Information (knowledge repository) is a key factor in the presentation of canny frameworks. The information base of the proposed framework is made out of an organized and compact portrayal of the information on space specialists of tropical prescriptions. The unpredictable information is worried about sureness, rule and presence of tropical sicknesses, which were as often as possible settled upon by specialists in the field of tropical prescription. For the assurance of this exploration, intestinal sickness as a realized central malady is dealt with.

Fuzzification

Fuzzification process is the dispose of identifying and changing a absolute scalar value into a fuzzy set. This is accomplished with various sorts of fuzzifiers. There are commonly four kinds of fuzzifiers, which are utilized for the fuzzification procedure. They are: Trapezoidal fuzzifier, triangular fuzzifiers, Singleton fuzzifier, and Gaussian fuzzifier [9]. Traingular fuzzifier which is broadly utilized will be utilized in this examination.

Fuzzification of information is completed by choosing input parameters into the flat pivot and anticipating vertically to the overhead limit of participation capacity to decide the level of enrollment.

The initial phase in the improvement of fluffy rationale-based master framework is to develop fluffy sets for the parameters. This is appeared in conditions underneath. Based on do-principle specialists' information, both information and yield parameters chose for this exploration were portrayed with four semantic factors (mild, i.e., gentle, moderate, severe: serious and very severe: extreme). The detail range of fuzzy value for all variables is given in the following Table 1:

Fuzzification starts with the upset of the crude information utilizing the capacities that are communicated in conditions beneath. During the activity, phonetic factors are look at utilizing triangular enrollment work and are go-to by consolidating level of participation running from 0 to 1 as appeared in conditions underneath. These recipes are unfaltering by the help of one and the other master specialists in the field of tropical prescription and writing.

The following stage in the fuzzification procedure is the advancement of fluffy (fuzzy) standards. The fuzzy logic principles for this examination were created with the help of area specialists. The knowledge-base of this system has so many fuzzy rules designed with the aid of consolidation theory but only the valid rules (20 in

Table 1 Range of fuzzy set values

Fuzzy linguistic variables	Fuzzy range values
Mild-gentle	$0.1 \leq x < 0.3$
Moderate	$0.3 \leq x < 0.6$
Severe-serious	$0.6 \leq x < 0.8$
Very severe-extreme	$0.8 \leq x \leq 1.0$

number) selected those are suggested by the field experts. Following are the some of the sample fuzzy rules for malaria.

Some of the Rules (Rules 1, Rules 2, 18 and Rule 20) Can Be Interpreted as Follows:

Rule1: IF fever = mild and headache = mild and nausea = mild and vomiting = mild and jaundice = mild and enlarge liver = mild and joint pain = mild and body weakness = mild, and dizziness = severe, and loss of appetite = mild and MP = mild THEN malaria = mild.

Rule 2: IF fever = moderate and headache = mild and nausea = mild and vomiting = mild and jaundice = mild and enlarge liver = mild and joint pain = moderate and body weakness = severe and dizziness = very severe, and loss of appetite = severe and MP = moderate THEN malaria = moderate.

Rule 18: IF fever = moderate and headache = Very Severe and nausea = Very Severe and vomiting = mild and jaundice = Severe and enlarge liver = Severe and joint pain = moderate and body weakness = severe and dizziness = very severe, and loss of appetite = very severe and MP = severe THEN malaria = very severe.

Rule 20: IF fever = very severe and headache = severe and nausea = severe and vomiting = moderate and jaundice = severe and enlarge liver = severe and joint pain = severe and body weakness = severe and dizziness = severe and loss of appetite = severe and MP = moderate THEN malaria = very severe.

Equations for fuzzification process:

$$\mu_{\text{sev}}(X) = \begin{cases} 0 \text{ if } x \leq 0.1 \\ \frac{x-0.1}{0.2} \text{ if } 0.1 \leq x \leq 0.3 \\ \frac{0.2-x}{0.1} \text{ if } 0.2 \leq x \leq 0.3 \\ 0 \text{ if } \ x \geq .0.2 \end{cases}$$

$$\mu_{\text{MSdrate}}\,(X) = \begin{cases} 0 \text{ if } x \leq 0.3 \\ \frac{x-0.3}{0.3} \text{ if } 0.3 \leq x \leq 0.6 \\ \frac{0.45-x}{0.15} \text{ if } 0.45 \leq x \leq 0.6 \\ 0 \text{ if } \ x \geq .0.45 \end{cases}$$

$$\mu_{\text{sin}\,w}(X) = \begin{cases} 0 \text{ if } x \leq 0.5 \\ \frac{x-0.6}{0.2} \text{ if } 0.6 \leq x \leq 0.8 \\ \frac{0.7-x}{0.1} \text{ if } 0.7 \leq x \leq 0.8 \\ 0 \text{ if } x \geq .0.7 \end{cases}$$

$$\mu_{\text{leps soner}}\,(X) = \begin{cases} 0 \text{ if } x \leq 0.8 \\ \frac{x-0.1}{0.2} \text{ if } 0.8 \leq x \leq 1.0 \\ \frac{0.2-x}{0.1} \text{ if } 0.9 \leq x \leq 1.0 \\ 0 \text{ if } \ x \leq 1.0 \end{cases}$$

4 Experimental Results

In the experimental analysis, total 30 patient's data is validated. It is observed that all the 20 rules are fired according to the input values of symptoms narrated by the patient. According to the result analysis, the output of this expert system is on the same line with the domain medical expert with + _ 5% differences as shown in the graph Fig. 2. For further understanding, one sample analysis is given here: seventeenth (17) rules were fired out for patient number 30, i.e., 17 rules generated nonzero minimum values from the fuzzy rule base for malaria. For all of the linguistic variables: mild, moderate, severe and very severe, the respective output membership function strength (range: 0–1) from the possible rules ($R1$–$R20$) are computed using RSS inference technique as shown in equations below:

Mild =	$\surd R\,6^2 + R\,10^2$ $\surd 0.25^2 + 0.25^2$ 0.3536
Moderate =	$\surd R\,2^2 + R\,5^2 + R\,8^2 + R\,9^2 + R\,13^2 + R\,19^2$ $\surd 0.25^2 + 0.25^2 + 0.25^2 + 0.25^2 + 0.25^2 + 0.25^2$ 0.6124
Severe =	$\surd R\,3^2 + R\,7^2 + R\,12^2 + R\,14^2 + R\,15^2$ $\surd 0.5^2 + 0.25^2 + 0.25^2 + 0.5^2 + 0.25^2$ 0.8291
Very severe =	$\surd R\,11^2 + R\,16^2 + R\,17^2 + R\,20^2$ $\surd 0.5^2 + 0.25^2 + 0.25^2 + 0.25^2$ 0.6614

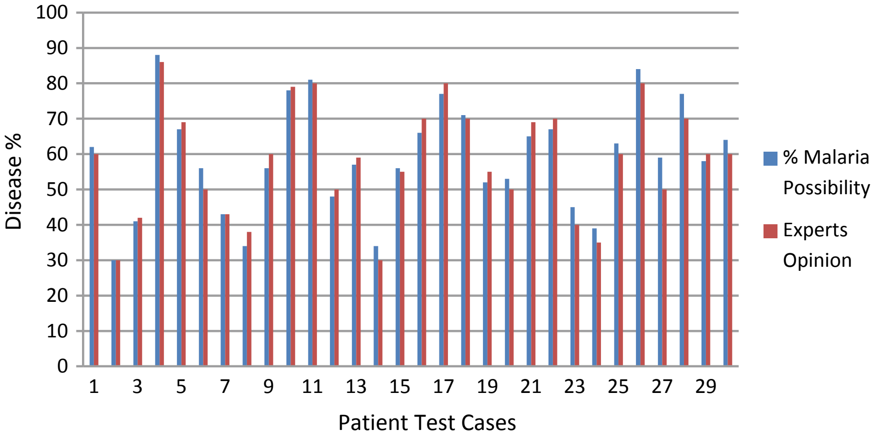


Fig. 2 Comparison of fuzzy expert systems output with medical experts opinion

The output from the fuzzylogic set calculated using the RSS later on defuzzified to get the crisp result for decision making. Then for defuzzification, it uses the discrete center of gravity technique; the sample calculation is as shown in equation below:

$$\begin{aligned}\text{Crisp Output} &= \frac{(0.5 * 0.2) + (0.75 * 0.4) + (0.79 * 0.62) + (1.22 * 0.98)}{(0.5 + 0.78 + 0.795 + 1.3)}\\ &= 0.62, \quad = 62\%\end{aligned}$$

5 Conclusion

In the research work presented, the fuzzy expert system is developed for malaria disease predication. The result analysis indicates that the fuzzy logic-based systems can be used for direct medical diagnosis. The function of fuzzy logic for healthcare diagnosis contributes an very helpful methodology to aid unsophisticated physicians to take a decision of malaria disease higher side instantaneously as well as precisely. This developed system is evaluated with the domain expert's opinion and it is found that the system resembles 95.9% with the expert's decision. Hence, the proposed technique developed in this study, if used reasonably, could be an efficient approach for diagnosing malaria. Furthermore, this work can be scaled up as multi-disease diagnosis system for detection other diseases showing similar kind of symptoms in patients.

References

1. Matui P, Wyatt JC, Pinnock H, Sheikh A, McLean S (2014) Computer decision support systems for asthma: a systematic review. Nat Partner J Pri Care Resp Med 24:14005. https://doi.org/10.1038/npjpcrm.2014.5
2. Apurba B, Arun KM, Anupam B (2007) A fuzzy expert approach using multiple experts for dynamic follow-up of endemic diseases. Artif Intell Expert Syst 19:53–73
3. Uzoka FME, Osuji J, Obot O (2010) Clinical decision support system (DSS) in the diagnosis of malaria: a case comparison of two soft computing methodologies. Expert Syst Appl 38:1537–1553
4. Doukidis GI, Cornford T, Foster D (1994) Medical expert system for developing countries: evaluation in practice. Expert Syst Appl 7:221–233
5. Uzoka FME, Barker K (2010) Expert systems and uncertainty in medical diagnosis: A proposal for fuzzy-AHP hybridisation. Int J Med Eng Inf 2:329–342
6. Devlin H, Devlin JK (2007) Decision support system in patient diagnosis and treatment. Fut Rheumatol 2:261–263
7. Mohd Zahari MK, Zaaba ZF (2017) Intelligent responsive indoor system (IRIS): a potential shoplifter security alert system. J Inf Commun Technol 16(2):262–282
8. Yang CL, Simons E, Foty RG, Subbarao P, To T, Dell SD (2016) Misdiagnosis of asthma in schoolchildren. Pediatr Pulmonol 52(3):293–302

9. Mohd Sharif NA, Ahmad N, Ahmad N, Mat Desa WLH, Mohamed Helmy K, Ang WC, Zainol Abidin IZ (2019) A fuzzy rule-based expert system for asthma severity identification in emergency department. J Inf Commun Technol 18(4):415–438
10. Morel CM (2000) Reaching maturity—25 years of the tropical disease research. Parasitol Today 16:522–528
11. Classen DC (1998) Clinical decision support systems to improve clinical practice and quality care. J Am Med Assoc 280:180–187
12. Briggs DJ (2008) A Framework for integrated environmental health impact assessment of systemic risks. Environ Health Malaria J 7:1186–1476
13. Tan CF, Wahidin LS, Khalil SN, Tamaldin N, Hu J, Rauterberg GWM (2016) The application of expert system: a review of research. ARPN J Eng Appl Sci 11(4):2448–2453. ISSN 1819–6608
14. Shortliffe EH (1997) Computer programs to support clinical decision making. J Am Med Assoc 258:61–66
15. Hudson DL, Cohen ME (1994) Fuzzy logic in medical expert system. IEEE Eng Med Bio 12:693–698
16. Uzoka FME, Famuyiwa FO (2004) A Framework for the application of knowledge technology to the management of diseases. Int J Health Care Q Ass 17:194–204
17. Adekoya AF, Akinwale AT Oke OE (2008) A medical expert system for managing tropical diseases. In: Proceedings of the third conference on science and national development, 74–86

Data Sharing Over the Cloud Based on Ring Signature

Priti Rumao, Mamta Meena, and Samidha Kurle

Abstract Cloud computing, which provides on demand delivery of services, wherever and whenever, mainly focuses on data sharing over the Internet. So, when data sharing among group of participants from same or dissimilar environment is involved, it may face compromising situation in terms with integrity, confidentiality and privacy of content/data as well as of data owner. One of the features of ring signature is to construct confidential what is more, unknown information sharing framework among the participants group. The ring signature can be used to provide signer's privacy as well as user's secrecy. But, for ring mark or signature, the certificate validity verification processing remains a pinch point which does not permit any framework to be flexible. So, to defeat this, the ID-based ring mark or signature can be used. Further, forward security is being appended to make ID-based ring mark or signature increasingly certain, on the grounds that regardless of whether single key used in one transaction have been undermined, still all recently created marks or signatures remain uncompromised and legitimate with the forward security theory use. In this, data sharing over the cloud based on ring signature system, ID-based ring mark or signature added with forward security is moreover improvised by appending Weil pairing, which keeps any key secure irrespective of its size, and it also has less time and space multifaceted nature and hence saves power. All these combined together is essential to construct scalable information sharing framework. So, this paper provides proficient and substantial application of the above-mentioned framework.

Keywords Cloud computing · Data sharing · Forward security · ID-based ring signature · Authentication · Weil pairing

P. Rumao · M. Meena (✉) · S. Kurle
Computer Engineering Department, Atharva College of Engineering, Malad West, Mumbai 400095, India
e-mail: mamtameena@atharvacoe.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_10

1 Introduction

1.1 Overview

The current era is of digitization and digitalization; hence, the focus is on e-information generated from all the things around, and hence, huge amount of data is being generated every minute. Cloud computing is recent big shift of IT companies in terms of how business used to work and how it works now with many features, among which data storage (IaaS) of hugely generated data and its sharing provides many benefits its every user legitimately or by implication. Cloud computing provides many services like SaaS, PaaS, IaaS, Naas, etc., but is generally utilized for information stockpiling and information distribution via distributed computing or cloud computing, and enormous number of cloud service users (customers) gains huge data. But many participants using same data and space may result in bargaining information uprightness, confidentiality what is more, protection. Hence, to conquer above-mentioned issues, Rivest, Shamir and Tauman projected a theory of ring mark or signature [1]. The catch here is, for ring signatures, their id requirement of certificates validity check as well as management, which results in spending more time and space on it. Therefore, the hypothesis of identity-based (Id-Based) cryptosystem [2] was acquainted by Shamir with overcome declarations confirmation just as the board issues. Here, Shamir presented that here for singular client, the open key can be registered from information identified with the client's freely known character. Along these lines, from its lord mystery key database for clients, a private key generator (PKG) generates unique random private keys for individual user doing any transaction. Hence, this property of Shamir, avoid the requirement of certificates to be validated what is more, goes with a basic open key to singular client inside the framework. So, basically, this Id-based framework takes out the need of authentications legitimacy check just as the executives and spares correspondence and computational (time) cost. The amalgam of ring mark or signature with Id-based framework gives the result which has upper hand over certificates-based ring mark or signature framework. The past authentication-based ring mark or signature theory is used to make verifier validate individual testament of the individuals (clients) present in the ring (framework). After endorsement legitimacy check is practiced, at that point, verifier confirms the message and client's mark. Now, for Id-based ring mark or signature theory, complete endorsement validity verification process is eradicated. Here, verifier simply needs to confirm client's identity (authentication) and associated message-key pair for that specific user. As certificate validity verification process is dispensed with, it spares the time and calculation cost hugely. But such system is salutary to the associations which generates and works with large data and has many participants. Zhang and Kim (2002) submitted the very first ID-based ring mark or signature theory [3]. Other existing ID-based ring mark or signature theories are proposed in [4–11].

Scalability of ring for data distribution is utmost important. Large members can be added for providing more protection, and it is like combination of all keys of users present in ring generate stronger key. But it has another side as well; by increasing users' number in the ring, probability of key exposure to outside world increases as even single advisory will compromise entire shared data. Basically, if signer's one private key of particular transaction is exposed, then all signatures generated by that particular underwriter become worthless; for example, future marks to be generated cannot be authenticated and what is more, recently gave marks cannot be convicted. So, when leakage of key by any means is recognized, then the key revocation mechanisms is be invoked effectively so that it will help to control the age of further mark utilizing the uncovered private key. All things considered, it is not adept for taking care of the issue of the past marks which were manufactured.

The hypothesis of forward secure mark or signature has been presented in 1997 by Anderson [12] to monitor the legitimacy and realness of the past marks or signatures despite the fact that the present mark or signature is uncovered and traded off, and answers for the equivalent were structured in 1999 by Bellare and Miner [13]. The model planned by Bellare and Miner is: Full meeting-based time (T) legitimacy of open (public) key is partitioned into littler time pieces (ti). Consequently, regardless of whether during the current time lump (ti) a key is uncovered, still then for enemy, it will be hard to discover recently created or the next keys dependent on it. Id-based ring signature and forward security joined to produce strong arrangement is ideated by Huang et al. [14]. Here, for ring marks or signatures, centrality of forward security is underlined and vigorous calculation has been conceptualized. Its absolute first full-verification plan of Id-based ring mark or signature with forward security and its security has been tried (demonstrated) under the standard RSA supposition, with the assistance of irregular prophet (random oracle) model. Still, it is constantly conceivable to decrease time cost for the framework.

In 1940, French mathematician Andre Weil ideated the concept of Weil pairing. The main focus of this theory was on elliptic curve cryptography (ECC) system [15]. Sakai et al. [16] proposed that Weil pairing can be used for key exchange as well. Boneh et al. [17] designed identity-based encryption system which has used random oracle model to pick up cipher text security, where encryption and decryption are done by two different algorithms: Encryption is performed by modified Elgamal algorithm and decryption is performed by modified Weil pairing computation.

Hence, for summarizing, data sharing over the cloud based on ring signature utilizes favorable circumstances of each of the three disparate methods: forward security, Id-Based ring mark or signature and Weil blending (pairing) to design a strong arrangement.

1.2 Proposed Methodology

A new concept called data sharing over the cloud based on ring signature which demonstrates its significance by building financially savvy (time based) bonafide and unknown framework for data sharing has been proposed in this paper.

- This system ideates concrete robust solution with the assistance of forward secure Id-based ring mark or signature using Weil pairing.
- Practical real-life implementation of mentioned system is possible, because:
  - Id-based ring mark or signature scheme has been used, so certificate validation cost has been taken out completely.
  - Even small-sized secret key can be used, and with that, key updating process is easy and generates required results as well.
  - Time consumption is further reduced by the use of Weil pairing, and it enhances security process as well.

2 Implementation

Content/data sharing over the cloud based on ring signature system introduces an effective and proficient information sharing framework for distributed computing or cloud computing condition which aids validation of information or data and signer's anonymity. The workflow is shown in Figs. 1 and 2. In this:

1. Data (cloud) owner initiates the process by registering and authenticating users into its system. The current prototype of system has single owner and multiple users.
2. After registration, authentication process of user into the framework (system) is done, and then, cloud owner uploads file (document) to be shared on the framework (system).
3. Once file/document is uploaded to the framework (system), cloud proprietor separates gathering of clients to shape a ring, among the pool of clients just with

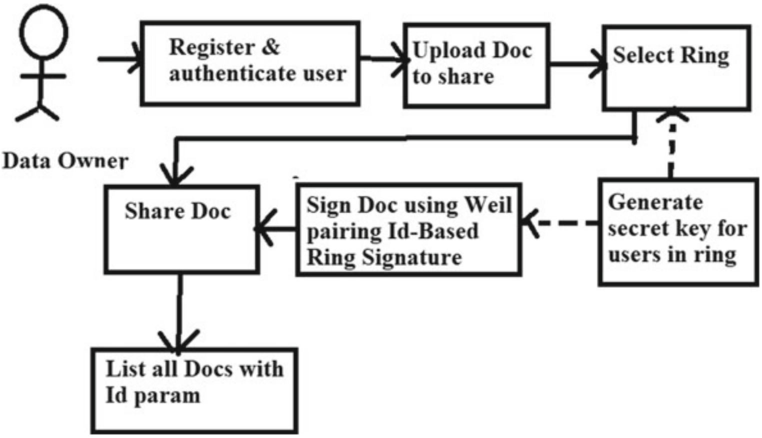

Data Owner
Register & authenticate user
Upload Doc to share
Select Ring
Share Doc
Sign Doc using Weil pairing Id-Based Ring Signature
Generate secret key for users in ring
List all Docs with Id param

Fig. 1 Workflow of uploading the document over the cloud

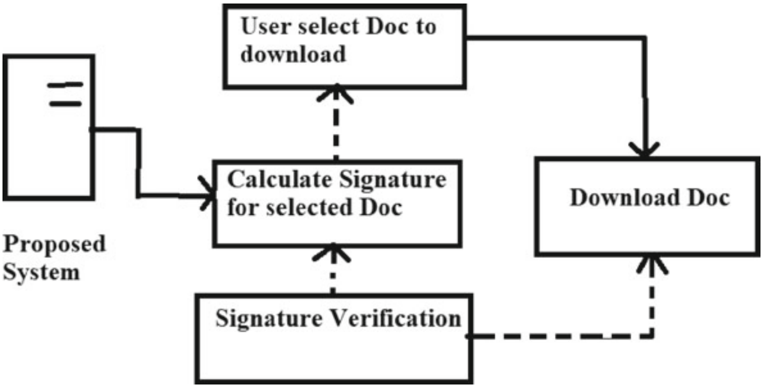

User select Doc to download
Calculate Signature for selected Doc
Download Doc
Proposed System
Signature Verification

Fig. 2 Workflow of downloading the document from the cloud

whom, document (information) is to be shared and generate the secret key for their ring.

4. For sharing the selected document, using ID-based ring mark or signature scheme with Weil blending (pairing), the document has been signed (S_i) and afterward shared. Here, all clients present in the framework (system) will be able to see all the shared and listed documents, but only, selected users (who are part of the ring) are allowed to download the said document
5. To download the record, recently, determined mark (Si) and mark or signature with respect to the user who wants to download the document, as well as selected document are being verified.
6. Succeeding the signature verification process, document can be downloaded.

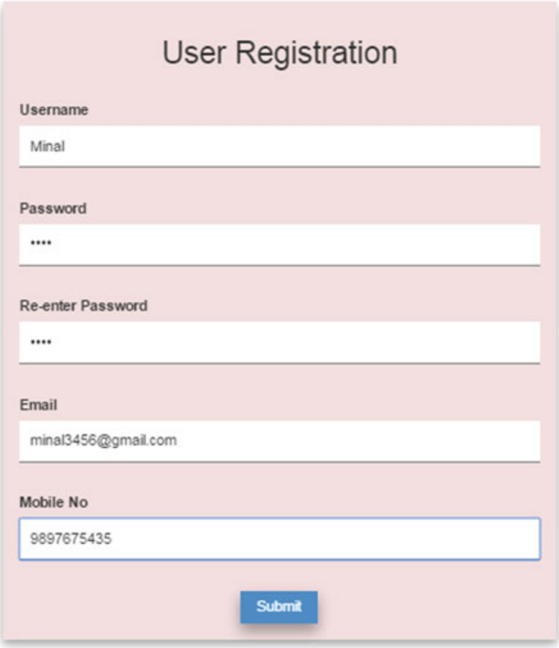
Steps below elaborate the workflow mentioned above:

Step 1: Registration and authentication of client (user): Here, data owner accepts and authenticates user into the system. While doing so, the idea of forward security is being implemented. Here, for each user against its public key, all-out legitimacy time of a year (may shift) is appointed. This associated time span has been separated into smaller time lumps (chunks). During those time pieces (chunks), private key, for example, password of the client (user) will stay legitimate. Before the end of ongoing time slot, system will seek for change of password. In spite of it, if password is not changed, then lattermost of the ongoing time slot if password is not changed instantly, system will not allow private key users' login (Refer Fig. 3).

Step 2: Uploads the document (in any suitable format) to be shared into the system (Refer Fig. 4).

Step 3: Select Ring: During this step, ring (gathering) signature has been created utilizing Weil pairing, and afterward, this mark or signature with document-related data (information) and client's (user's) public key, one more secret key is generated and sent to users of ring, and file is listed in common data (Refer Fig. 5).

Step 4: Share document: Here, document is shared and all applied algorithms become effective in this step. (Refer Fig. 6).



User Registration

Username: Minal

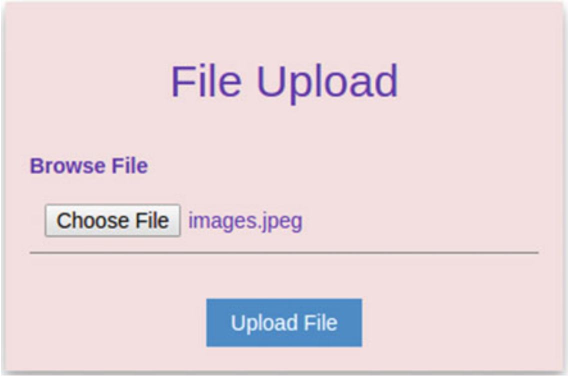
Password: ****

Re-enter Password: ****

Email: minal3456@gmail.com

Mobile No: 9897675435

Fig. 3 Registration form



File Upload

Browse File

images.jpeg

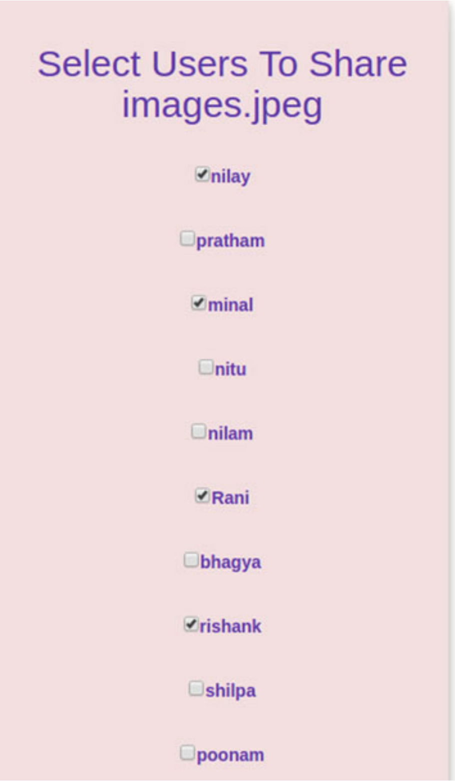
Fig. 4 Upload file (independent of format up to 10 MB)

Step 5: Documents shared so far are listed to view by all users. (Refer Fig. 7).

Step 6: User selects document to download: In regards with selected document and user information who wants to download the report (document), mark or signature is being produced, and afterward, presently created signature and recently produced signature related with that record (document) are being validated.

Step 7: After ensuring that both the marks and signatures are coordinated at that point for that particular user, selected document will be downloaded. User whose

Fig. 5 Select ring



Select Users To Share images.jpeg

☑ nilay

□ pratham

☑ minal

□ nitu

□ nilam

☑ Rani

□ bhagya

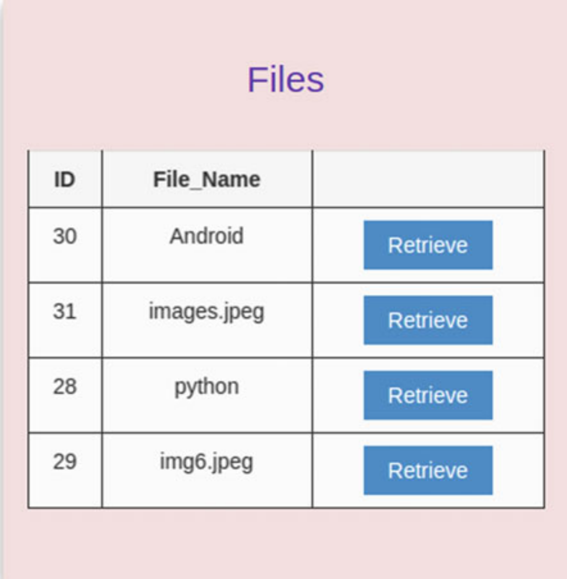
☑ rishank

□ shilpa

□ poonam

Fig. 6 Document shared

File Shared Successfully !!!



Files

ID	File_Name	
30	Android	Retrieve
31	images.jpeg	Retrieve
28	python	Retrieve
29	img6.jpeg	Retrieve

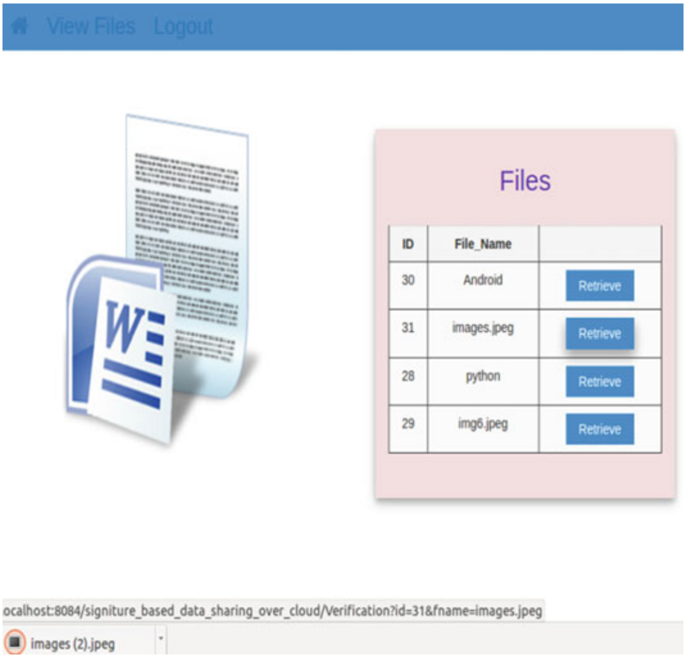
Fig. 7 All shared documents

signature is not equated is not permitted to get to that specific archive (document) (Refer Fig. 8).

3 Comparative Studies

Here, comparison and investigation of the proposed framework (system) concerning leaving framework (system) [14] are appeared. This examination is directed by methods for of execution time parameter (time unpredictability). The current framework (system) which is forward secure ID-based ring mark or signature framework (system) [14] works on two distinct ideas: forward security and ID-based ring mark or signature, though the proposed framework (system) centers around Weil blending (pairing), forward security and id-based ring mark. The proposed framework (system) theorizes on and actualizes the idea of blending (pairing) for secret key generation, though the existing framework (system) keeps away from idea of matching entirely.

The results of examination of existing and proposed framework are shown in Fig. 10 (an) and (b). It plainly shows that archive (document) recovery time for the existing framework is higher when compared with the proposed framework (system). So, arguably, the proposed system provides time-efficient solution when compared to the existing system.



Files

ID	File_Name	
30	Android	Retrieve
31	images.jpeg	Retrieve
28	python	Retrieve
29	img6.jpeg	Retrieve

ocalhost:8084/signiture_based_data_sharing_over_cloud/Verification?id=31&fname=images.jpeg

images (2).jpeg

Fig. 8 Document downloaded after match

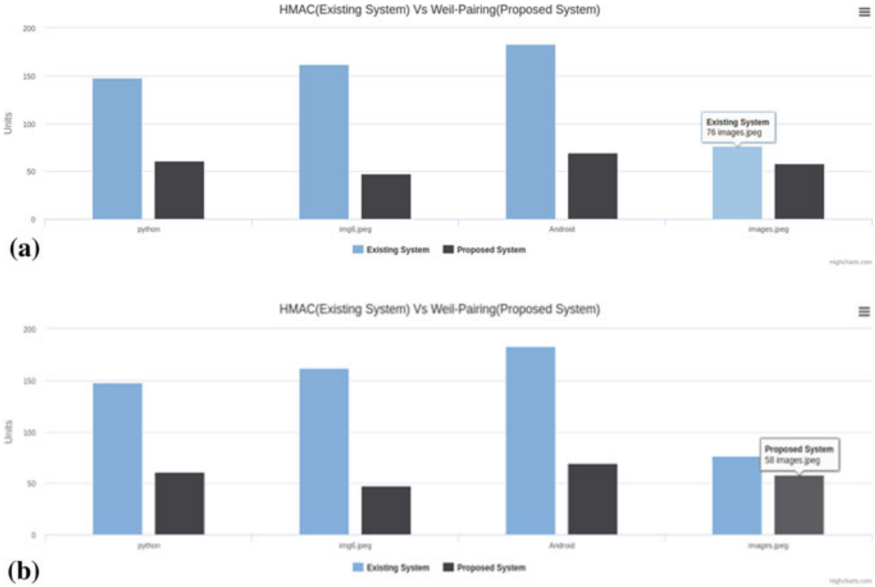


Fig. 10 **a b** Bar chart of comparison of time required to retrieve documents between the existing and proposed systems

4 Conclusion

This paper proposed a concept called data sharing over the cloud based on ring signature to fulfill the need of data sharing within group of users securely. It consolidates forward secure ID-based ring mark or signature and Weil blending (pairing). This concept stipulates authentication or validation of data anonymity of signer (user) while reduces execution time as well. Weil blending (pairing) further provides more secure signature generation structure, hence making system more efficient. This concept will be very useful in terms of security aspect in many real-life applications and where authentication, authorization and user's anonymity play vital role. The current focus of the proposed concept is on reducing execution time and is successful doing so. Providing more security in terms of confidentiality about shared data and still minimizing cost regarding time and space parameters as well as increasing size of data or information to be uploaded are open issue (problem) and future research work.

References

1. Shamir A, Rivest R, Tauman T (2001) How to leak secret. In: Asiacrypt'01 LNCS 2248, 552,565.
2. Shamir A (1984) Identity-based cryptosystems and signature schemes. In: Crypto 1984, volume 196 of lecture notes in computer science. Springer, pp 47–53
3. Zhang F, Kim K (2002) Id-based blind signature and ring signature from pairings. In: Asiacrypt 2002, volume 2501 of lecture notes in computer science. Springer, pp 533–547
4. Chien, H-Y (2008) Highly efficient Id-based ring signature from pairings. In: APSCC, pp 829–834
5. Awasthi AK, Lal S (2005) Id-based ring signature and proxy ring signature schemes from Bilinear pairings. Corr, Abs/Cs/0504097
6. Zhang J (2007) An efficient identity-based ring signature scheme and its extension. In: ICCSA (2), volume 4706 of lecture notes in computer science. Springer, pp 63–74
7. Chow SSM, Yiu S-M, Hui LCK (2005) Efficient identity based ring signature. In: ACNS 2005, volume 3531 of lecture notes in computer science. Springer, pp 499–512
8. Chow SS, Lui RW, Hui LC, Yiu S (2005) Identity based ring signature: why, how and what next. In: Chadwick, D, Zhao G (eds) Europki, volume 3545 of lecture notes in computer science. Springer, pp 144–161
9. Nguyen L (2005) Accumulators from bilinear pairings and applications. In: Menezes AJ (ed) Ct-Rsa 2005, volume 3376 of lecture notes in computer science. Springer, pp 275–292
10. Herranz J (2007) Identity-based ring signatures from RSA. Theor Comput Sci 389(1–2):100–117
11. Tsang PP, Au MH, Liu JK, Susilo W, Wong DS (2010) A suite of non-pairing Id-based threshold ring signature schemes with different levels of anonymity (Extended Abstract). In: Provsec, volume 6402 of lecture notes in computer science. Springer, pp 166–183
12. Anderson R (2000) Two remarks on public-key cryptology. Manuscript, Sep. 2000. Relevant material presented by the author in an invited lecture at the fourth ACM conference on computer and communications security, 1997
13. Bellare M, Miner S (1999) A forward-secure digital signature scheme. In: Crypto'99, volume 1666 of lecture notes in computer science. Springer-Verlag, pp 431–448

14. Huang X, Liu J, Tang S, Xiang Y, Liang K, Xu L, Zhou J (2015) Cost-effective authentic and anonymous data sharing with forward security. Comput IEEE Trans 64(6)
15. Frey G, Muller M, Ruck H (1999) The Tate pairing and the discrete logarithm applied to elliptic curve cryptosystems. IEEE Trans Info Th 45:1717–1718
16. Sakai R, Ohgishi K, Kasahara M (2000) Cryptosystems based on pairings. In: Proceedings of symposium on cryptography and information security. Japan
17. Boneh D, Franklin M (2003) Identity-based encryption from the weil pairing. Appears Siam J Comput 32(3):586–615

Data Analytics on Columnar Databases in Big Data Environment

Priyanka Bandagale and Vinayak Ashok Bharadi

Abstract Traditional approaches to data warehousing have significant drawbacks in terms of effectively delivering a solution to various businesses requirements. They were also designed primarily for online transaction processing (OLTP), making them unsuited for the growing need for analytics and ad hoc queries. Significant other drawbacks include high licensing and storage cost, slow query performance against large data volumes, and difficulty in providing access to all of the data stored in multiple databases. Columnar databases can be very helpful in your big data project. We have described and implemented the column-oriented approaches for OLTP and shown how it is efficient when compared to row-oriented databases.

Keywords OLTP · Columnar orientation · Infobright · ICE · DPN · PHP

1 Introduction

Over the past decade, business intelligence has emerged as one of the highest priority items on chief information officer (CIO) agendas. Businesses and government agencies know that mining information from the increasingly huge volumes of data they collect is critical to their business or mission. Conventional database technologies were not formulated to cater for such a huge amount of data and hence their performance deteriorates significantly as volumes increase. Some of the reasons for this are:

- The primary limiting factor is disk I/O, while the cost of disk has decreased and data transfer rates have not changed. Therefore, accessing large structures such as tables or indexes is slow.

P. Bandagale · V. A. Bharadi (✉)
Finolex Academy of Management and Technology, Mumbai University, Ratnagiri, Maharashtra, India
e-mail: vinayak.bharadi@famt.ac.in

P. Bandagale
e-mail: Priyanka.bandagale@famt.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_11

- A row-oriented design mandates the database to retrieve all column data regardless of its use to resolve the query.
- Data loading speed degrades as the indexes required to be recreated as data is appended; this causes huge number of sorts (another very slow operation).
- The ideal design to address this problem must reduce the disk I/O, access only the data required for a query, and be able to operate on the data at a much higher level to eliminate the volume of the data.

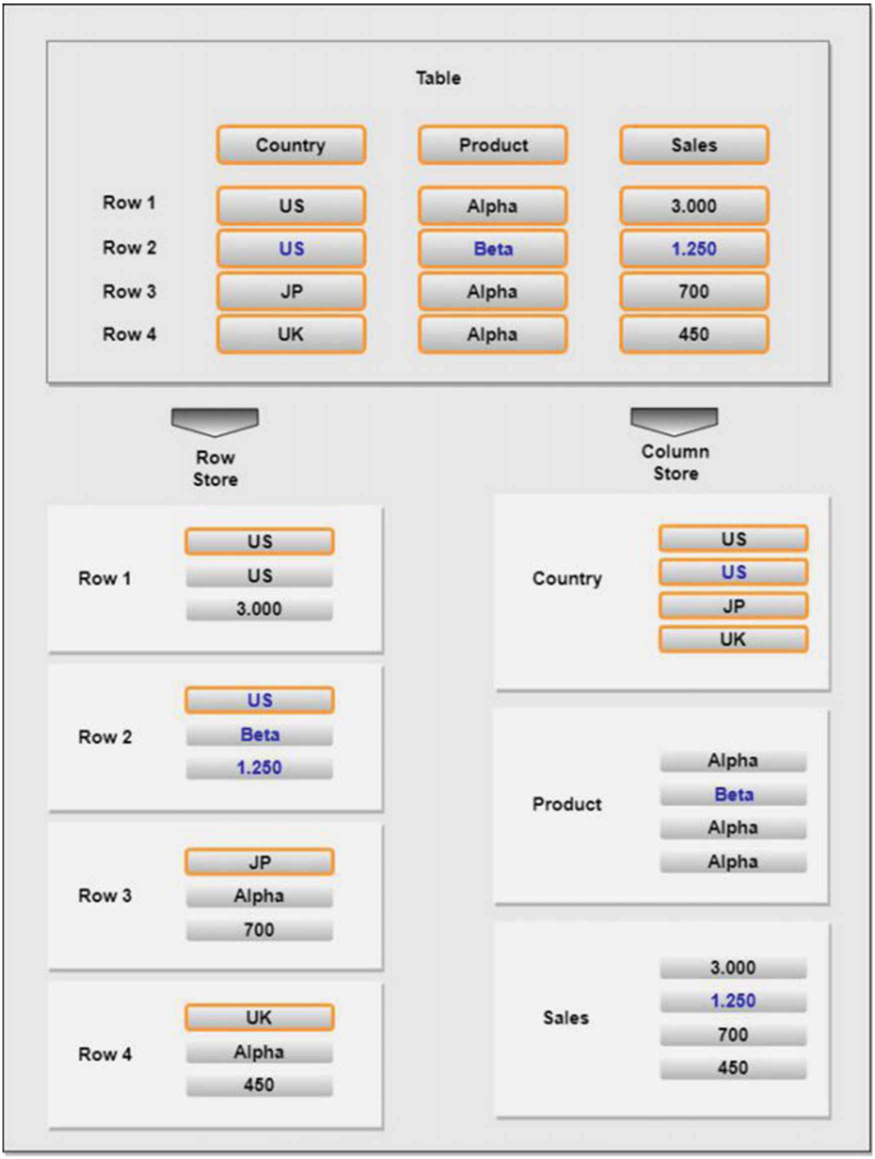
1.1 Columnar Versus Row Databases

Column database stores data in columns instead of rows. So, when we query data, the row store has to go through the whole row, whereas in column store, we can just work with the specific column(s) we are interested in. This enables massive performance benefits in scaling of the database [1, 2]. Figure 1 displays organization of data on disk in column and row stores.

1.2 Advantages of Columnar Databases [3]

- **Better analytic performance**: Column-oriented approach gives better performance in the execution of a large number of simultaneous queries.
- **Rapid joins and aggregation**: data access streaming on the column-oriented data results in incrementally computing the results of aggregate functions, and this is a crucial aspect for data warehouse applications.
- **Suitability for compression**: The storage of multiple indexes is removed along with views and aggregations. This addition facilitates fast improvement in compression.
- **Rapid data loading**: In columnar data arrangement, the framework essentially allows to segregate storage by columns. This means each column is built in one pass and stored separately, causing the database system to load columns in parallel using multiple threads.
- Finally, since the columns are stored separately, entire table columns can be added or dropped without downing the system and without the need to re-tuning the system following the change.

The paper is organized in four sections followed by references. Section 2 describes the open-source software which we have used to implement the system. Section 3 describes the implementation showing comparison of row and column databases. Section 4 summarizes the paper.



Table

	Country	Product	Sales
Row 1	US	Alpha	3.000
Row 2	US	Beta	1.250
Row 3	JP	Alpha	700
Row 4	UK	Alpha	450

Fig. 1 Actual storage on the hard disk

2 Infobright Community Edition (ICE)

ICE is an open-source DBMS developed to deliver a scalable data warehouse management [4, 5].

2.1 *Key Benefits*

- Supports large volumes, suitable for data volumes up to 30 TB
- Industry leads data compression (from 10:1 to over 40:1), and this compression mechanism heavily reduces disk I/O (improving query performance) and gives significantly less storage than the available options
- Low latency for complex analytic queries
- Query and load performance remain unchanged as the storage requirements increase
- Not dependent on for specific database schemas, e.g., star schema
- Specific materialized views are not required, and it has complex data partitioning strategies or indexing
- Simple to implement and manage, requiring little administration
- Less number of servers are required, and this effects in lower CAPEX and OPEX. It can be deployed on the low cost, off-the-shelf hardware
- Easy integration with major business intelligence tools such as Pentaho, JasperSoft, Cognos, Business Objects, and others.

2.2 *Infobright's Architecture*

It is based on the following concepts:

- Column orientation
- Data packs and data pack nodes
- Knowledge nodes and the knowledge grid
- The optimizer
- Data compression

The architecture is shown in Fig. 2.

2.2.1 Column Orientation

Infobright by the design is a greatly compressed column-oriented database. Column orientation brings lots of advantages, such as the capability to do more with efficient data compression because each column stores a single data type (as opposed to

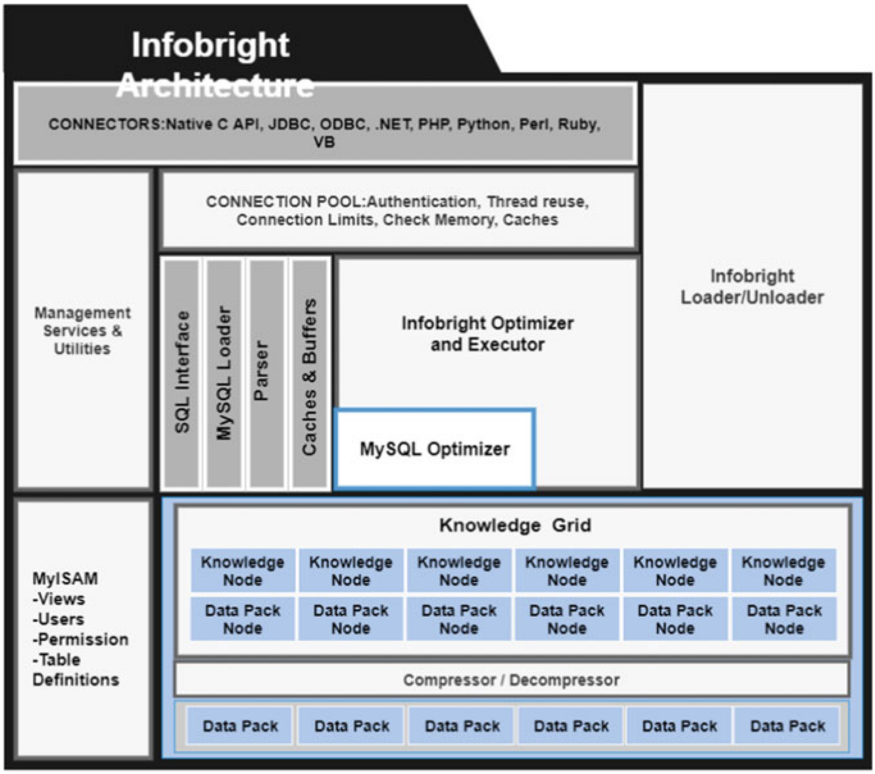

Infobright
Architecture
CONNECTORS:Native C API, JDBC, ODBC, .NET, PHP, Python, Perl, Ruby, VB
CONNECTION POOL:Authentication, Thread reuse, Connection Limits, Check Memory, Caches
Management Services & Utilities
SQL Interface
MySQL Loader
Parser
Caches & Buffers
Infobright Optimizer and Executor
MySQL Optimizer
Infobright Loader/Unloader
MyISAM
-Views
-Users
-Permission
-Table Definitions
Knowledge Grid
Knowledge Node
Data Pack Node
Compressor / Decompressor
Data Pack

Fig. 2 Architecture of infobright

rows that typically contain several data types), further, infobright compression to be optimized for each particular data type, drastically decreasing the disk I/O.

2.2.2 Data Organization and the Knowledge Grid

Infobright deploys the data into three layers:

Data Packs

The data inside the columns is stored in 65,536 item groups referred as the data packs. The deployment of the data packs improves data compression as they are smaller subsets of the columnar data and the compression algorithm can be executed based on the specific data type as shown in Fig. 3.

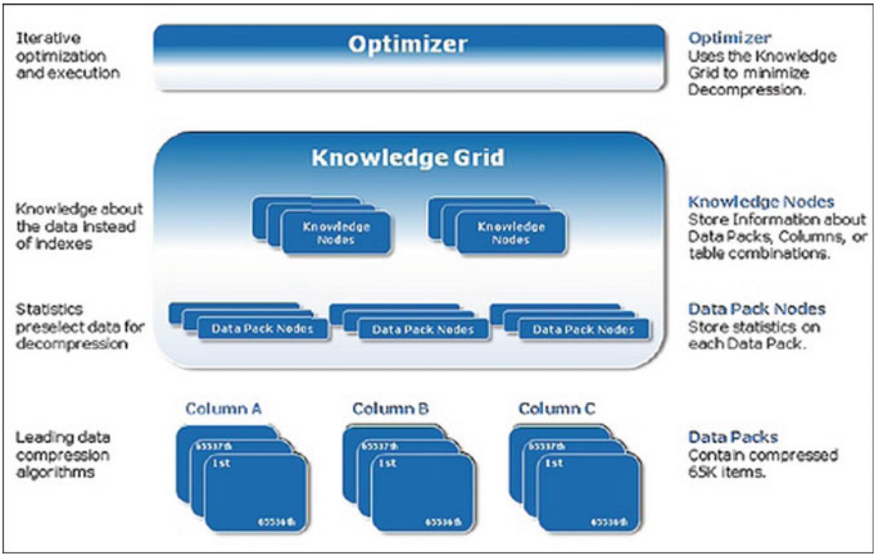

Iterative optimization and execution
Optimizer
Optimizer
Uses the Knowledge Grid to minimize Decompression.
Knowledge Grid
Knowledge about the data instead of indexes
Knowledge Nodes
Knowledge Nodes
Knowledge Nodes
Store Information about Data Packs, Columns, or table combinations.
Statistics preselect data for decompression
Data Pack Nodes
Data Pack Nodes
Data Pack Nodes
Data Pack Nodes
Store statistics on each Data Pack.
Column A
Column B
Column C
Leading data compression algorithms
65537th
1st
65536th
Data Packs
Contain compressed 65K items.

Fig. 3 Data organization

Data Pack Nodes (DPNs)

Data pack nodes consist of a set of statistics related to the data which is stored and compressed in each of the above-mentioned data packs. It always has a 1-to-1 relationship between data packs and DPNs. DPNs are always present and hence the Infobright has some knowledge about all the data in the database, in contrast to the conventional databases wherein the indexes are created for only a column's subset.

Knowledge Nodes

This is a set of metadata related to column relationships or data packs. They are more insightful on the data, detailing the ranges of value occurrences, or they can be retrospective, detailing how they relate to other data in the datastore. Most KNs are generated at the database service loading time, but others are generated in response to the specific queries for the purpose of performance optimization.

2.2.3 The Infobright Optimizer

The optimizer is the topmost degree of intelligence in the architecture. It makes the use of the knowledge grid to estimate the lowest possible set of data packs, which need to be decompressed in view to satisfy a given query with the lowest latency.

2.2.4 Data Loading and Compression

The process of creation and storage of data packs and their DPNs is described in Fig. 4. Unlike the conventional row-based data warehouses, data is stored by column, allowing efficient compression algorithms to be neatly designed to the column data type. Furthermore, for each column, the data is split into data packs with each storing up to 65,536 values of a given column. The mechanism then applies a set of innovative compression algorithms that are optimized by automatically self-adjusting various parameters of the algorithm for each data pack. Inside Infobright, the compression ratio may vary depending on data types and content. In addition, some data parts may be more repetitive than others and hence compression ratios may be as high as 40:1. On an average, a compression ratio of 10:1 a common thing.

3 Implementation and Results

The PHP Hypertext Preprocessor (PHP) is a programming platform for the creation of dynamic content that communicates with databases. This server-side scripting

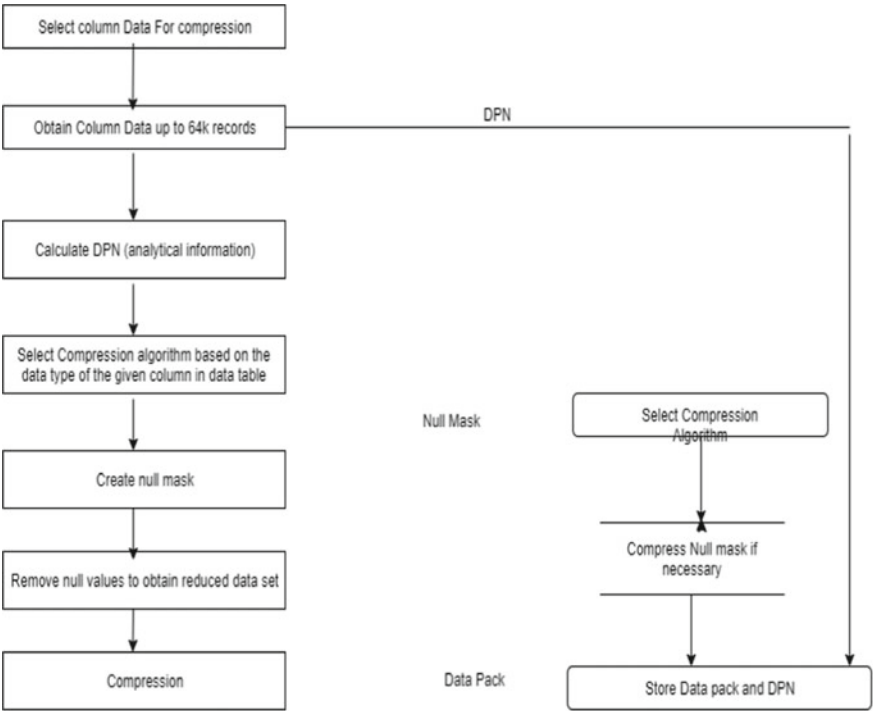

Select column Data For compression
Obtain Column Data up to 64k records
DPN
Calculate DPN (analytical information)
Select Compression algorithm based on the data type of the given column in data table
Null Mask
Select Compression Algorithm
Create null mask
Compress Null mask if necessary
Remove null values to obtain reduced data set
Compression
Data Pack
Store Data pack and DPN

Fig. 4 Data storing and compression

language designed for Web development is used as a general-purpose programming language [6]. Since it provides easy connectivity to databases and its ease to understand, it is chosen for implementation. XAMPP is used which provides Apache server to run the scripts and MySQL which is a row-oriented database [7].

3.1 Database Used for Implementation

Once Infobright is downloaded and installed, you are ready to establish a connection to the Infobright columnar database. It is simply a matter of executing the following PHP code as you would with MySQL. In the rudimentary example, we will use root for the user and ‘password’ as blank [8].

3.2 Loading Data in Database

Infobright includes a dedicated high-performance loader that differs from the standard MySQL Loader. The Infobright Loader is designed for speed, but supports less LOAD syntax than the MySQL Loader, and only supports variable length text formatted load files [9].

3.2.1 Infobright Loader Syntax

MySQL Loader: Loads from a flat file include text and field delimiters, support more features for escaping embedded characters, error handling, and transforming data using functions at the time of load then the Infobright Loader. This loader is set using an environment variable (@bh_dataformat = ‘mysql’).

Import your data into an Infobright table by using the following load syntax (all other MySQL Loader syntax is not supported):

```
LOAD DATA INFILE ‘/full_path/file_name’
INTO TABLE tbl_name
[FIELDS
[TERMINATED BY ‘char’]
[ENCLOSED BY ‘char’]
[ESCAPED BY ‘char’]
];
Example: Loading data in OrderDetails Table in the database:
LOAD DATA LOCAL INFILE ‘e:/datafiles/orderdetails.txt’ INTO TABLE OrderDetails
FIELDS TERMINATED BY ‘,’ ENCLOSED BY ‘"’ LINES TERMINATED BY ‘\r\n’;
```

To load data in the database, a form is provided to enter the name of the table.

Considering we have entered Orderdetails in the form above, in MySQL, the query executes in 0.211844 s.

In Infobright, the same query executes in 0.43 s. Note: The performance is affected in Infobright due to limitations of ICE (single threaded). In ICE, due to multithreaded operation, it is faster than MySQL.

3.3 Compression

One of the major advantages of a column-based approach is its effect on compression. Since data is stored by column, and each column consists of a single data type, it is possible to apply optimal compression algorithms for each column. This may seem like a small change but the difference in database size can be very significant when compared to other approaches. The compression information is stored in information schema database [10].

Compression in Infobright takes place automatically whereas in MySQL we need to use pack command to compress and unpack to uncompressing. Once compressed in MySQL the data becomes readable only. Hence, for every update, we need to uncompressing it.

3.3.1 Syntax for MySQL Compression is

```
myisampack c:/xampp/mysql/data/classicmodel/orderdetails.MYI
```

The compression is Fig. 4.55%. Hence, the file size becomes 43.55 KB. This can be shown from the information stored in information schema database. The query to show the size of the data file stored after compression is given by the query.

```
SELECT table_name AS “Table”,round(((data_length + index_length)/1024), 2) “Size in KB” FROM information_schema. TABLES WHERE table_schema = “classicmodels” AND table_name = “orderdetails”;
```

In Infobright, compression takes place during loading and the compression is nearly 80%. Hence, file size becomes 15.39 Kb. The same query above can be used to get the file size of the compressed file.

3.4 Data Analysis

Column-oriented databases are a good option for the data analytics, in the said cases only portions of each record are required. By grouping the data together in columnar way, the database only needs to retrieve columns that are related to the query, heavily reducing the overall disk I/O required. Query showing the sum of all products ordered is given as:

```
SELECT sum(quantityordered) from orderdetails
```

Query execution time in MySQL is 0.006655 s, whereas same query when executed in Infobright takes 0.002024 s.

3.4.1 Displaying Data Analyzed in Graphical Format

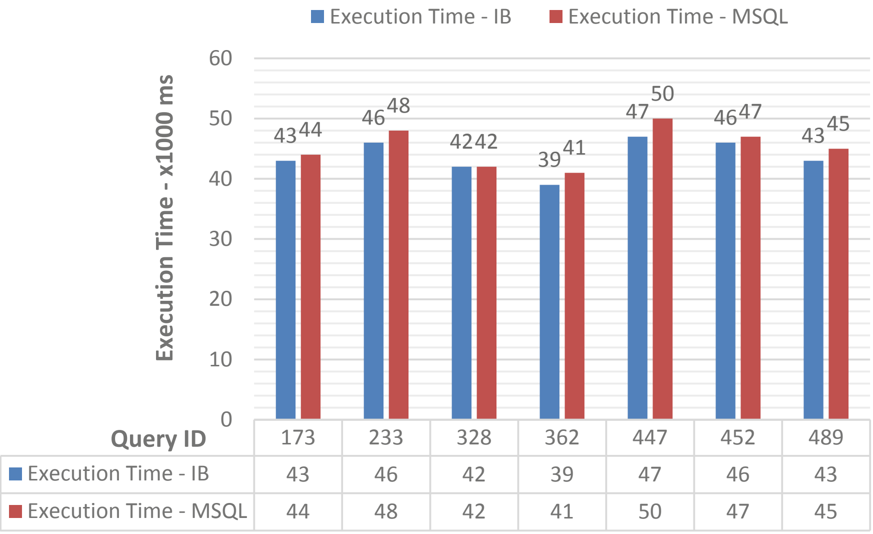
JpGraph is an object-oriented graph drawing library for PHP 5.1 and above. Some of the important features are: multiple scale formats; anti-aliasing of the lines; color-gradient fills; support for PNG, JPG, and GIF standards; and support for multiple Y axes. JpGraph can dynamically generate various types of graphs based on data you feed it from a database/array.

With JpGraph, it is easy to draw both 'quick and complex' graphs with a less of code and complex professional graphs which require a very fine grain control. JpGraph is equally well suited for both scientific and business type of graphs.

Query

"SELECT sum(creditlimit), customernumber from customers group by customer number having sum(creditlimit) between 40,000 and 50,000";

In Infobright, the execution time and bar chart for the same query is shown in Fig. 5, InfoBright has less execution timing as compared to MySQL.



Query ID	173	233	328	362	447	452	489
🟦 Execution Time - IB	43	46	42	39	47	46	43
🟥 Execution Time - MSQL	44	48	42	41	50	47	45

Fig. 5 Infobright and MySQL query execution performance

4 Conclusion

The working of columnar databases for OLTP operations is better and is presented in this paper on loading compression and data analysis speed. This is implemented using ICE as backend and PHP as frontend. The columnar databases for large volume of data are analyzed. Infobright database is having better performance 10% faster as compared to MySQL.

References

1. Kanade AS, Gopal A (2013) Choosing right database system: row or column-store. In: IEEE international conference on information communication and embedded systems, pp 16–20, 21–22
2. Gopal A, Bhagat V (2012) Comparative study of row and column-oriented databases. In: Fifth international conference on emerging trends in engineering and technology (ICETET), published in IEEE, pp 196–201
3. Infobright Community Edition 4.0.3 GA Quick Start by www.infobright.org
4. Infobright Analytic Database: Architecture, 2012 by www.infobright.org
5. Download software at www.infobright.org/download
6. Dalibor DD (2007) Installing, configuring, and developing with xampp
7. XAMPP at https://www.apachefriends. org/download.html
8. Infobright (2010) Data Loading Guide: Revision 2.1
9. Database from www.eclipse.org/birt/documentation/sample-database.php
10. Upgrading your row-based database to a columnar database in Infobright, 2010.

Real Time Multiple Object Detection on Low-Constrained Devices using Lightweight Deep Neural Network

Namdeo Baban Badhe and Vinayak Ashok Bharadi

Abstract Over the past few decades, deep neural networks have gained attention due to their widespread use in filtering, combinatorial advancement, style transfer, pattern recognition, computer, and some different territories. Object detection is one of the demanding tasks in the area of computer vision. There are numerous well-known object recognition approaches that focus on deep neural networks. The popular ones are region-based convolution neural networks (RCNN), single-shot multi-object locator (SSD), and you only look once (YOLO), centered in Area. Each of these models requires huge processing and power consumption, which is a major challenge for implementations of multiple object detection in real-time on resource devices, such as mobile phones or embedded gadgets. The collaborative integration of SSDlite and mobile networks provides better results for the identification and tracking of artifacts on resource-constrained networks. In this paper, the faster_rcnn_inception_v2,ssd_mobilenet_v1_fpn and SSDlite-mobilenetV2 coco models are trained on GPU GTX 1660 GeForce GTX 1660 Ti using MS-COCO dataset and checked on using RabberiPi 3 B + . The parameters are accuracy and frame per second (FPS) are used to compare performance of the models on GPU as well as CPU was tested.

Keywords Object detection · MobileNet · Ssdlite

1 Introduction

Object detection (OD) [1] is the most challenging part of any program for the image processing.OD outputs single and multiple objects with a class mark and its position (usually in the form of bounding box coordinates) given an input image. It is

N. B. Badhe (✉)
IT Department TCET, Mumbai University, Mumbai, India
e-mail: namdeob.badhe@thakureducation.org

V. A. Bharadi
IT Department FAMT, Mumbai University, Ratnagiri, India
e-mail: vinayak.bharadi@famt.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_12

generally helpful to find real-world object instances such as car, bike, TV, flowers, and humans in still images or videos. This can be done by studying the particular features that each object possesses. Some common object detection applications include pedestrian detection in road safety system, self-driving cars, emotional detection of facial expression, Human activity recognition, object tracking in surveillance system, people counting in crowded areas, medical imaging relief and rescue operations, understanding aerial images, etc. The images can have instances of the objects from the same category, separate classes, or no instances. There are many problems present, such as variance in size, dim lighting, occlusion, rotation, low resolution, shape, complex context other factors all will seriously affect the efficiency of object detection. The conventional method of object detection such as background subtraction, temporal differencing, optical flow, Kalman filtering, support vector machine, and contour matching [1] is generally used to manually locate image features. But not all are able to accurately detect objects.

During the past few decades, deep neural networks have pulled in great consideration in view of their wide applications in filtering, combinatorial advancement, style transfer, pattern recognition, computer, and some different territories. Deep neural networks have been shown to be effective in various regions. Deep neural networks in these fields have achieved state-of-the-art efficiency compared to traditional approaches based on manually crafted visual features. The convolution neural network (CNN), a popular deep learning algorithm centered on the artificial neural network, is used to identify visual patterns directly from pixel images with heterogeneity [2]. CNNs are similar to traditional neural networks but have deeper layers. CNNs model temporal and spatial associations widely used for image classification and face recognition. This uses a set of layered functions (i.e., convolution layers) on all possible parts of an input value array to compute the output. In 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [3], Hinton and his student Krizhevsky [4] applied CNN to image classification and obtained top 5 error 15.3 versus 26.2% of traditional system.

2 Related Work

Methods for object detection can be divided into two main types: two stage-methods and one stage-methods. The model first proposes a set of regions of interest (ROI) in two stage-methods and then a classifier processes only the candidates for the region. The other one-stage method skips the proposal stage of the area and identifies possible locations directly over a dense sampling.

Ross Girshick recommended a two-stage object detection using region based convolution neural networks(R-CNN) [2]. R-CNN is a special type of CNN that is fit for finding and identifying objects in pictures .The output is usually a set of bounding boxes closely representing each object detected, as well as a class output for each object detected. The picture below illustrates identification from video stream of single and multiple objects (Figs. 1 and 2).

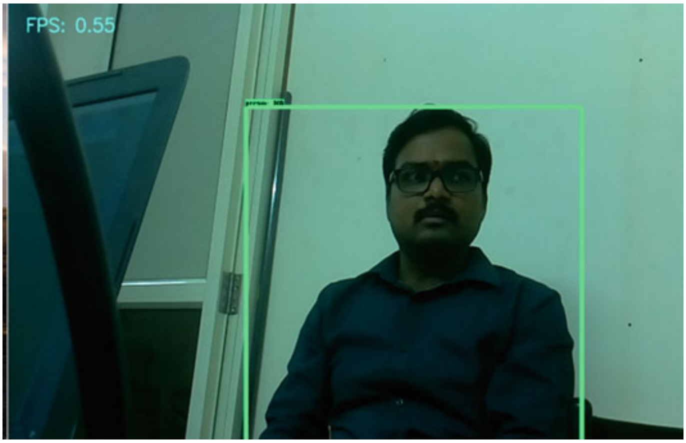


Fig. 1 Single object detection example

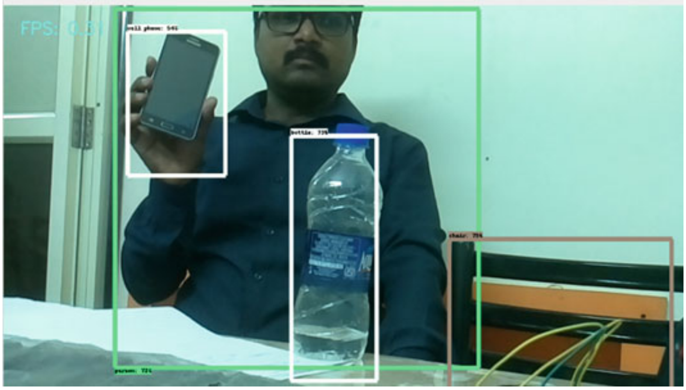


Fig. 2 Multiple object detection example

Region-based CNN models can discover a lot of boxes in the image and test whether any of those crates contain any objects. R-CNNs uses selective search to extract these boxes (these boxes are called regions) from an image. There are a number of challenges that need to be tackled to offer a more reliable system of detection. R-CNNs have problem with the heavy and long datasets. Object detection takes place in two stages region of interest and then real classification and the performance may decrease for unlabeled dataset. R-CNNs do not take a full look at the image. Instead, image sections that have high probability of containing the object. [3] You only look once (YOLO) regression-based model detects objects in a single stage. YOLO has a single convolutional network that predicts the bounding boxes and the class probabilities for these boxes in one evaluation directly from full images. The problem with YOLO model is that it predicts

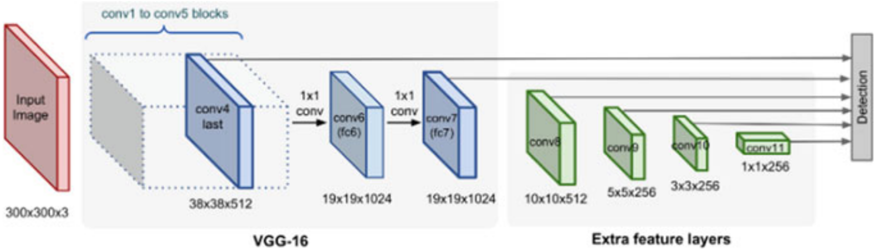


Fig. 3 SSD: single-shot multi-box detector [4]

only 1 type of class in one grid, thus degrading the output with small objects inside the image, for example, detecting a flock of birds.

Single-shot multi-box detector (SSD) was proposed by Wei Liu, Dragomir Angue lov [4]. SSD's primary purpose is to detect objects in video stream data in real time. The SSD detects objects in two steps:

Step 1 Extract useful image features using ImageNet-pre-trained VGG16 architecture and.

Step 2 Detect objects using extra convolution Conv4_3 layer (Fig. 3).

SSD model is a convolutional feed-forward network which produces a fixed-size set of bounding boxes and scores for the presence of object class instances in those boxes, followed by a non-maximum suppression step to produce the final detections. SSD model takes one single shot to distinguish multiple objects inside the image, while R-CNNS requires two shots, one to generate region proposals and one for detecting the object of each proposal. So, compared with two-shot R-CNN versions, SSD is much faster. But there are also some drawbacks of SSD models as it not only confuses objects with similar categories (e.g., animals) but also results in worse performance on smaller objects as they may not appear on all feature maps. The MobileNet model [5] is a lightweight architecture proposed by Google. It is widely used in many computer vision applications such as object detection, face attributes recognition, and landmark recognition, that run efficiently on low end devices with less computing power. This architecture uses depth wise separable convolutions with less number of parameters. MobileNets is built on a simplified architecture that uses separate depth wise convolutions to create deep neural networks of light weight. MobileNet can also be used in modern object detection systems, as an effective base network. MobileNet can also be used in modern object detection systems as an effective base network.

Object detection using region-based CNN, YOLO, and SSD models requires massive processing and power consumption, which is a most important challenge for multiple object detection executions on resource constrained devices, such as mobile phones or embedded gadgets, in real time. A resource constrained scenario means that a computing task must be carried out with a restricted supply of resources, such as processing time, storage space, and battery power. SSDlite is a lightweight model mostly utilized for object location and Mobilenet classifies the object.

SSDlite requires only 2.1M parameters as compared to base SSD model (14.8M) [5] SSD_Mobilenet_v2 is a combination of MobileNetV2 and SSDLite can produce the multiple object detection. MobileNetV2 + SSDLite is 20× more efficient and 10 × smaller while still outperforms YOLOv2 on COCO dataset [6].

3 Proposed Work

The system works in two phases.

Phase No. I Online Phase:

(a) Load and read image dataset
(b) Extract and resize the image dataset
(c) Convert RGB images into grayscale
(d) Train and test the images using COCO-trained models on GPU GTX 1650 Ti server
(e) Store the results into knowledge base.

Phase No. II Offline Phase:

(a) Deploy COCO-trained models on Raspberry Pi B3 version kit.
(b) Load video from webcam capture the image from video stream
(c) Extract the frame from the image
(d) Detect required foreground content from the captured image
(e) Convert RGB images into grayscale
(f) Train and test the images using COCO model
(g) Display the detected objects on screen (Fig. 4).

4 Dataset and COCO Models

As benchmark, we utilized Microsoft COCO Dataset. It includes all 164 K images from COCO 2017 (training 118 K, validation 5 K, test-dev 20 K, test-challenge 20 K). It covers 172 classes: 80 thing classes, 91 stuff classes, and 1 class unlabeled. For learning, we used COCO-trained tensoflow models. A Raspberry 3 Model B kit is used as an edge server where a camera is connected to load a file, and the same model extracts and checks appropriate foreground frames. Raspberry 3 Model B has features a quad-core 64-bit ARM Cortex A53 clocked at 1.2 GHz.We have used tensorflow pretrained models on Raspberry 3 Model B

- faster_rcnn_inception_v2_coco Model [7]
- ssd_mobilenet_v1_fpn_coco Model [7]
- ssdlite_mobilenet_v2_coco Model [8]

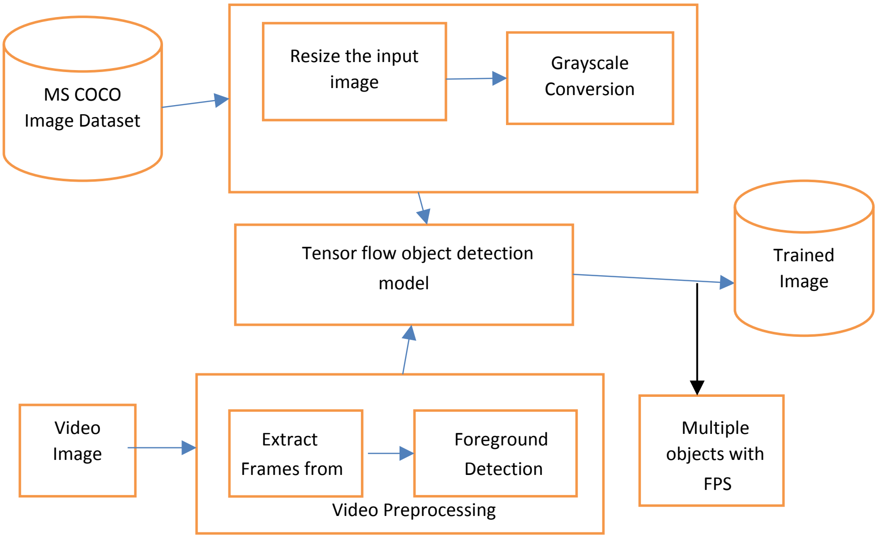


Fig. 4 Object detection system

5 Results and Discussion

The COCO models have tested on Raspberry Pi B3.The system is tested using frames per second (FPS) and accuracy (Figs. 5, 6 and 7).

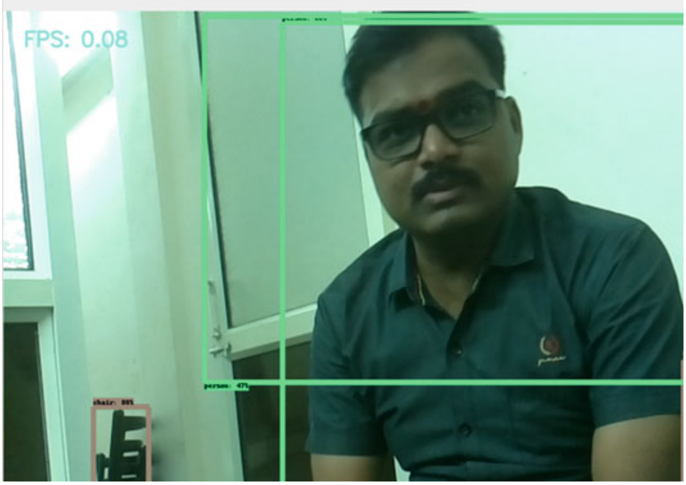


Fig. 5 Faster_rcnn_inception_V2_coco model result

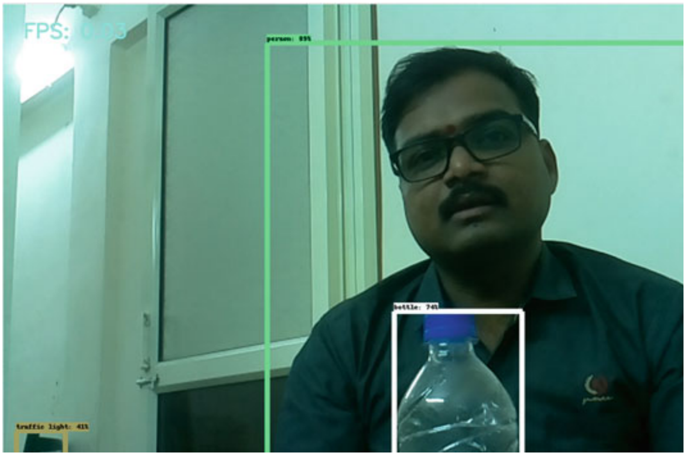


Fig. 6 ssd_mobilenet_v1_fpn_coco model result

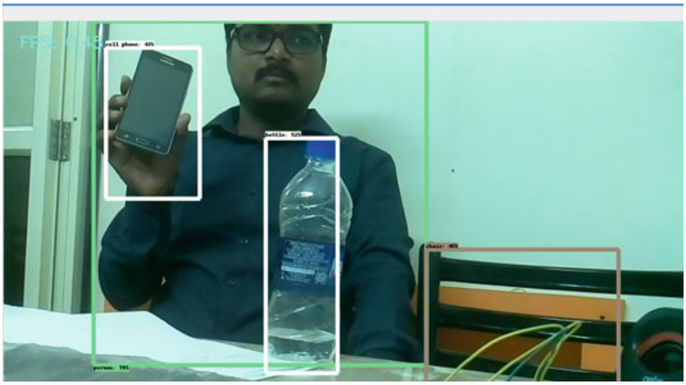


Fig. 7 ssdlite_mobilenet_V2_coco model result

Table 1 Comparison table of COCO model

Model name	Frame per Second (FPS)	Accuracy (%)	Outputs
Faster_rcnn_inception_V2_coco Model	0.08	47	Boxes
ssd_mobilenet_v1_fpn_coco Model	0.03	89	Boxes
ssdlite_mobilenet_V2_coco Model	0.46	78	Boxes

The accuracy of the trained model on COCO database and its respective speed of execution are shown in Table 1. Compared with other versions, the combined version of ssdlite and mobileNet gives high FPS. But less is accuracy.

6 Conclusion and Future Work

We compared three models in this study, and demonstrated their results. Our research indicated that an excellent result is the deep learning approach in object detection tasks. We were able to achieve higher accuracy through deep learning approach. In this research, faster_rcnn_inception_v2, ssd_mobilenet_v1_fpn and SSDlite-mobilenetV2 coco models are trained and implemented on Raspberry Pi 3 B3 kit. This detects multiple single objects in live video, as well. In terms of FPS, model performance on GPU is better compared to low-restricted devices. In the future, due to low processing speed and storage space if we run the model for long time, the model restarts again. In the future, we will focus on how to improve the FPS by reducing the number of layers and normalization techniques. In the future, we will use tensoflowlite models on real-time dataset and will try to get maximum FPS.

Reference

1. Parekh HS, Thakore DG, Jaliya UK (2014) A Survey on Object Detection and Tracking Methods. Int J Inn Res Comput Commun Eng 2(2):2970–2978
2. Girshick R, Donahue J (2015) Region-based convolutional networks for accurate object detection and segmentation. In: 2015 IEEE transactions on pattern analysis and machine intelligence.https://doi.org/10.1109/TPAMI.2015.2437384
3. Redmon J, Divvala S (2016) You only look once: unified, real-time object detection. In: 2016 IEEE conference on computer vision and pattern recognition
4. Liu W, Anguelov D (2016) SSD: single shot multi box detector. In: 2016 international conference on engineering and technology (ICET). https://doi.org/10.1109/icengtechnol.2017.8308186
5. Mark S, Andrew H MobileNetV2: inverted residuals
6. COCO Dataset https://cocodataset.org/#detection-2019
7. Janahiraman TV, Subuhan MSM (2019) Traffic light detection using tensorflow object detection framework. In: 2019 IEEE 9th international conference on system engineering and technology (ICSET). Shah Alam, Malaysia
8. Zhao Z-Q Object detection with deep learning: a review. In: IEEE transactions on neural networks and learning systems for publication

Researcher Framework Using MongoDB and FCM Clustering for Prediction of the Future of Patients from EHR

Harish Barapatre, Yogesh Kumar Sharma, Julli Sarode, and Vaishali Shinde

Abstract Biomedical engineering field is one of the most important research areas for patient diagnosis and prediction of diseases using old history of various patient information. Data collection and data analysis models changed business trends over the past few years. By using BigData analytics, we can predict the effects of drugs and how drugs develop disease on mankind. Many machine learning algorithms like cluster computing environment, classification, etc., analyze the content of healthcare. The proposed framework has developed C-means clustering algorithm in developing biomedical engineering applications. The data is collected from machine learning repository. BigData framework MongoDB database is used to analyze the data. Here, we have modules like doctor, administrator, and analyst. By using machine learning and BigData, the analyst module identifies the chronic disease and help in reducing the medical expenses. Doctor by observing the predicted data symptoms predicts related disease. Administrator's role is to add or remove the users in the database. The researcher finds the patient parameter by using fixed, alphanumerical, variable field data, or combination data which design the exact query which executed on MongoDB server and shows the search result that result has processed by using modified FCM cluster, and finally, we have calculated the accuracy on the basis of search parameter which always helps the researcher for diagnosis of the patient history.

Keywords MongoDB · FCM · BigData

H. Barapatre (✉) · J. Sarode · V. Shinde
Department of Computer, Mumbai University, Mumbai, India
e-mail: harishkbarapatre@gmail.com; computerscience@jjtu.ac.in

J. Sarode
e-mail: julli.sarode01@gmail.com

V. Shinde
e-mail: vaishalishinde22@gmail.com

Y. K. Sharma
Department of Computer, JJTU University, Rajasthan, India

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_13

1 Introduction

BigData is widely used in biomedical engineering domain in the recent trends. The research studies have been carried out in BigData analysis so as to work on the domains like biomedical engineering. Before discussing all this, it has to be discussed about the reason why we are choosing BigData [1–4]. This paper highlights supervised learning techniques where there is a need of training data that is acquired with training examples. Supervised learning is a machine learning techniques which always maps the input with the output basing on the examples that were provided to it. Supervised learning always works with supervisor or trainer that we provide to it. Supervised learning builds model, and later, it evaluates it with predictions as base. It trains the input data with the given examples using the stated algorithm and make reasonable predictions. So as to solve problem of supervised learning, the following steps are performed:

Step 1 Initially, we have to decide the type of training examples.
Step 2 Secondly, there is need for collecting a training set.
Step 3 Determine the learned function for the input representation.
Step 4 Learned function structure is determined with corresponding learning algorithm.

The above-mentioned research paper, Researcher Framework using MongoDB and FCM clustering for Prediction of the Future of Patients from EHR, is said to help the patients.

2 Objectives

The main objective of this masters research project is to examine different clustering algorithms in order to detect groups in a real-world, high-dimensional EHR dataset. The goal of the work is to find algorithms that detect high-quality clusters accuracy within a reasonable computational time. In order for the chosen algorithms to scale with an increasing size of the patient sample size and/or features, it is desirable. To find an algorithm that achieves high-quality accuracy with minimal preliminary processing of the data, the choice of methods for the identification of groups in the dataset imposes challenges due to several aspects:

High-dimensionality: The dataset originally has about 1200 features, which constrains the choice of algorithms and evaluation measures.

Missing data: The set contains missing values that are not missing at random.

Accuracy: Data is collected through health questionnaires, and the accuracy of this self-reported data is unknown.

3 Literature Survey

In recent years, several major breakthroughs in medical science and engineering are prompting a convergence between the healthcare industry and engineering industry. These have rapidly led to more collaborative relations among patients and their physicians. BigData analytics (BDA) has a key role in developing those improved relations [5]. Healthcare professionals and pharmaceutical companies now have the capability to explore and analyze data not only for an individual, but also for an increasing number of patients in specific population studies [5]. By combining the health science with computing and engineering, a new dimension is added within the domain of biomedical engineering [5]. Researchers can not only predict the risk of disease formation in the human body, but also with the help of healthcare data, they can provide solutions. Applying machine learning techniques (clustering, regression, classification, etc.) in biomedical engineering, researchers are now able to identify when cell damage or tumor will occur in human body. Engineering professionals have been working in healthcare sectors for several years, and it is well established that clinical and engineering researchers share a similar mind-set when it comes to develop a new biomedical system which involves medical and engineering knowledge [6]. Both in conventional engineering and in medicine, the area is defined in terms of problem solving. According to [6], the motto in both disciplines is whatever works. But there is a vital difference: engineering professionals deal with problems related to phenomena on computing and physics. If reliable theory is not available, then alternatively engineers use empirical models, as far as they can solve the problem at hand. Due to this, they are left with a sense of fragility and mistrust, and they try to replace them as soon as possible with theory-based mechanistic models, which are both predictive and explanatory. Healthcare professionals and clinical researchers deal with problems for which there is a much less well-established body of knowledge. Moreover, this knowledge is often qualitative or semi-quantitative and obtained from highly controlled clinical experiments. By including BigData in biomedical engineering with a new generation of technologies and architectures, the healthcare research and development domain has been shifted a lot [7]. It revolutionized the medical device development techniques and disease predictions using machine learning and artificial intelligence. Now, researchers have a very good grip on both the data storage demands, and the hefty server processing required analyzing large volumes of clinical data in a protected manner [8]. According to [8], clinical data is unstructured information and is not classically easy for traditional databases to analyze. Therefore, the predictive power of BigData has been explored recently in healthcare field. According to [9], BigData-related computing tools are improving the biomedical engineering research. These computing tools gather knowledge from unstructured clinical data, and then, it uses mathematical algorithms for analyzing. For example, data repositories are created to help the doctors and patients at the same time. Patients suffer from diseases such as cancer which can find the appropriate drug for their disease type. The tool is called My Cancer Genome, and it was developed by a group of researchers at Vanderbilt University in the USA [9]. Machine learning is considered as a sub-discipline of artificial intelligence (AI). It focuses on algorithms those capable of learning and/or adapting their structure based on a set of experiential data [10]. Machine learning has been the subject of great interest in the biomedical community because they offer potential for improving the understanding the diagnosis of disease. Also, it helps the healthcare professionals to decide what to do based on the output from the system that uses machine learning features [10, 11]. The significance of the impact of machine learning is much greater than ever given the increase in biomedical data. Machine learning can provide new tools for interpreting the complex datasets with which the clinician is challenged. Machine learning in biomedical engineering requires development and analysis of linear methods for supervised and unsupervised feature extraction and classification [4]. Linear methods become very popular among the healthcare professionals as it is easier to analyze and make decisions. On the other hand, nonlinear methods can be difficult to interpret sometimes [4].

4 Existing System

Indeed hierarchical clustering algorithms generally seems to perform better than partitioned methods in creating high-quality clusters. As opposed to partitioned methods, they are also able to discover nonspherical clusters. Hierarchical clustering algorithms also offer a good visualization in the form of a dendogram, which is suitable for interpretation. The hierarchical methods has the main drawback that the time complexity is quadratic O(n2). Because of the popularity in literature and widespread use of the two traditional methods, hierarchical clustering and C-means clustering will be of interest in this study. Although our high-dimensional dataset might implicate lengthy running times for the hierarchical method, leading to lengthy experiments, it compensates this drawback through producing high-quality clusters, which is of main interest for this work.

5 Problem Definition

Effective prediction of clinical risks of ACS patients via their heterogeneous EHR data is still an intricate problem and remains a major challenge for health care management, mainly due to high clinical complexity and the natural heterogeneity in EHR data. Therefore, one of the most important tasks in clinical risk prediction is to develop robust prediction models that can effectively handle high-dimensional heterogeneous EHR data and accurately classify different clinical risks levels based on the acquired EHR data. It has particular advantages such as rapid inference and the ability to reconstruct features (a.k.a. clinical risk factors) yielding good classification accuracy. In order to ensure that features reconstructed by FCM are useful to the HER in problem for each patient, two regularization constraints are added on FCM to capture characteristics of patients at similar risk levels to make the reconstructed features of patients within the same risk level as close as possible and preserve the discriminating information across different risk levels to make the reconstructed features of patients as separated as possible. After this, we append FCM clustering layer on the top of the resulting reconstructed feature representation layer, which is tailored to the EHR problem. Our proposed model learns more discriminative patient feature representations and thus improves the performance of EHR dataset.

6 Role of BigData and Machine Learning in Biomedical Engineering

BigData and machine learning helped biomedical engineering in many aspects [1–4, 10, 12, 13]. Sajid et al. [1] artificial intelligence came into existence from the study pattern recognition and computational learning theory [2]. Machine learning creates algorithms that learn from many examples, and it makes predictions on the data [3]. Machine learning algorithms were trained to identify the disease basing on symptoms. It helps to diagnose Parkinson's disease based on dysphonia measures. It also provides a review of recent researches on brain tumor, etc. [4]. Machine learning techniques are used for identifying retinal image analysis [12]. Machine learning algorithms supported in finding obstructive sleep apnea detection using neural network with the help of deep learning framework [10]. Machine learning algorithms help in detecting cell damage and tumors [13]. Researches not only detects the disease but they also try to provide remedies to the concerned disease.

7 Mathematical Notation

Defination 1: Accuracy of C-means Clustering Algorithm Calculation Formula
Accuracy can be calculated as below

$$\text{Acc} = (\text{Pc}/\text{Ds} - \text{Pc}) * 100$$
$$(\text{Accuracy} = \text{predicted class}/\text{dataset})$$

Acc Stands for accuracy.
Pc Stands for correctly predicted class.
Ds Stands for dataset.
NM Not match user variable.

Dunn was the first to implement fuzzy C-means clustering algorithm in 1973, later Bezdek improved this algorithm in 1981. For fuzzy C-means, the membership function is represented by matrix (U) having values between 0 and 1 that represents the membership of data points to every cluster, while hard C-means uses only 0 and 1 values for membership function. Fuzzy C-means algorithm is based on an objective function.

8 Proposed System

The proposed framework contains modules like administrator, doctor (clinician), and analyst/researcher. The administrator role is to add or remove users, i.e., update user information. The doctor's module role is to name the disease and their symptoms. All the disease regarded information are stored in database. Analyst module role is to provide the parameter values for the analysis, and it has to apply relevant C-means algorithm to the provided data. The parameters can be like names, dates, gender, or age (Fig. 1).

Analyst after choosing required parameter, he/she can select the required representation method to get the desired output. Proposed framework where the BigData

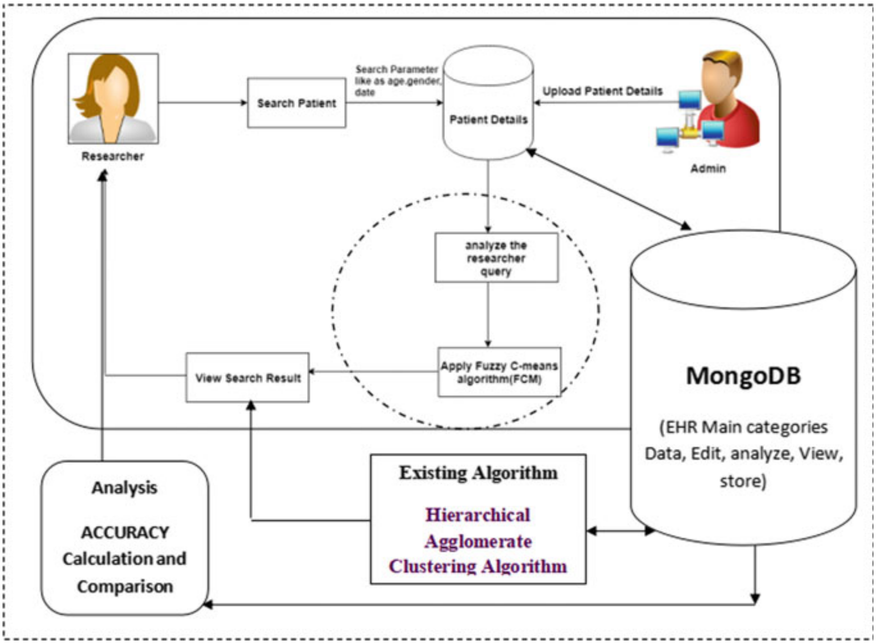

Researcher
Search Patient
Search Parameter like as age,gender, date
Patient Details
Upload Patient Details
Admin
analyze the researcher query
Apply Fuzzy C-means algorithm(FCM)
View Search Result
MongoDB
(EHR Main categories Data, Edit, analyze, View, store)
Analysis
ACCURACY Calculation and Comparison
Existing Algorithm
Hierarchical Agglomerate Clustering Algorithm

Fig. 1 Framework of the system

trained with machine learning using fuzzy C-means algorithm. Hierarchical clustering iteratively joins the two nearest clusters beginning from singleton clusters or with the complete set. After joining of two clusters, distance between all the other clusters and a new cluster that is joined is recalculated.

8.1 Patient Data

The hospital dataset used in this study contains real-life hospital data, and the data are stored in the data center. To protect the patient's privacy and security, we created a security access mechanism. The data provided by the hospital include EHR, medical image data, and gene data. The inpatient department data is mainly composed of structured and unstructured text data. The structured data includes laboratory data and the patient's basic information such as the patient's age, gender, and date. While the unstructured text data includes the patient's narration of his/her illness, the doctor's interrogation records diagnosis, etc. In order to give out the main disease which affects this region, we have made a statistics on the number of patients, the sex ratio of patients, and the major disease in this region every year from the structured and unstructured text data.

8.2 Patient Details Prediction

We obtain the main cancer disease in this region. The goal of this study is to predict whether a patient is among the cerebral infarction high-risk population according to their medical history. More formally, we regard the risk prediction model for cerebral infarction as the supervised learning methods of machine learning, i.e., the input value is the attribute value of the patient, $A = (a_1, a_2, \ldots, a_n)$ which includes the patient's personal information such as age, gender, the prevalence of symptoms, and other structured data and unstructured data.

8.3 Algorithm Accuracy

The accuracy determines and predicts the measure of how machine learning classification algorithm is working correctly. It finds the number of true corrected documents and calculates the number of cluster which we get correct result of the researcher. Machine learning (ML) algorithms help software applications to predict outcomes accurately. Machine learning follows the same strategies of data mining and predictive modeling. Machine learning is widely used in many domains like fraud detection, network security threat detection, etc.

Defination 1: Fuzzy C-means Clustering Algorithms Steps Are:
In cluster, populations or data points are divided into different groups. Data points of same group are similar, and data points of dissimilar group are not similar. C-means clustering algorithm works by assigning membership to each data point. The member ship to data point is given basing on data point distance with cluster center. If the data point is nearest to cluster center its membership is more. We can conclude that the data point nearest to cluster center serves as prototype of cluster.

Fuzzy C-means Clustering Algorithms Steps Are:

1. Import the all dataset and preprocess also stored on Mongo server.
2. Compute the searching parameter and contract the query with different dimension for calculating distance from one cluster to another cluster.
3. Sort all documents as dimensionwise and form a group based on new dimension. Under one dimension, it consists of collection of records called cluster. We have called the different function on the basis of different cluster
4. Compute the distance between the clusters and find center of cluster.
5. Also calculate the predicted and non-predicted documents which are executed by cluster.
6. Reused the cluster center using member function and calculate the accuracy
7. Repeat step 2 to 6, until the distance is zero.
8. Exit.

9 MongoDB

1. The storage pattern in MongoDB is document-oriented storage. Here, the data is stored in the form of JavaScript Object Notation (JSON) pattern style documents.
2. *MongoDB uses* multikey *indexes*. The *indexing of multikey index is used to store* the content in arrays. *Field is indexed* that holds an array value.
3. *MongoDB* creates variable *index* entries for each and every element of the array. Multikey *indexes* allow queries to select the documents which contain arrays by matching. A replica set in MongoDB is a group of MongoDB processes that can maintain the same dataset. Replica sets in MongoDB provides redundancy and high availability.
4. Sharding feature of MongoDB is the process of distributing data across multiple servers for storage purpose. MongoDB software uses sharding to manage huge massive data growth. Sharding feature in MongoDB allows adding many servers to our database to support the data growth. MongoDB automatically balances data and load across its various servers.

9.1 Advantages of MongoDB Over RDBMS

MongoDB database is a document-based schema less in which database where there is collection that holds variety of documents. The number of fields, the content, and size of document differ from one another. Structure of object in MongoDB is very clear. MongoDB database have no complex joins to work with. MongoDB supports dynamic queries when we are working with documents. This is called document-based query language which is as powerful as that of structured query language. MongoDB is much easier to scale. There is no need of object to database mapping in MongoDB. MongoDB uses internal memory for storing the working set and that helps in fast access of data.

10 Result Analysis

(1) This C-means algorithm gives best results for overlapped datasets when compared with k-means algorithm.
(2) In k-means algorithm, data points are assigned to single cluster center. In c-means algorithm, data points are assigned to its membership to more than one cluster.

The search result of gender parameter using FCM are shown in Fig. 2.

Figure. 2 analysis said that the researcher research on gender parameter, it may be male or female, while we select any one the six types of cluster was created based on gender. The six types of cancer category were created, and each category consists of number of documents which was processed by FCM.

Here, the researcher constructs the query which was executed by MongoDB server with variable parameter. The query was processed by the FCM and analyzed all fields like age between or near, name, or gender which was constructed by the researcher, and researcher helps those contents for the prediction of patient history. The analysis graphs are shown in Figs. 2 and 3. After successfull analysis, researcher got the

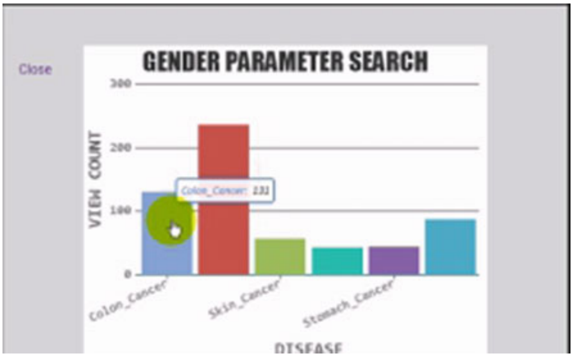


Fig. 2 Analysis with gender parameter

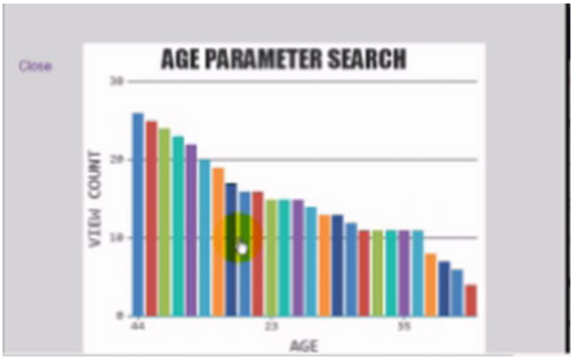


Fig. 3 Analysis with age parameter

Table 1 Data set input parameter

Input detail			Accuracy	
Dimension	Parameter	Prediction	HAM	FCM
Age	MIN:10 and MAX:40	384	38.3	62.1
Gender	Male	401	40.9	66.4
Date	S:30/01/2015 E: 30/03/2020	147	17.2	36.7
Age and gender	MIN:10 and MAX:40 and Male	136	NS	48.6
Age and date	S:30/01/2015 E: 30/03/2020 and MIN:10 and MAX:40	103	NS	35.0
Gender, date and age	Male, date, and age	87	NS	20.9

accuracy details file which can be downloaded by researcher and analyzed which cluster shows the highest number of documents. Each category of age is treated as cluster that means it consist of number of documents which processed by group of cluster.

10.1 Input/Output Description

Table 1 describes the different kinds of parameter as input [14], and when we apply on specified documents now, predicted results were shown with accuracy. Hierarchical was not supported on two parameter and three parameter searching techniques.

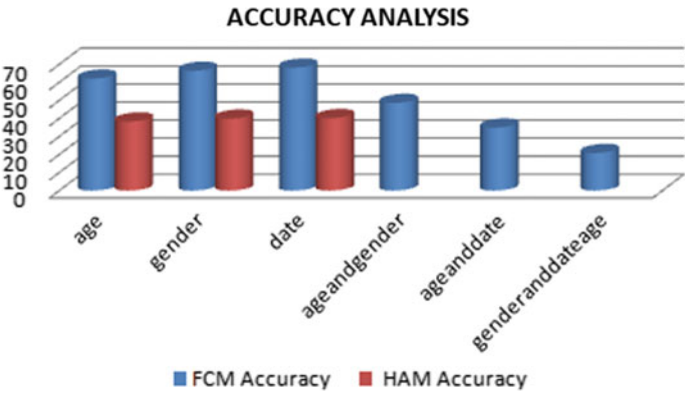


Fig. 4 Accuracy analysis

The overall system analysis graph is as shown.

In the fixed variable searching, we got the documents regarding this fixed query of not getting exact prediction from EHR also, but when we compare the existing system, we got best accuracy of documents processing as compared to existing system. The various existing system does not support the two or three parameter searching method, but our proposed system support one, two, and three parameter searching and shows exact prediction and improves 30 percent accuracy which is shown in the accuracy analysis which was very helpful for the researcher and doctors for the prediction of research from EHR dataset.

11 Conclusion

The aim of this paper was to propose a framework using BigData and FCM clustering method to select drugs for patients. Biomedical engineering applications are trained with FCM clustering algorithm. This paper helps the researcher for doing various kinds of research and prediction from existing database. Here, we have used the one, two, and three parameter searching techniques which processed on number of clusters using FCM and HAM cluster algorithm. FCM shows the more accurate results as compared to hierarchical, and also by using group of cluster, we have improved 70% accuracy as treated as single cluster. **MongoDB** is a NoSQL database **used** for high volume data storage, and it is document oriented which gives high data security; here, in this research, C-means algorithm is taken for training purpose as it works faster. FCM algorithm assigns its values to multiple clusters with accuracy calculation, and finally, we compare this result with hierarchical algorithm. Our proposed system achieves the highest accuracy as compared to existing system and gives accurate prediction of patient for the researcher.

Future Work In the proposed framework, the work has been done on C-means clustering algorithm. So, selecting C-means made the programming part simple. In future, this above-said framework will be developing healthcare data solutions, healthcare intelligence, data accuracy, etc.

References

1. Sajid I, Khan UG, Saba T, Rehman A (2018) Computer-assisted brain tumor type discrimination using magnetic resonance imaging features. Biomed Eng Lett
2. Chen M, Hao Y, Hwang K, Wang L, Wang L (2017) Disease prediction by machine learning over big data from healthcare communities. IEEE Access 5:8869–8879
3. Tomasev N, Radovanovi M (2016) Clustering evaluation in high-dimensional data in unsupervised learning algorithms. Springer, Cham, pp 71–107
4. Auffray C, Balling R, Barroso I et al (2016) Making sense of big data in health research towards an education plan. Genome Med 8(1):71
5. Lupton D, Jutel A (2015) A critical analysis of self-diagnosis smart-phone apps . Soc Sci Med 133:128–135
6. Mao R, Xu H, Wu W, Li J, Li Y, Lu M (2015) Overcoming the challenge of variety: big data abstraction. The next evolution of data management for all communication systems. IEEE Commun Mag 53(1):42–47
7. Costa FF (2014) Big data in biomedicine . Drug Disc Today 19(4):433–440
8. Claeys OF, Dupont M, Kerckhove T, Verhoeve W, Dhaene P, Turck D (2013) A probabilistic ontology-based plat form for self-learning context-aware healthcare applications. Expert Syst 40:7629–7646
9. Pedregosa F et al (2011) Scikit-learn: machine learning in python . J Mach Learn Res 12:2825–2830
10. Mishra NK, Celebi ME (2016) An overview of melanoma detection in dermoscopy images using image processing and machine learning. arXiv: https://arXiv.com/1601.07843
11. Houle ME, Kriegel HP, Kroge P (2010) Can shared-neighbor distances defeat the curse of dimensionality. In: Proceedings of SSDBM, pp 482–500
12. Eskofier BM, Lee SI, Daneault JF et al (2016) Recent machine learning advancements in sensor-based mobility analysis. In: IEEE 38th annual international conference of the deep learning for Parkinson's disease assessment in Engineering in Medicine and Biology Society (EMBC). IEEE, pp 655–658
13. Yao Q, Tian Y, Li PF, Tian LL, Qian YM, Li JS (2015) Design and development of a medical big data processing system based on hadoop. J Med Syst 39(3):23
14. Mishu MM (2019) A Patient oriented framework using big data & C-means clustering for biomedical engineering applications. In: 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). Dhaka, Bangladesh, pp 113–115. https://doi.org/10.1109/ICREST.2019.8644276

Performance Level Evaluation of Cryptographic Algorithms

Ashwini P. Parkar, Madhuri N. Gedam, Nazneen Ansari, and Shanthi Therese

Abstract Cryptographic Algorithms persuade a significant contribution in data security. But they require notable portion of resources like processor time, system memory, power consumption, latency time etc. So, it is needed to find out best cryptographic algorithm by evaluating performance of each of them. This paper evaluates the performance of cryptographic algorithms namely AES, triple DES, Blowfish, RSA, MD5 as well as SHA based on parameters like time required to encipher, time required to decipher and memory used by them. The literature surveys of these algorithms are made and then the system for the measurement of performance is proposed and implemented.

Keywords Cryptographic algorithms · AES · RSA · MD5

1 Introduction

Most of the systems are under the threats of attacks by the attacker as the widespread use of internet and electronic communication. For the development of most of the systems, we need to maintain confidentiality, authentication and integrity of

A. P. Parkar (✉)
Department of Computer Engineering, SLRTCE, Mira Road, Mumbai, India
e-mail: parkarashwini14@gmail.com

M. N. Gedam
Department of Computer Engineering, Veermata Jijabai Technological Institute, Matunga, Mumbai, India
e-mail: madhuri.gedam@gmail.com

N. Ansari
Department of Information Technology, SFIT, Borivali, Mumbai, India
e-mail: nazneenansari@sfit.ac.in

S. Therese
Department of Computer Engineering, TSEC, Bandra, Mumbai, India
e-mail: shanthitherese123@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_14

the system using various methods for data security in modern digital communication [1, 2]. One of them is ‘Cryptography’. Cryptography is broadly stratified into symmetric/secret key cryptography and asymmetric/public key cryptography and hashing [3]. An encryption system in which the data transmitter and recipient share a single shared key that is used to encipher and decipher a piece of information is known as symmetric key cryptography. Lengthy key is difficult to break than shorter key. Asymmetric key cryptography is the process implemented using public and private keys to encipher and decipher data respectively [4]. So symmetric algorithms are less complicated to use than the asymmetric algorithms. Hashing is a one way function of cryptography that helps to change piece of data into a distinctive form. It is irreversible. It gives a unique fingerprint for unique data. Errors can be detected with the help of hashing because if one bit in data is changed the hash value will also be varied. Most common hashing functions are MD5, SHA1 [5, 6].

The present work evaluates and measures the performance of symmetric (AES, 3DES, BLOWFISH) and asymmetric (RSA) and hashing (MD5, SHA-1) cryptographic algorithms. The comparison has been conducted for text files on three parameters such as enc time, dec time and memory usage. Many researchers are working on the improvement of various available cryptographic techniques.

In Sect. 2 of this paper, the literature review has been discussed. Section 3 gives overview of cryptographic algorithms methodology. Implementation and result analysis discussed in Sect. 4, performance discussion done in Sects. 5 and 6 gives final result of analysis which demonstrates the efficiency of each algorithm.

2 Literature Review

Many researchers have compared different types of cryptographic algorithms to analyze performance level of each of them.

Proposed a system to see the effects on performance of the AES and Blowfish algorithms by changing the input data of different sizes and types [7]. They implemented the system using Python, Pycrypto package. From the result, they have shown that the speed of Blowfish slower than AES with a difference of 200–300 ms and the regression slope for the Blowfish more than the AES for different range of ASCII values. Blowfish algorithm enciphered data greater in size than the AES.

Analyzed DES, 3DES, AES and Blowfish on the basis of average encryption time, average decryption time, memory usage, power consumption, etc. [8]. Finally, came to the conclusion that, the best in all parameters is Blowfish, AES was the second best algorithm but the 3DES, the most inadequate algorithm among them.

References [4, 9] concluded that the RSA takes more encryption time than decryption time implemented the program using JAVA library functions [9].

Proposed a system using C language to show the comparison between the MD5 and the SHA-1 algorithm [6] . The comparison made on collision rate and the length of hash value. They analyzed that the MD5′s fingerprint is 32 bits greater than the SHA-1. It shows SHA-1 algorithm is more fortified. But the SHA-1 takes 80 steps and the MD5 requires only 64 steps. So, the SHA-1 takes more time than the MD5 on the same system. So, for small data we can select the MD5 rather than SHA-1.

3 Overview of Algorithms

As shown in Fig. 1, different types of cryptographic algorithms include symmetric, asymmetric and hashing.

3.1 Symmetric Key Cryptography Algorithms

DES. As shown in Fig. 2 Input of 64 bits clear text goes to DES, with the key length 56 bits, gives output of 64 bits cipher text [10].

DES uses two methods of classical cryptography which includes substitution and transposition techniques. It incorporates 16 rounds (steps). Each round performs substitution and transposition [11]. The rounds in DES are as follows:

1. 64-bit plain text goes through Initial Permutation function.
2. The Initial Permutation is performed.
3. Two halves of permuted blocks LPT and RPT are generated.
4. Each of them has to go through 16 rounds of encryption.
5. Before Final Permutation LPT and RPT are rejoined
6. In the last step 64 bit cipher text is produced.

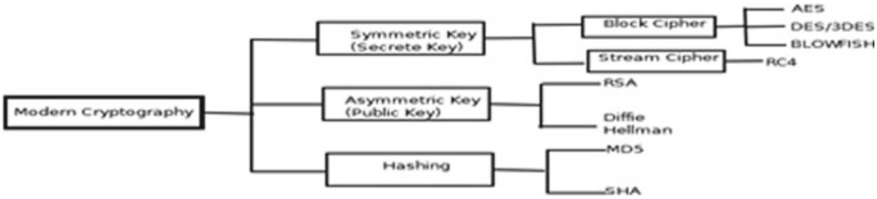


Fig. 1 Types of cryptographic algorithms

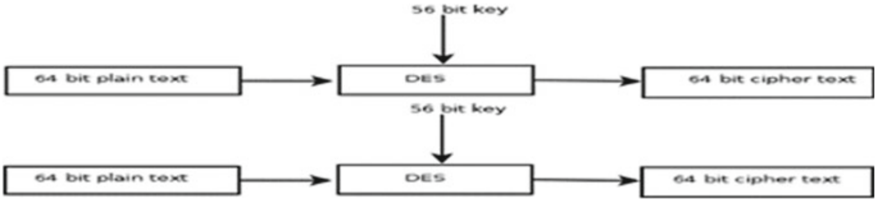


Fig. 2 Working of DES

3DES. DES is insufficient to encipher data. The key size of DES is extended for 3DES by applying the algorithm 3 times with 3 different keys to improve the security. 168 bits (3 * 56 bits) is combined key size. Iterations are repeated 3 times to increase the encryption level [12].

AES. Advanced Encryption Standard is 6 times faster than 3DES. Clear text block of 16 bytes forms a matrix organized in 4 rows and 4 columns. Key size decides the number of rounds [12] i.e. ten rounds for 128 bits key, twelve rounds for 192 bits key and fourteen rounds for 256 bits key. Each of round involves four steps:

Substitution bytes—Performs byte by byte substitution.

Shift rows—A simple permutation.

Mix column—Output of previous step is multiplied by the matrix of algorithm.

Add round key—The key is XORed with piece of data.

BLOWFISH. BLOWFISH uses 16 rounds with block size 64 bits and size of key spans from 32 to 448 bits for encryption. 3 steps in Blowfish [13].

Key expansion-Keys should be computed before encryption and decryption. The P-array consists of 18, 32 bit sub keys. $P1 \ldots P18$.

Data Encryption-Blowfish Encryption Steps:

Divide 64 bit into two 32 bit halves Lx, Rx.

For $I = 1$ to 16.

$Lx = Lx$ XOR Pi.

$Rx = F(XL)$ XOR Rx.

Swap Lx and Rx.

Swap Lx and Rx (Undo last swap).

$Rx = Rx$ XOR $P17$.

$Lx = Lx$ XOR $P18$.

Rejoin Lx and Rx.

Data Decryption-As similar as encryption process but reverse order used for $p1 \ldots p18$.

Table 1 shows the overall comparison of symmetric key cryptography algorithms using dimensions like size of key and block of data, security, number of rounds, encryption/decryption time and applications.

3.2 Asymmetric Key Cryptography Algorithms

RSA. Asymmetric key cryptography Rivest, Shamir, Adelman popularly known as RSA algorithm implemented with two different keys, a shared key and a secrete key to encipher and to decipher data respectively. Key size varies from 1024 to 4096 bits.

Using two distinct prime numbers shared key and secrete keys are generated [1]. It works in three steps.

1. Generates key
2. Process to encipher
3. Process to decipher

Table 1 Comparison between symmetric key algorithms [1, 7, 8, 13]

Developed in year/by	Size of key (bits)	Security	Size of input block (bits)	Rounds	Encipher/decipher time	Applications
1974/IBM	56	2^{56}	64	16	High/medium	Smart cards, Sim cards
1998/IBM	112 or 168	2^{168}	64	48	High/medium	Electronic payment industry, BlackBerry Enterprise Server
1998/Vincent Rijmen, Joan Daemen	128, 192, 256	2^{128}, 2^{192}, 2^{256}	128	10, 12, 14	High/high	HTTPS, FTPS, OFTP
1993/Bruce Schneier	32–448	Upto 2^{448}	64	16	High/low	Linux

3.3 Hashing Algorithms

MD5. Message Digest5 designed by Rivest. It is improved version of MD series algorithm has 128 hash values. It provides prevention from data modification or alteration. It generates Message Digest (fingerprint) through MD5 algorithm. One of its applications is password security [6]. Steps in MD5 shown in Fig. 3.

SHA-1. Secure Hash Algorithm 1 produces 160 bits hash value known as message digest or fingerprint. It is used in cryptography for data integrity IPSecurity, Secure Socket Layer, Pretty Good Privacy Protocols [14, 15] (Table 2).

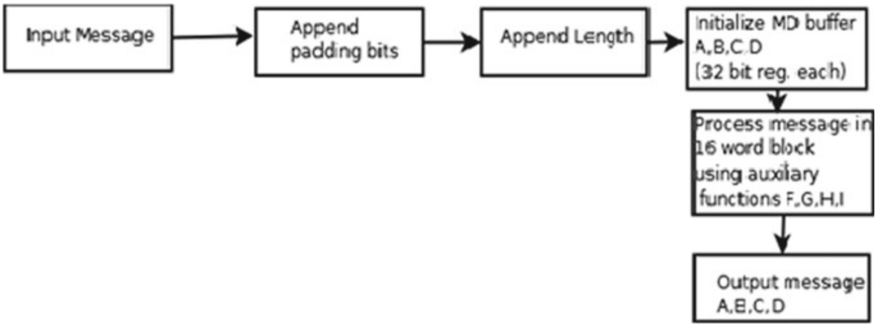

Input Message
Append padding bits
Append Length
Initialize MD buffer A,B,C,D (32 bit reg. each)
Process message in 16 word block using auxilary functions F,G,H,I
Output message A,B,C,D

Fig. 3 Steps in MD5

Table 2 Summary of hashing algorithm [6, 16]

Name of algorithm	Developed by/in year	Input block size	Rounds	Memory limit (bits)	Hash code size
MD5	Ronald Rivest/1992	512	4	264	128
SHA 1	NSA/1995	512	80	264	160

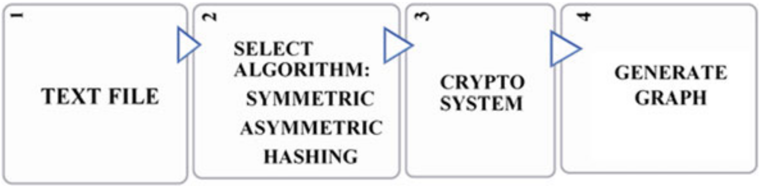


Fig. 4 Operational block diagram

4 Proposed System for Measurement of Performance

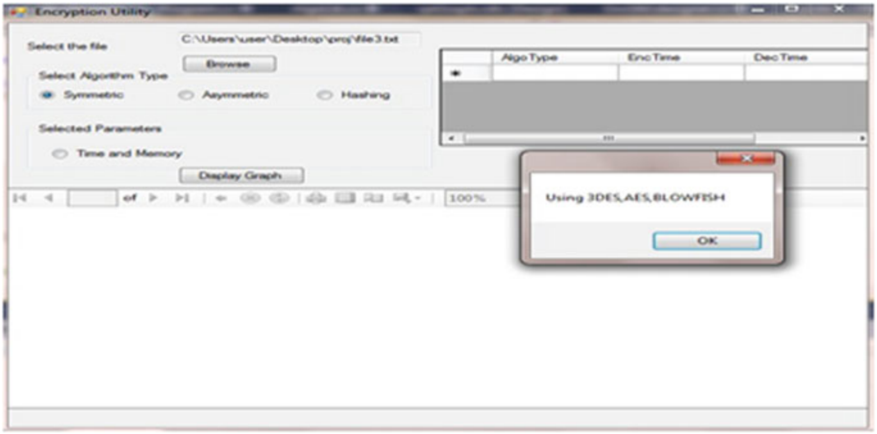
Figure 4 shows the operational diagram of the system to be implemented. Cryptographic algorithms gathered from reliable sources for analysis using three phases. Text file used as an input. In phase one the commonly used symmetric key cryptography algorithms compared and presented using bar graph based on parameters like enc time (time for encryption), dec time (time for decryption) and memory usage. In phase two the asymmetric key cryptography algorithm RSA analyzed and compared with respect to symmetrical key algorithms on the parameters mentioned above. In phase three the hashing algorithms analyzed on the different parameters like length of hash value, execution speed.

5 Implementation and Results

The simulation implemented on a laptop with Windows 32 bit, Intel® core™ 2 Duo CPU 2.00 GHz with 3 GB RAM, C# language using MS Visual Studio 2010 choice for implementation which uses Security Cryptography Package. The performance comparison was conducted on different algorithms mentioned above using text file. The parameters are

- Enc Time-Time taken for encryption
- Dec Time-Time taken for decryption
- Memory Usage-Memory taken by the algorithm.

Figure 5 shows the options of cryptographic algorithms to be compared. We can select any option that we want to compare.


Encryption Utility

Select the file C:\Users\user\Desktop\proj\file3.txt

Select Algorithm Type ● Symmetric ○ Asymmetric ○ Hashing

Selected Parameters ○ Time and Memory

Algo Type	Enc Time	Dec Time

Using 3DES,AES,BLOWFISH

Fig. 5 Select the algorithms to be compared

Figure 6 shows the encrypted data of text file upon selection of algorithm.

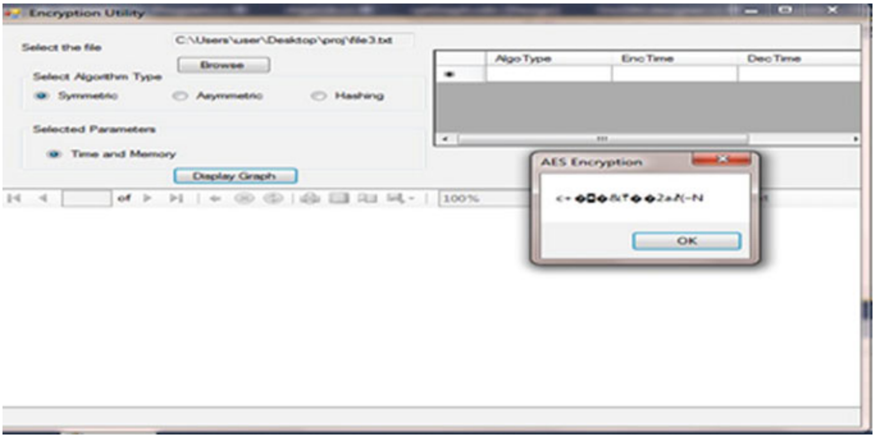
Figure 7 gives the graph of result and table of comparison for different cryptographic algorithms.

Figures 8, 9 and 10 shows the graph for symmetric, asymmetric and hashing algorithm on the basis of parameters like Enc, Dec Time and memory usage.

We concluded enc, dec time and memory usage of cryptographic algorithms in Table 3.

Brief analysis of system results given below:

Result analysis of phase one concluded that Blowfish algorithm took minimum amount of enc time and 3DES the most. Blowfish also consumes least dec time.3DES took the larger dec time. If we consider both encryption time and decryption time


Encryption Utility

Select the file C:\Users\user\Desktop\proj\file3.txt

Select Algorithm Type ● Symmetric ○ Asymmetric ○ Hashing

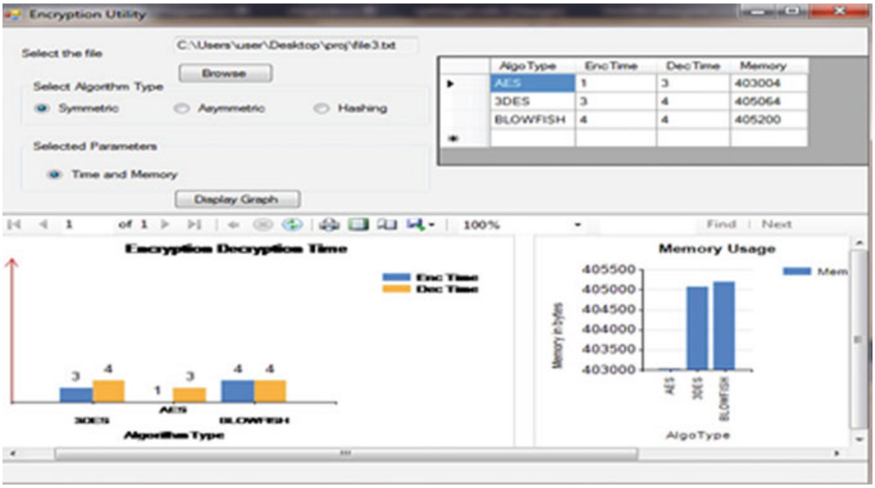
Selected Parameters ● Time and Memory

Algo Type	Enc Time	Dec Time

AES Encryption

<•◆□◆&T◆◆2aÅ(–N

Fig. 6 Display the encrypted text



Encryption Utility

Select the file: C:\Users\user\Desktop\proj\file3.txt

Select Algorithm Type: ● Symmetric ○ Asymmetric ○ Hashing

Selected Parameters: ● Time and Memory

AlgoType	EncTime	DecTime	Memory
AES	1	3	403004
3DES	3	4	405064
BLOWFISH	4	4	405200

Fig. 7 Display the graph of result and table of comparison

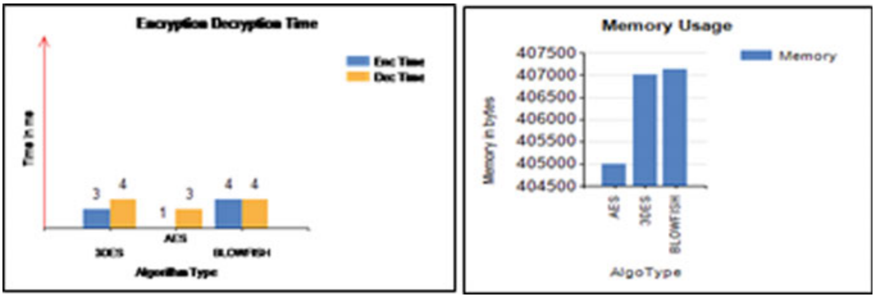


Fig. 8 Enc, dec time and memory usage for symmetric key algorithms

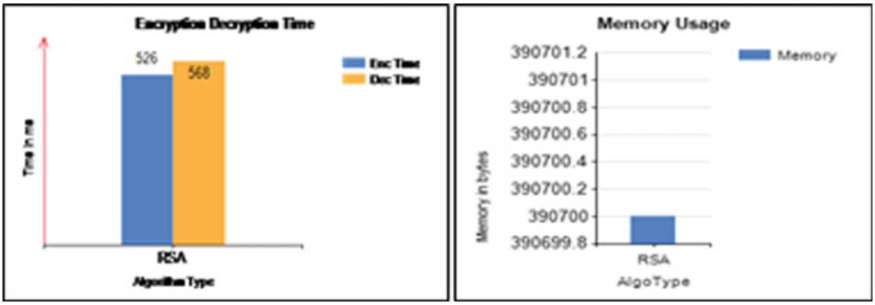


Fig. 9 Enc, dec time and memory usage for asymmetric key algorithm

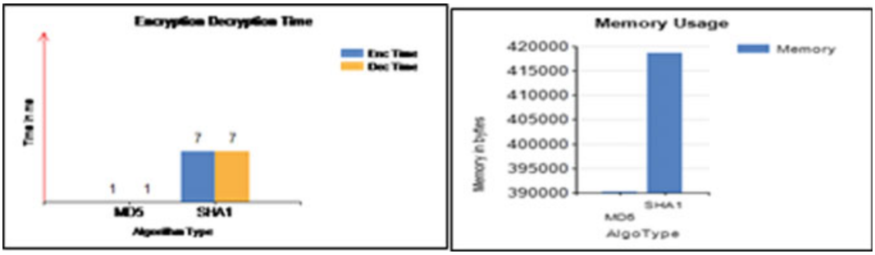


Fig. 10 Enc (HashTime), dec (Rehash Time) and memory usage for hashing algorithms

Table 3 Output table for enc time, dec time and memory taken by text file

Algorithm type	Enc time (in ms)	Dec time (in ms)	Memory usage (in bytes)
AES	1	3	404,980
3DES	3	4	407,008
BLOWFISH	4	4	407,144
RSA	526	568	390,700
MD5	1 (HashTime)	1 (ReHashTime)	390,144
SHA1	7 (HashTime)	7 (ReHashTime)	418,720

Blowfish more stable compared to AES.But ratio of encryption time and decryption time showed that AES performed faster than Blowfish algorithm [7, 17]. Comparison graph of memory usage showed that BLOWFISH took the more memory than the other two. Result analysis of phase two showed that AES algorithm took not as much time to encrypt the text file and RSA took more time to decrypt the same data or text but consumed less amount of memory. Phase three result analysis using parameters like length of hash value and speed of execution showed that the longer hash value is difficult to break as MD5 generates fingerprint/hash of 128 bits and SHA-1 generates 160 bits hash value. SHA-1 takes 80 steps and MD5 takes 64 steps with 128 bits buffer and 160 bits buffer respectively makes it slower than MD5 [6].

6 Discussion

In AES algorithm key size (128 bits, 192 bits and 256 bits) is quite enough to protect the information. Attacks on security proved unsuccessful on AES [10] and reviewed by NSA strong and stable on non classified data by U.S. government.

As per analysis RSA must be improved because of computational overhead [4] as it takes more encryption time than decryption. Longer encryption time is disadvantage of RSA [9]. Also, the attacker uses different methods (Brute Force, Oracle attack, Timing attack) which makes RSA imperfect in the term of security [10].

The speed of SHA-1 was slower than MD5 and message digest of MD5 can be used when small data size needed [6].

7 Future Work Based on Conclusion

This paper evaluates the performance analysis of cryptographic algorithms which includes symmetric key cryptography, asymmetric key cryptography and hashing algorithm. From the previous work done and related analysis it is concluded that AES algorithm is faster, secure with moderate power consumptions and memory used in comparison with DES, 3DES and Blowfish algorithm.

If we analyze all the parameters of RSA which is an asymmetric key cryptography algorithm it has proven inadequate safe algorithm. It can also be concluded that symmetric key cryptographic system shows better performance than asymmetric key cryptography algorithms.

And finally, for hashing algorithms can be concluded that though SHA-1 message digest is 32 bits shorter than MD5. SHA-1 takes 80 steps and MD5 takes 64 steps to generate message digest. Hence it is proven that MD5 runs faster than SHA-1. MD5′s run time is higher than SHA-1, we can use MD5 algorithm to secure small size data. A proposed way for future work could be a faster encryption using features of the above cryptographic algorithms for their improvement in performance.

References

1. Panda M (2016) Performance analysis of encryption algorithms for security. In: 2016 IEEE international conference on signal processing, communication, power and embedded systems (SCOPES), pp 278–284
2. Gedam M, Meshram B (2019) Vulnerabilities and attacks in SRS for object-oriented software development. WCES (2019), San Francisco, USA, 22–24 Oct 2019
3. Bhardwaj A, Som S (2017) Study of different cryptographic technique and challenges in future. 2016 IEEE 1st international conference on innovation and challenges in cyber security (ICICCS2016), pp 208–212
4. Meelu P, Meelu R (2012) Implementation of public key cryptographic system: RSA. Int J Inf Technol Knowl Manage 5(2):239–242
5. Debnath S, Chattopadhyay A, Dutta S (2017) Brief review on journey of secured hash algorithms. In: 2017 4th international conference on Opto-electronics and applied optics (Optronix). IEEE
6. Wang Z, Cao L (2013) Implementation and comparison of two hash algorithms. In: International conferences on computational and information sciences. IEEE, pp 721–725
7. Raigoza J, Jituri K (2016) Evaluating performance of symmetric encryption algorithms. In: 2016 IEEE international conference on computational science and computational intelligence, pp 1378–1381
8. Mota A, Azam S, Shanmugan B, Yeo K, Kannoorpatti K (2017) Comparative analysis of different techniques of encryption for secured data transmission. IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI-2017). IEEE, pp 231–237

9. Bonde S, Bhadade U (2017) Analysis of encryption algorithms (RSA, SRNN and 2 key pair) for information security. IEEE, pp 1–5
10. Saleh M, Tahir N, Hisham E, Hashim H (2015) An analysis and comparison for popular video encryption algorithms. IEEE, pp 90–94
11. Zhang J, Jin X (2012) Encryption system design based on DES and SHA-1. In: 11th international symposium on distributed computing and applications to business, engineering and sciences. IEEE, pp 317–320
12. Singh G, Kinger S (2013) Integrating AES, DES, and 3-DES encryption algorithms for enhanced data security. Int J Sci Eng Res 4(7)
13. Blowfish https://www.commonlounge.com/discussion/d95616beecc148daaa23f35178691c35
14. SHA-1 https://en.wikipedia.org/wiki/SHA-1
15. Makkad R, Sahu A (2016) Novel design of fast and compact SHA-1 algorithm for security applications. IEEE, pp 921–925
16. Jayapandian N, Menagadevi R, Abinaya S, SriSampoorani O (2017) To enhance consumer privacy and security for online shopping using MD5 algorithm. In: 2017 international conference on innovations in information, embedded and communication systems (ICIIES). IEEE
17. Kofahi N (2013) An empirical study to compare the performance of some symmetric and asymmetric ciphers. Int J Secur Appl

Driver Activity Monitoring Using MobileNets

Garima Tripathi, Deval Srivastava, Priyank Shah, and Saim Shaikh

Abstract This paper presents a method to monitor the driver's activity and continuously look for red flags such as distracted driving, overuse of mobile phones while driving, drowsiness, sleeping. This is achieved by using a camera-based system and the MobileNet neural network which has been fine tuned on our self-made dataset.

Keywords Computer vision · Machine learning · Activity monitoring · MobileNets · Image processing · Raspberry pi · Neural networks · Deep learning

1 Introduction

In our current world as driving technology continues to grow the driving effort required decreases. Hence, our drivers become more and more careless resulting in loss of life in many circumstances. The proposed system aims to solve this problem by developing a system to monitor driver's activity and warn them whenever necessary. The method involves deploying a neural network trained on various categories such as talking or texting on the phone, talking to co passengers, operating the radio and drinking water. In our paper we have extensively tested performance of different neural networks [1] such as Resnet-50 [2], Inception [3] and MobileNets [4]. Throughout the development our focus has been to make a system which replicates the driver's real life conditions. Hence, our network will receive the images from an IR camera allowing our system to perform during night time. Our system can be easily fitted to any existing vehicle very easily and will be intuitive to use. Our system has been developed such that it can work even in regions having extremely poor internet connectivity. The system will also be equipped with sensors to detect rash driving and will consist of security features such as fencing, fingerprint authentication to prevent thieving of the vehicle.

From the previous work and research done in this domain it can be concluded that the most popular computer vision methods include detecting driver inattention

G. Tripathi (✉) · D. Srivastava · P. Shah · S. Shaikh
Fr. Conceicao Rodrigues College of Engineering, Mumbai, India
e-mail: garima@frcrce.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_15

using head pose, eye gaze estimation or simply checking eye closure rate as well as measures such as EEG, electrocardiogram, etc. We will discuss these methods and other techniques that have been used in the next section.

The paper has been organized in the following manner, the next section will discuss related work in the given domain in depth after which we will present our method of approaching this condition and finally we will discuss all the results we have obtained from the proposed system.

2 Literature Survey

According to the research conducted by us it can be concluded that the most popular methods to solve this problem involve either driver biological measures, driver physical measures, driving performance measures or some kind of a hybrid measure [5].

Driver biological measures include biological signals like EEG, electrocardiogram (ECG), electro-oculography (EOG). These signals are collected through electrodes in contact with the skin and then analysed for fatigue and drowsiness. Physical measures involve eye closure detection and blink frequency, face position, driver gaze to detect inattention.

Driver performance measures involve various measures such as steering angle and other driving criteria. Most research that has been done in the related field has been focused on detecting driver inattention using eye gaze tracking and head pose estimation. These methods rely only on head and eye movement to detect inattention whereas in real life a driver can be distracted doing various tasks that cannot be detected by head movement alone. It has been observed that current driver monitoring systems employ statistical machine learning methods to detect driver distraction and work on a limited dataset. Some research has been done on applying deep learning technologies to solve this problem but such systems cannot be deployed in a vehicle in a cost effective manner nor do they work in night time conditions [5, 6].

3 Algorithm

For the development of our system we made use of the MobileNet Algorithm [7]. MobileNet [7] is a neural network that was developed by Google to perform on low powered devices lacking graphical GPUs that are known to accelerate neural network performance. MobileNets are small, low-latency, low-power models parameterized to meet the resource constraints of a variety of use cases, one of those use cases is that it can also be deployed on a Raspberry pi which we intend to do.

A standard convolution, filters and combines inputs into a new set of outputs in one step, but in a case of MobileNets it first uses depth wise convolution [7] that applies a single filter to each input channel. The point wise convolution then applies a 1 ×

1 convolution to combine the outputs the depth wise convolution. The depth wise separable convolution splits this into two layers, a separate layer for filtering and a separate layer for combining. This factorization has the effect of drastically reducing computation and model size, this modification allows MobileNet to be faster than its other counterparts.

For our application we have employed a MobileNet v2 [7], its the second iteration of MobileNets and now along with the depth wise separable blocks it also uses bottleneck residual layers and also adds a 1 × 1 expansion block whose purpose is to expand the number of channels in the data before it goes to the next block. In the proposed system we have used MobileNet v2 as it's much better than its older version, mobilenet was trained on a self made dataset of driver's performing distracted activities.

4 Proposed method

4.1 Implementation

Figure 1 describes the entire workflow of the project right from the hardware setup to the user interface. The Pi camera is mounted on an appropriate position in the dashboard of the vehicle. It is then connected to the camera port on the Raspberry Pi. The SM808 GSM + GPS module is connected to the Raspberry Pi via USB TO RS232 serial port. The GPS antenna is connected to the module and placed outside

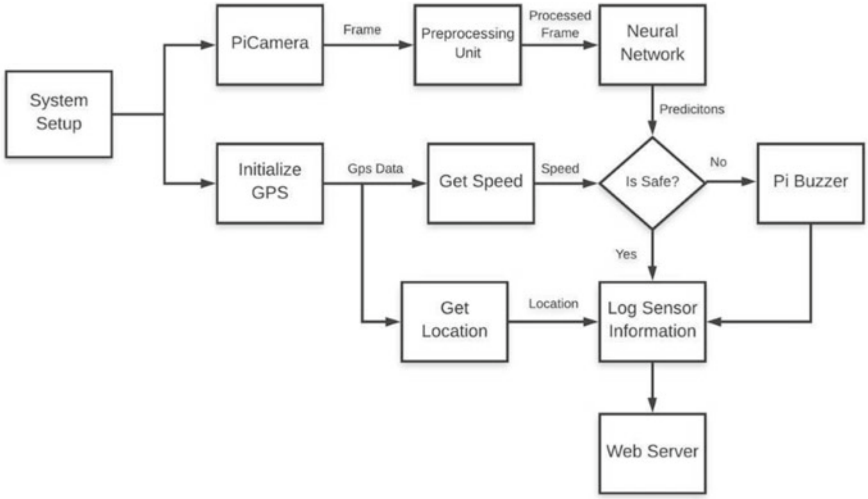


Fig. 1 Implementation flowchart

the vehicle with a clear view of the sky. The Raspberry Pi is then connected to a portable power supply via the micro USB port.

The Pi camera records footage of the driver and sends the frames to the Preprocessing Unit. The footage is recorded at a resolution of 640 * 480 at 25 frames per second. The preprocessing unit then performs basic image processing, noise reduction on each frame and resizes them to 224 * 224 * 3 then it is forwarded to the neural network which predicts the class of that image and depending on that result we declare the driver as distracted or not. If the driver is distracted the buzzer is rung to alert the driver.

For night time footage the Pi Camera is aided by 2 IR bulbs which help provide clear frames even in pitch black conditions. The frames received from the Pi Camera are processed. Alongside this we also calculate the speed of the vehicle. Once both of these operations are complete data is pushed to the administrative web server.

4.2 Dataset and Training

When we were looking for a dataset that would be able to suffice our needs for the classification tasks we came across many publicly available datasets one of them being the NTHU Driver Drowsiness Detection Dataset [8], Kaggle Statefarm Dataset [9] but none of them proved to be satisfactory as our trained neural networks were not performing as expected in real world conditions [10]. After this realization we started working on creating a real world dataset curated to our task. We recorded 6 drivers performing various distracted activities across multiple cars. We recorded drivers performing activities such as talking on the phone, texting while driving, drinking, sleeping, yawning. All of these activities were performed in simulated environments where the drivers are not actually driving. The Dataset was recorded using a pi camera as it's the camera that will be feeding images to the neural network. We recorded images for the night dataset by using an IR camera and IR lights. The database statistics are presented in Table 1.

Figures 2 and 3 we can see the dataset samples from night and from day. The images are in order of Sleeping, Talking on the phone, Drinking, Texting, Yawning.

Table 1 Database statistics

S. No.	Class	Count (day + night)
1	Safe driving	5000
2	Talking on phone	5000
3	Texting on phone	5000
4	Drinking	5000
5	Sleeping	5000
6	Yawning	5000

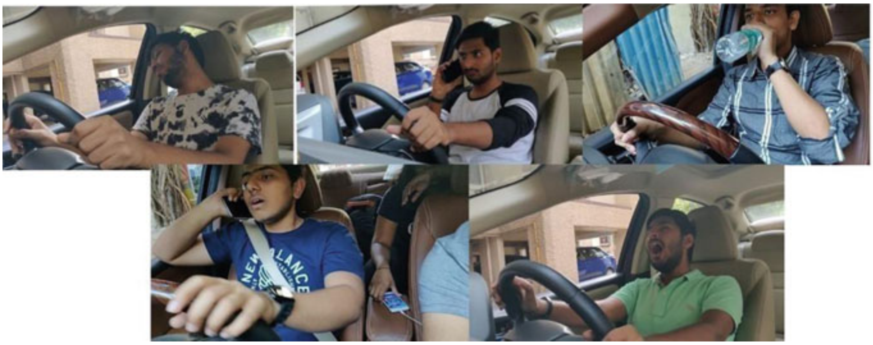


Fig. 2 Dataset samples taken during day

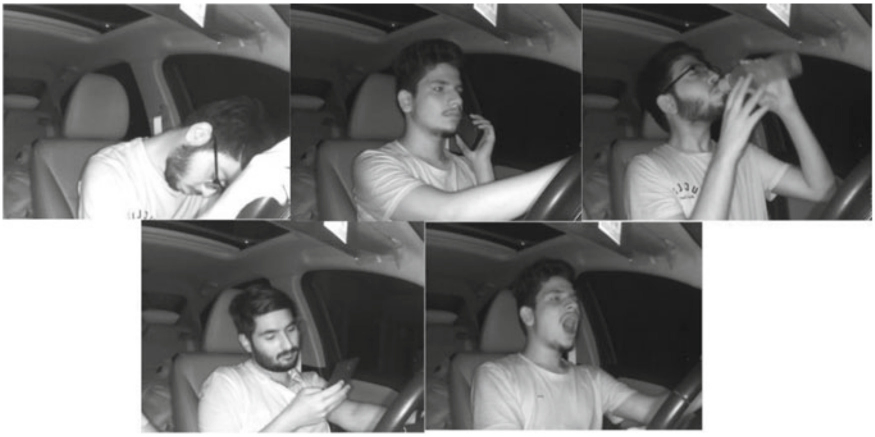


Fig. 3 Dataset samples taken during night

This section will discuss how we trained our model on the above dataset. We have used transfer learning to train our neural network as we have used the model of MobileNet v2 and only trained the last few layers to get best results and quicker training times.

We used a 60, 20 and 20% split for training, validation and testing respectively for our model. To further simulate real world conditions we added data augmentation to our model. This allows our model to perform better in difficult scenarios like low lighting, improper camera alignment etc. After experimenting and testing with a lot of different kinds of augmentations we found the following augmentations gave us best results were random zoom that generates extra images that are zoomed in randomly upto 20%. In the same way we added random crops upto 20% and also random brightness for varied conditions. Augmentations effectively increase the size of our dataset and also makes sure the model works better in unknown conditions.

Table 2 Comparison of different models

S. No.	Model name	CPU usage (%)	FPS	Accuracy (%)
1	Inception v2	90	1	98
2	VGG-16	93	1	96
3	Resnet-50	91	2	97
4	**MobileNet v2**	**66**	**6**	**93**

We fine-tuned the MobileNet v2 model on our custom dataset on a computer with a Nvidia GTX1070TI with Cuda acceleration. Our model was trained for roughly 6500 steps with a batch size of 32 which took 4hrs to train and we stopped when we had a validation accuracy of 95.6%.

5 Results and Discussion

In this section we will discuss the results we have received after implementing the system. Our model received 93.6% accuracy on our test dataset and along with training MobileNets we also trained our model on various other architectures such as Inception v2, ResNet-50 and VGG-16 [11] model on the same dataset in order to compare the performance we receive on the Raspberry pi. The results given below are an average of all the 5 runs done on the same Raspberry pi using the above mentioned architectures. Thus after looking at the results we can come to a conclusion that the MobileNet v2 is the most ideal neural net model for our use case. We believe that MobileNet v2 was the best model for our application as we want to have the system to be almost real time which will enable it to prevent accidents. As MobileNet v2 requires the least amount of CPU usage some processing ability of the limited compute on a Raspberry pi cpu can also be used for other tasks such as calculating speed and sending data to web server. We further tested our system in real world conditions and found satisfactory results (Table 2).

6 Conclusion and Future Scope

On successful implementation the system will provide a robust and efficient method to monitor driver activities and thus prevent accidents that occur due to distracted driving, overuse of mobile phones while driving, texting on phone, drowsiness, sleeping etc. When such a system is in place it will enforce the drivers to be more careful and drive responsibly which will prevent loss of lives and will promote a safer driving experience for other drivers on the road.

Once applied over a large number of vehicles the system can also be used to create a network of vehicles to share important information. In the future this network will be able to collect huge amounts of data and this data can be used to plan routes better. Moreover since we have deployed a hardware platform more and more features can be added in due time. Features such as facial recognition for authentication and various kinds of analysis can be done using the data of our platform. The algorithm in the proposed system relies on a neural network to detect driver's activity which performs well but an object detection approach can be used to detect specific distracting objects which will theoretically perform even better than standard neural net approaches.

References

1. LeCun Y, Bottou L, Bengio Y, Haffner P Gradient based learning applied to document recognition
2. He K, Zhang X, Ren S, Sun J Deep residual learning for image recognition
3. Szegedy C et al (2015) Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, pp 1–9
4. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural net
5. Dong Y, Hu Z, Uchimura K, Murayama N (2011) Driver inattention monitoring system for intelligent vehicles: a review. IEEE Trans Intell Transp Syst 12(2):596–614
6. Vicente F, Huang Z, Xiong X, Torre F, Zhang W, Levi D (2015) Driver gaze tracking and eyes off the road detection system. IEEE Trans Intell Trans Syst 16(4):2014–2027
7. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City, UT, pp 4510–4520
8. Weng C-H, Lai Y-H, Lai S-H (2016)D river drowsiness detection via a hierarchical temporal deep belief network. In: Asian conference on computer vision workshop on driver drowsiness detection from video. Taipei, Taiwan
9. Kaggle Statefarm Dataset https://www.kaggle.com/c/state-farm-distracted-driver-detection
10. Koesdwiady A, Bedawi S, Ou C, Karray F (2017) End-to-end deep learning for driver distraction recognition. In: Image analysis and recognition, 14th international conference, ICIAR 2017, July 2017
11. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations

A Review of Deep Learning Techniques Used in Breast Cancer Image Classification

Vidya Sarode, Amogh Chaudhari, and Fabian T. R. Barreto

Abstract Human society is faced with the increasing global burden of cancer as we enter the new decade. According to the Globocan 2018 report, in India alone, breast cancer has the largest number of new cases (27.7%) as compared to the world incidence rate of 24.7%. Breast cancer detection after clinical examination includes two tests, namely breast imaging and breast histopathology. Breast cancer has different stages or spread, aggressiveness, and genetic makeup. An end-to-end system that helps in early detection and prevention would increase the survival rates. Traditionally, machine learning techniques have been used to detect malignancy in breast cancer images. However, machine learning is limited in its ability to process natural data in its raw form. The limitation is due to the need of domain experts who can carefully handcraft features to feed a classifier. Deep learning, a subfield of machine learning, however, automatically learns multiple levels of representation and abstraction that help to analyze breast cancer images at a greater depth. Deep convolutional neural networks have achieved remarkable breakthroughs in processing images. In our paper, we review the deep learning techniques and some of the models that have given extraordinary performance for both breast cancer image and breast cancer histopathology image classification and thus helped in early detection of breast cancer.

Keywords Breast cancer image classification · Convolutional neural networks · Deep learning · Machine learning

V. Sarode (✉) · F. T. R. Barreto
Department of Electronics and Telecom, Xavier Institute of Engineering, Mumbai, India
e-mail: vidya.s@xavier.ac.in

F. T. R. Barreto
e-mail: frfabiansj@xavier.ac.in

A. Chaudhari
MGM Institute of Health Sciences, Mumbai, India
e-mail: amogh.chaudhari@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_16

1 Introduction

As we begin a new decade, we are faced with the increasing global burden of cancer. It remains an unresolved cause for increased morbidity and mortality of human beings. This, in turn, impacts social well-being and economy. The GLOBOCAN database, compiled by the International Agency for Research on Cancer (IARC), provides estimates on national cancer incidence and mortality [1]. In India, according to the GLOBOCAN 2018 report, breast cancer has the largest number of new cases (27.7%) as against the world incidence rate of 24.7%.

The last decade has seen an intensified research efforts in the field of prevention and early detection of this disease. As the wave of rapid urbanization sweeps India, one observes drastic changes in dietary habits and lifestyle, causing at times high stress levels. Studies in high-income countries have indicated that from one-third to two-fifths of new cancer cases could be avoided by eliminating or reducing exposure to known lifestyle and environmental risk factors [2].

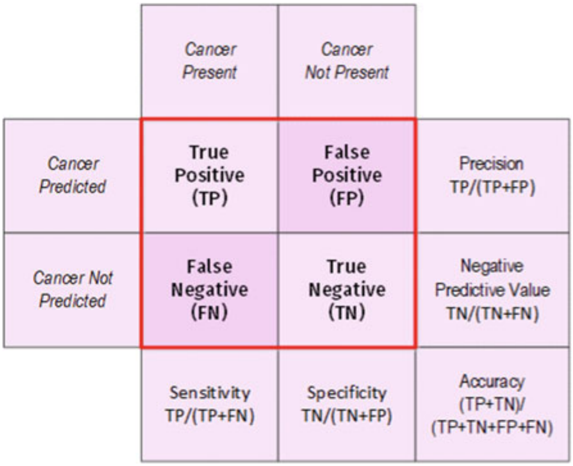
The risk factor for most cases of breast cancer is due to a combination of genetic and environmental factors. The breast cancer detection process, after physical examination, includes two tests, breast imaging and histopathology. Early detection, true diagnosis, and prompt treatment are crucial for a better prognosis [3]. This is made possible by effective and efficient screening programs. Early diagnosis of cancer in the natural progression of the disease before it spreads to local regions or to distant organs can result in higher five-year survival rates.

Imaging modalities may be classified as primary (simple) or secondary (complex). Primary imaging modalities are used in the regular screening of patients. This is the first option to be used among the different diagnostic tools, as they are cheap and provide reliable and repeatable results under diverse conditions. Mammography is one of the commonly used primary imaging modalities. Any anomaly is categorized as regular or irregular, and capsulated (benign—not cancerous) or noncapsulated (malignant). Mammography techniques may be screen-film mammography (SFM) or full-field digital mammography (FFDM). SFM may not detect benign cancer properly. FFDM performs better in case of misdiagnosed cancer samples as the images are processed using image processing techniques.

The confusion matrix as shown in Fig. 1 is used to check the performance of the DL algorithms. The true positives, in this case, give the correct predictions of the presence of cancer, and the true negatives give the correct predictions of the absence of cancer. Accuracy is the ratio of correctly predicted to the total observations. Precision, also called positive predictive value, is the ratio of true positives to the total number of positives. Sensitivity or recall is the ratio of true positives to the total actual positives. The F1 score combines sensitivity and precision and is the ratio of twice (precision $\times$ recall) to (precision $+$ recall).

A reduction in breast cancer mortality has been observed after using screening mammography [4]. It is reduced by about 20–35% in women aged 50–69 years. For women between 40 and 49 years, it is slightly less. The other primary imaging modalities include ultrasound, thermography, and electrical impedance. A further

Fig. 1 A confusion matrix



	Cancer Present	Cancer Not Present	
Cancer Predicted	True Positive (TP)	False Positive (FP)	Precision TP/(TP+FP)
Cancer Not Predicted	False Negative (FN)	True Negative (TN)	Negative Predictive Value TN/(TN+FN)
	Sensitivity TP/(TP+FN)	Specificity TN/(TN+FP)	Accuracy (TP+TN)/ (TP+TN+FP+FN)

examination and confirmation of diagnosis can be done using imaging modalities such as magnetic resonance imaging (MRI), computation tomography (CT), and scintimammography. At present, digital breast tomosynthesis (DBT, also called 3D mammography) is replacing full-field digital mammography (FFDM). It enables cross-sectional visualization of breast tissue and thus reduces recall rates.

An independent risk factor for breast cancer is breast tissue density. Masking (superimposition) of the dense tissue can lead to reduced mammographic accuracy. Hence, a supplemental imaging method is used to improve the sensitivity of screening [5]. Mammographic breast density is visually assessed by radiologists in routine mammogram image reading. They use four qualitative breast imaging and reporting data system (BIRADS) breast density categories. They often find it difficult to consistently distinguish the two most common and most variably assigned BIRADS categories, i.e., “scattered density” and “heterogeneously dense” [6]. Hence, the need is for computer-aided detection (CADe).

It is important to understand that breast cancer screening has potential limitations and drawbacks. These include false-positive, false-negative test results, and also overdiagnosis.

2 Machine Learning and Deep Learning

2.1 Machine Learning

Machine learning (ML) uses handcrafted features that are learned from the training data. ML algorithms helps to understand the structures and patterns within datasets. It is used for classification, clustering, and prediction of data points. Classification segregates in categories. Clustering divides a large set of data points into clusters having common properties. Prediction uses the past data and builds models that are used to forecast a value in the future.

Conventional ML techniques are limited in their ability to process raw data. The limitation is because of the need of domain expert who has to handcraft features for the classifier. It is easy to handcraft the low-level features for some data and tasks. However, designing effective features for new data and tasks requires new domain knowledge, as most handcrafted features cannot simply be transferred to new conditions [7].

ML algorithms use support vector machines (SVMs) with different kernel functions like Gaussian radial basis function (RBF) or a polynomial function. SVMs are "nonparametric" models where the parameters are not predefined, and their number depends on the training data used. An SVM classifier performs binary classification. Multiple classifiers may be combined using the bagging and boosting techniques to improve performance.

The other ML techniques that have been used are k-nearest neighbors (k-NN), random forest (RF), and Bayesian networks (BNs) [8]. k-NN captures the idea of similarity (distance or closeness) and makes a decision for the most frequent label in case of classification. RF combines many decision trees to ensemble a forest of trees. This gives a better stability and is also makes it insensitive to the noise of the input data. BN is a subfield of probabilistic graphical models (directed acyclic graph—DAG) that are used for prediction and knowledge representation of uncertain domains.

In ML, the classification is done in two stages, namely training and testing. Often, 80% of the data is used for training and 20% for testing. During training, the image is acquired and preprocessed, and features are extracted and selected. Similar steps are followed during testing. The preprocessing step helps to remove the noise and thus improve the quality of the image. Finally, in the classification stage, different classifiers are applied for diagnosis and prognosis.

2.2 Deep Learning

Deep learning (DL) techniques are representation-learning methods that help to understand data such as text, images, and sound. DL algorithms learn multiple levels of representation and abstraction. Figure 2 shows the difference in learning the features for DL and ML techniques. These features are obtained by composing simple but nonlinear modules that each transform the representation at one level (beginning with the raw input) into a slightly more abstract higher level representation. These transformations learn very complex intricate functions. A deep-learning architecture thus computes nonlinear input–output mappings.

DL uses deep neural network (DNN) architecture as shown in Fig. 3. The power of DNNs is due to the highly optimized hardware that is the graphics processing units (GPUs) which can handle lots of parallel computations using thousands of cores and possess a large memory bandwidth to deal with the data for these large

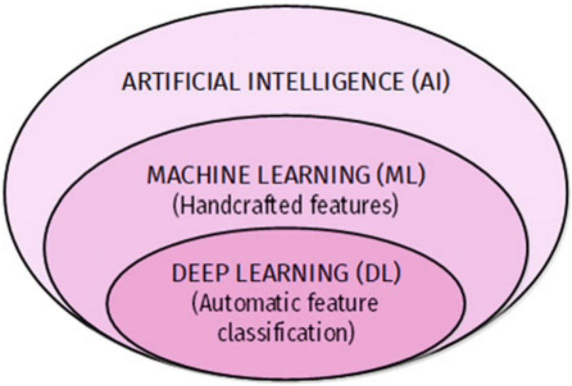

ARTIFICIAL INTELLIGENCE (AI)
MACHINE LEARNING (ML)
(Handcrafted features)
DEEP LEARNING (DL)
(Automatic feature classification)

Fig. 2 Difference between ML and DL

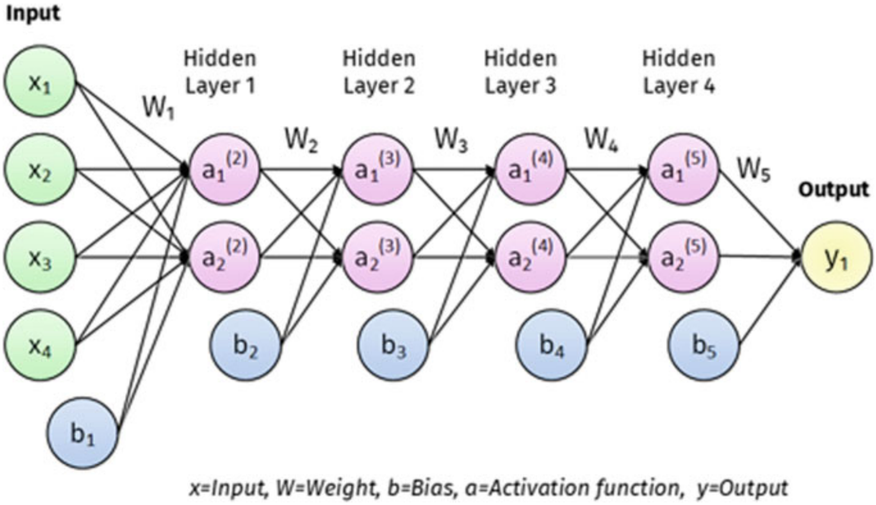


Fig. 3 Deep neural network (DNN) architecture

computations. They work on the massive training data and learn its multilevel feature representation.

Learning can be classified as supervised, unsupervised, or semi-supervised. Supervised learning selects a feature by using labelled data. Unsupervised learning evaluates feature relevance by exploiting the innate structures of the data. A semi-supervised feature selection integrates a small amount of labelled data into unlabelled data. This gives additional information, thus improving the performance of an unsupervised feature selection.

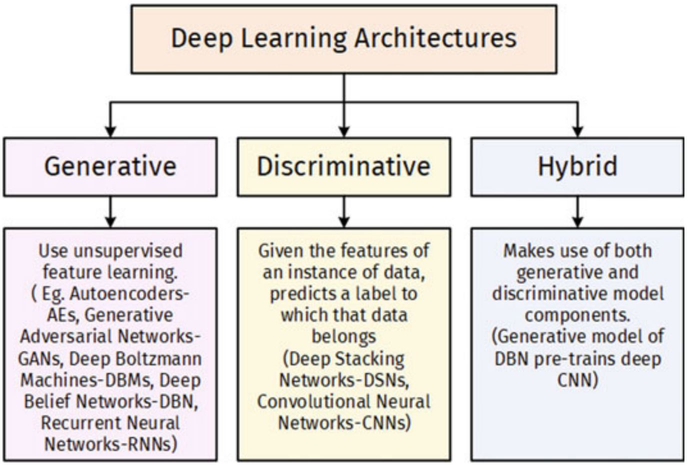

Deep Learning Architectures
Generative
Discriminative
Hybrid
Use unsupervised feature learning. (Eg. Autoencoders-AEs, Generative Adversarial Networks-GANs, Deep Boltzmann Machines-DBMs, Deep Belief Networks-DBN, Recurrent Neural Networks-RNNs)
Given the features of an instance of data, predicts a label to which that data belongs (Deep Stacking Networks-DSNs, Convolutional Neural Networks-CNNs)
Makes use of both generative and discriminative model components. (Generative model of DBN pre-trains deep CNN)

Fig. 4 Deep learning architectures

Deng [9] categorizes DL architectures into three areas as shown in Fig. 4. They are generative, discriminative, and hybrid, depending on how the architectures and techniques are intended for use, e.g., synthesis/generation or recognition/classification. Generative deep architectures characterize the high-order correlation properties of the observed data for pattern analysis.

3 Datasets

In the past three decades, several image databases have been created for breast cancer research, some of which are public and others private. Public databases provide a common reference for researchers to test and compare their methods. The image database is crucial to the type of study that one wishes to undertake. Often, the collected datasets are class imbalanced. There is a large class where the patients do not have cancer and a small class of patients with cancer. Necessary precautions need to be taken while conducting research by choosing the proper research design.

Datasets are available both for breast imaging and breast histopathology. Based on the imaging results, a doctor may suspect that a person has breast cancer. A biopsy is then performed to diagnose if cancer is really present. Among the different biopsy techniques, the most common ones are fine needle aspiration (FNA), core needle biopsy, vacuum-assisted, and incisional or excisional surgical biopsy [10]. The surgical biopsy, though reliable, is invasive and costly.

The designed classifying system needs an adequately large and preferably public database. In case of deep learning, this database has to be large to get meaningful classification results. The author in [11] gives the details of the requirements for mammographic databases. These include the case selection, ground truth, patient's associated information, and importantly the organization and distribution of the database. In [12], a case is made for computer-aided detection (CADe) as against a human person interpreting screening mammograms because the nature of the task is highly repetitive and also highly subjective, resulting in large intra- and inter-interpreter variability.

Breast cancer Wisconsin (diagnostic) dataset [13] which was made public in 1992 has been widely used especially for creating good machine learning classifiers [14]. It has 32 attributes, 569 instances and two classes.

Another database is the mini-MIAS (mammographic image analysis society) database of mammograms [15]. The dataset contains 322 images, with each image falling into one of seven categories: calcification, circumscribed masses, spiculated masses, architectural distortion, asymmetry, other ill-defined masses, and normal. The size of each image is 1024×1024 pixels.

A publicly available database, released in 1997, is the digital database for screening mammography (DDSM) [16] which is available from the University of South Florida. The authors in [17] have developed a new dataset called INbreast. There are 115 cases with a total of 410 images. 90 cases are of those with both breasts affected with four images per case. 25 cases are from mastectomy patients with two images per case.

The breast cancer histopathological image classification (BreakHis) dataset [18] has 9109 microscopic images of breast tumor tissue. It has been collected from 82 patients from a clinical study which ran for a year from January 2014 to December 2014. Different magnifying factors (40×, 100×, 200× and 400×) were used which gave a uniqueness to the dataset. It contains 2480 benign and 5429 malignant samples. These images are 700×460 pixels and are colored three-channel RGB with an 8-bit depth in each channel. The format used is portable network graphics (PNG) format.

4 Deep Learning Models

There are many variants of CNN architectures based on the requirements for the specific application. In general, the architecture consists of three layers, convolutional, maxpooling and fully connected. The convolutional layer is made up of kernels (filters) that learn the different feature maps. More abstract features are learnt with more levels of convolutional layers.

AlexNet [19] is a deep CNN that classified 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. The neural network has 60 million parameters and 650,000 neurons. It consists of five convolutional layers, two of which are followed by maxpooling layers, and at the end are three fully-connected layers with a final softmax layer as shown in Fig. 5. The

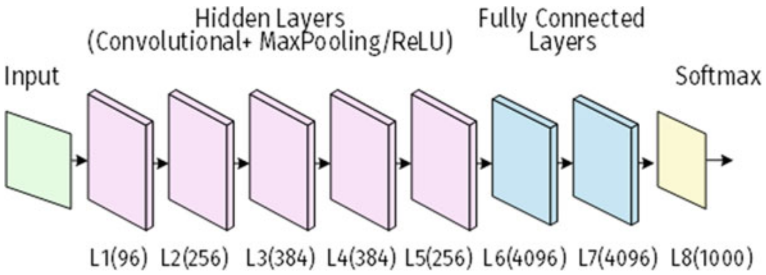


Fig. 5 AlexNet architecture

authors in [6] used an improved version of the AlexNet model. They used the two-class CNN model to classify the two BIRADS breast density categories: “scattered density” vs. “heterogeneously dense.” They have suggested following future work in the area: identifying more likely to be misclassified images, comparing other CNNs model structures, developing approaches to deal with noisy or potentially inaccurately labeled data and testing by using larger multicenter datasets. They used transfer learning from nonmedical images to medical imaging-based applications. They used the AlexNet model pretrained on ImageNet and fine-tuned it with their mammogram data.

U-Net [20] describes a convolutional neural network (CNN) that was developed for biomedical image segmentation at the computer science department of the University of Freiburg, Germany. The architecture works with fewer training images and yields more precise segmentations. The name comes from the U-shaped architecture consisting of a contracting (Convolution + Maxpooling + Rectified Linear Unit-ReLU) and an expanding path (up-convolutions and concatenations).

The authors in [21] have developed a deep learning model called grouped-Resaunet (GRA U-Net) which segments the nipple region with very high accuracy from other similar looking nonnipple regions such as a tumor, resulting in better segmentation results for prognosis.

The authors in [22] developed an efficient deep learning framework. It identifies, segments, and classifies cell membranes and nuclei from human epidermal growth factor receptor-2(HER2)-stained breast cancer images with minimal user intervention. The also used long short-term memory (LSTM) to use pixel information from the multiple frames in making the semantic segmentation and classification cases. Their model achieved 96.64% precision, 96.79% recall, 96.71% F-score, 93.08% negative predictive value, 98.33% accuracy, and a 6.84% false positive rate. Their framework was run for seven days (24 × 7 = 168 h) for training which was done on a parallel computing platform. It took on average 1.6 seconds for cell membrane and nucleus segmentation, classification, and HER2 scoring on test data.

Thus, different models are used to compute better results for breast cancer image and breast cancer histopathology image classification. One practical problem for researchers is the availability of Indian datasets. Often, the ground reality of the

Indian patients may be far different than the ones in the Western world. Hence, the need is to develop Indian datasets to make it more relevant to our context.

5 Conclusion

Researchers from different disciplines are fighting against time to discover the Holy Grail to cure cancer. Multidisciplinary research efforts in the field of prevention and early detection have shown progress on many fronts. As of today, great progress has been achieved in increasing the survival rate of patients. For example, transitioning from standard digital images to the tomosynthesis modality has helped in the early detection of cancer cases and shown the potential to reduce the number of false positives. Deep learning-based CNN models have exhibited extraordinary performance for both breast cancer image and breast cancer histopathology image classification. As deep learning techniques progress, we hope for an end-to-end cancer detection system, which along with other multidisciplinary approaches, will not only help in early detection but also help prevent cancer in the first place.

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Znaor A, Bray F (2019) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer 144:1941–1953
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394–424
3. Early Detection of Breast Cancer https://www.breastcancerindia.net/screening/guidelines.html, last accessed 2020/01/10
4. Elmore JG, Armstrong K, Lehman CD, Fletcher SW (2005) Screening for breast cancer. JAMA 293:1245–1256
5. Wender RC, Brawley OW, Fedewa SA, Gansler T, Smith RA (2019) A blueprint for cancer screening and early detection: Advancing screening's contribution to cancer control. CA Cancer J Clin 69:50–79
6. Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S (2018) A deep learning method for classifying mammographic breast density categories. Med Phys 45:314–321
7. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35:1798–1828
8. Bazazeh D, Shubair R (2016) Comparative study of machine learning algorithms for breast cancer detection and diagnosis. In: 2016 5th international conference on electronic devices, systems and applications (ICEDSA)
9. Deng L (2012) Three classes of deep learning architectures and their applications: a tutorial survey. APSIPA Trans Signal Inf Process 1
10. Biopsy https://www.breastcancer.org/symptoms/testing/types/biopsy
11. Nishikawa M (1996) Development of a common database for digital mammography research. University of Chicago, Chicago, IL
12. Nishikawa RM, Gur D (2014) CADe for early detection of breast cancer—current status and why we need to continue to explore new approaches. Acad Radiol 21:1320–1321

13. Breast Cancer Wisconsin (Diagnostic) Data Set https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic), last accessed 2020/01/12
14. Bennett KP, Mangasarian OL (1992) Robust linear programming discrimination of two linearly inseparable sets. Optim Methods Softw 1:23–34
15. The mini-MIAS database of mammograms https://peipa.essex.ac.uk/info/mias.html, last accessed 2020/01/11
16. University of South Florida Digital Mammography https://www.eng.usf.edu/cvprg/Mammography/Database.html, last accessed 2020/01/10
17. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012) Inbreast: toward a full-field digital mammographic database. Acad Radio 19:236–248
18. Breast Cancer Histopathological Database (BreakHis) https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/, last accessed 2020/01/11
19. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems
20. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention
21. Zhuang Z, Raj ANJ, Jain A, Ruban N, Chaurasia S, Li N, Lakshmanan M, Murugappan M (2019) Nipple segmentation and localization using modified U-net on breast ultrasound images. J Med Imaging Health Inform 9:1827–1837
22. Saha M, Chakraborty C (2018) Her2net: a deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. IEEE Trans Image Process 27:2189–2200

Prediction of nMAG in PMIPv6 with the Help of MN Positions

Nitesh M. Tarbani, A. S. Alvi, G. R. Bamnote, and V. S. Gulhane

Abstract The Proxy Mobile IPv6 (PMIPv6) is a protocol which manages mobility. It uses signaling and home agent's activity of MIPv6 through a proxy mobility agent in a localized network. In any type of wireless communication, handover delay should be as less as possible. Tremendous work has been done to minimize the handover delay in PMIPv6, and many solutions have been proposed by the many researchers. Almost all of the solutions have one thing in common that the authentication information of the MN should be sent to new MAG in advance, and to send information to new MAG, it is mandatory to anticipate the new MAG. This paper proposes an algorithm to predict new MAG and also provides the results achieved by simulating the proposed algorithm on NS-2.29.

Keywords PMIPv6 · Mobile node · AAA server · MAG

1 Introduction

In the current period of wireless technology, mobility is gaining popularity for enabling users to access resources while roaming. There are mainly two models which provide mobility by the use of mobile IP. The first is network-based mobility management in which the functions related to mobility management reside only in network entities, and it does not need involvement of MN in any mobility managing

N. M. Tarbani (✉) · V. S. Gulhane
Sipna College of Engineering, Amravati, Maharashtra, India
e-mail: ntarbani@gmail.com

V. S. Gulhane
e-mail: vijaygulhane27@gmail.com

A. S. Alvi · G. R. Bamnote
Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, India
e-mail: asalvi@mitra.ac.in

G. R. Bamnote
e-mail: grbamnote@mitra.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_17

activity. That means, MN remains completely unaware of its mobility. Another one is host-based mobility management which needs active participation of MN in functions related to mobility management. Therefore, an additional software stack should be installed in MN. Thus, host-based mobility management requires updating current mobile node to fit into the network, whereas there is no such requirement in network-based mobility management. For this reason, Internet Engineering Task Force (IETF) has standardized Proxy Mobile IPv6 (PMIPv6) as the network-based mobility management protocol.

In PMIPv6, the mobile node is privileged to roam within that localized network provided by the PMIPv6 domain. Thus, the network administrator may have overall knowledge-set of complete network, and also the resources/assets are made gettable at every PoAs and their adjoining PoAs as well. In PMIPv6, there are three network entities which take care of functions related to mobility management: local mobility anchor (LMA), mobile access gateway (MAG), and authentication, authorization, and accounting (AAA) Server. To use the network, MN needs to get registered to LMA and then MN can access network through MAG. While moving in the network, MN may get disconnected from current MAG and get connected to the new MAG. Each time, when MN gets connected to MAG, it verifies authentication of MN from LMA which increases handover delay significantly.

To reduce handover delay, many researchers have proposed that current MAG should send authentication information of MN to new MAG. This is possible when new MAG should be known to current MAG. This paper focuses on the implementation of algorithm that predicts new MAG with the help of MN movement. This paper also presents results achieved.

2 Literature Review

Academic and industries have done widespread research work to reduce handoff delay in host-based mobility management protocols and network-based mobility management protocol.

D.Johnson et al. proposed MIPv6 which describes functioning of MIPv6. MIPv6 is host-based mobility model in which each mobile node is always identified by its home address, regardless of its current point of attachment to the Internet [1]. Mobile IPv6 allows a MN to be connected to the Internet while travelling from one access point to another, a process is called as handover. During handover, there is a time period when the MN cannot communicate to the Internet due to various operations such as link switching, IP protocol tasks, movement detection, new care of address configuration, and binding update. Above-mentioned handover delay as a result of typical Mobile IPv6 operations is mostly not tolerable to real-time traffic.

To reduce this handover delay, R. Kodali et al. has proposed fast handover for Mobile IPv6. In this, various factors are focused to reduce handover delay such as how to enable a MN to send packets as soon as it finds a new subnet link and how to deliver packets to a MN as soon as it becomes reachable [2].

Soliman et al. also proposed Hierarchical Mobile IPv6 Mobility Management in which handover delay has been reduced in HPMIPv6 by introducing one more level in hierarchy and one more MIPv6 node called as Mobility Anchor Point [3].

Similarly, in network-based mobility management, several works have been performed to minimize handover delay. V. Deverapalli et al. proposed "Proxy Mobile IPv6" which is envisioned for delivering mobility management support to a mobile node, which does not require the participation of the MN in IP mobility-related operations. The main components of PMIPv6 are local mobility anchor (LMA) and the mobile access gateway (MAG). In PMIPv6, AAA server can be used to store policy profile of MNs to keep track of genuine MN [4].

Ming-Chin Chuang et al. proposed FH-PMIPv6, defines predictive handoff mode in which the next MAG is informed about the handover information in advance. To decrease time required for handover, registration, and authentication phases are executed simultaneously, in FHPMIP [5].

Ahmad Rasemy et al. have proposed protocol called Optimized Proxy MobileIPv6 (O-PMIPv6) in which route optimization is also performed while handoff is taking place. In their proposed protocol various parameters of network such as signaling cost, network utilization, and handover delay are improved. Moreover, the performance of PMIPv6 and F-PMIPv6 also optimized in this proposed protocol [6].

Geumsan Jo et al. have proposed protocol which anticipates the nMAG by comparing the history of MN movement with the current location of the MN. They also proposed to maintain history of MN movement local mobility anchor (LMA) [7].

3 Proposed Methodology

As discussed earlier, to reduce handover delay in PMIPv6, many researchers have proposed that authentication information of MN must be received by nMAG by current MAG in advance. But actually, every MAG in the network is surrounded by many other MAGs. It means that present MAG should anticipate the new MAG to which the MN would connect next. Following is the algorithm proposed here to predict the new MAG.

In this work, the area around current MAG is divided into eight different parts. Once the MN is about to go out of coverage area of current MAG, the MAG will find out the current location of MN using GPS. Upon knowing current location of MN, MAG can find out the nearest MAG to the MN using well-known Haversian formula. The simulation of all entities (MAG, LMA, AAA) of PMIPv6 is avlaible in NS2.29 [8]. As the proposed methodology has been simulated in the NS-2.29, following formula is used to find out distance between MAG and current location of MN.

$$D = \sqrt{(X1 - X2)^2 + (Y1 - Y2)^2}$$

where ($X1$, $Y1$) and ($X2$, $Y2$) are locations of MN and MAGs, respectively.

Following is the pseudo code for finding the nearest MAG to MN current location.

```
if(dist >threshold && mniflag==0)
{
  for(i=0;i<n;i++)
  {
     dx=xpos-amxpos[i];
     dy=ypos-amypos[i];
     dxs=dx*dx;
     dys=dy*dy;
     dist=dxs+dys;
        if(dist<min)
        {
             min=dist;
             mini=i+2;
        }
     }
   mniflag=1;
   }
```

where dist is the distance of MN from current MN.

Threshold is used to decide whether MN is likely to leave coverage area or not, xpos is x coordinate of MN, and amxpos is the array which contains the MAG addresses.

4 Result and Analysis

For simulation, one AAA server, one LMA, nine MAGs, one corresponding node (CN), and one MN are considered, and CN needs to send data to MN. Topology is simulated in such a way that MN will get connected to MAG which is surrounded by eight other MAGs. When MN is about to get disconnected from current MAG, current MAG uses above-proposed methodology to predict the next MAG. The important parameters of simulation are listed in Table 1.

The correctness of proposed methodology is evaluated on the basis of predicted new MAG and actual new MAG. The TCL script is run again and again with new destination of MN, so that each time the new MAG for MN should be different. Each time, readings of predicted new MAG and actual new MAG are noted down. The readings, in which the predicted new MAG and actual new MAG are same, show that the prediction was correct, whereas the readings, in which the predicted new

Table 1 Simulation parameters

NS2 version	NS-2.29
Simulation duration	20 s
Total no. of LMAs	1
Total no. of MAGs	9
Size of packet	1000 B
Protocol used for routing	PMIPv6 with AAA
Traffic type	CBR
Destination of MN	Each time, different destination is set

MAG is not same as actual new MAG, show that the prediction was not correct. To evaluate the correctness of proposed methodology, following formula was used.

$$\text{Correctness Percentage} = \frac{\text{CPr}}{\text{TPr}} * 100$$

where CPr is number of correct predication got and TPr is total number of predictions taken.

Handoff delay also was observed for correct predictions and incorrect predictions. Handoff delay is nothing but the time difference between the time at which the MN got disconnected from current MAG and connected to new MAG [9, 10]. Figure 1 explains the handoff delay.

Figure 1 shows $t1$ is the time at which MN got disconnected from current MAG (MAG1), at time $t2$, MN discovered the new MAG (MAG2). Though MN discovered new MAG, MN cannot use the network until new MAG receives PBA from LMA. At time $t3$, MAG2 receives PBA from LMA, and at $t3$, MN now can use the network.

Fig. 1 Handoff delay

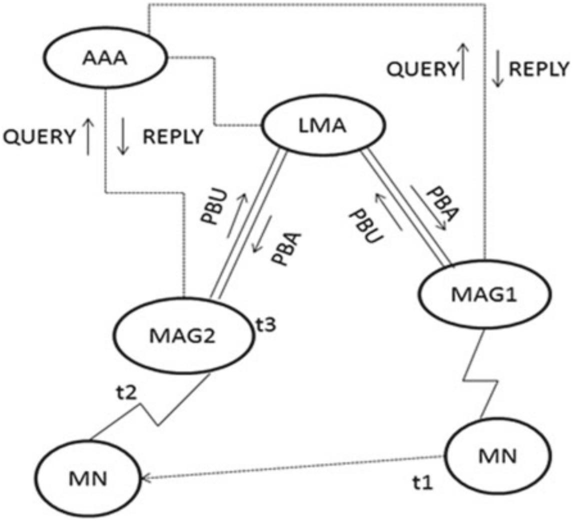


Table 2 Handover delay

Situation	Handoff delay (ms)
Without prediction	8.044
Correct prediction	**4.024**
Incorrect predictions	8.044

So, handoff delay in this case can be given as follows.

$$\text{Handoff delay} = t3 - t1.$$

Table 2 shows the handoff delay observed in various situations.

The table shows that three scenarios were observed. First one, when prediction algorithm is not used, handoff delay found in this scenario is 8.044 ms. In second scenario, when prediction algorithm is used and predicted MAG was the same as of new MAG, the handoff delay observed was 4.024 ms (almost half of the previous case). In third scenario, the predicted MAG was not same as new MAG to which MN got connected. In this case also, handoff delay observed was 8.044 ms. Moreover, to find correctness of implemented algorithm, the code is run against 200 distinct destination addresses and out of 200 runs, 184 times, the predicted address, found to be correct. This allowed us to conclude that the correctness of algorithm is 92%.

5 Conclusion

The PMIPv6 is one of the promising network-based mobility management algorithm. However, the handover delay is the place where this algorithm can be optimized more and made more prominent. Many research works have been done in this field, to reduce handoff delay of PMIPv6. The backbone of reducing handoff delay is that new MAG must get the authentication information of MN from present MAG in advance. To achieve this, current MAG should predict new MAG of MN. This paper proposed an algorithm to predict the new MAG and also presented the result achieved. It is observed that use of prediction algorithm leads to reduction in handoff delay to almost half. Moreover, to find correctness of prediction algorithm, the algorithm is run 200 times with distinct destination addresses of MN. The correctness of algorithm is found to be 92%.

References

1. Johnson D, Perkins C (2004) Mobility support in IPv6, RFC3775, June 2004
2. Koodli R (2005) Fast handovers for mobile IPv6. Internet engineering task force request for comments—4068, July 2005

3. Soliman H, Castelluccia C, El Malki K, Bellier L (2005) Hierarchical Mobile IPv6 mobility management (HMIPv6). Internet engineering task force request for comments—140, Aug 2005
4. Devarapalli V, Chowdhury WK (2008) Proxy Mobile IPv6, RFC 5213, Aug 2008
5. Chuang M-C, Lee J-F (2011) FH-PMIPv6: a fast handoff scheme in proxy mobile IPv6 networks. In: International conference on consumer electronics, communications and networks IEEE, Nov 2011
6. Rasemy A, Makaya C, St-Hilairey M (2012) O-PMIPv6: efficient handover with route optimization in proxy mobile IPv6 domain. In: International conference on wireless and mobile computing, networking and communications (WiMob) IEEE, Dec 2012
7. Jo G, Choe HJ, Choo H (2013) Predictive handover scheme using mobility history in PMIPv6. Res Adapt Converg Syst 13
8. Tarbani NM, Chandavarkar BR (2011) Implementation of AAA server for PMIPv6 in NS-2. In: International conference on parallel, distributed computing technologies and applications, PDCTA 2011, Springer, Tirunelveli, India, pp 523–531, 23–25 Sept 2011. ISBN 978-3-642-24037-9
9. Tarbani NM, Alvi AS (2015) Use of global positioning system to track movement of mobile node in proxy mobile internet protocol version 6. In: International conference on computer and communication technologies. Springer, pp 103–108, online ISBN 978-81-322-2517-1, Sept 2015
10. Tarbani NM, Alvi AS, Bamnote GR, Pattalwar SV (2016) Prediction of next mobile access gateway by tracking locations of mobile node. In: International conference on information and communication technology for competitive strategies. ACM, ISBN: 978-1-4503-3962-9, Mar 2016

Surveillance System for Intruder Detection Using Facial Recognition

Mohammed Umraan Shaikh, Deepali Vora, and Abhishek Anurag

Abstract Facial recognition system is used widely to identify and verify the person's face from image or video source. With the continuous expansion of the surveillance system, surveillance cameras not only bring convenience, but also produce a massive amount of monitoring data, which poses huge challenges to storage, analytics, and retrieval. The smart monitoring system equipped with intelligent video analytics technology can monitor as well as pre-alarm abnormal events or behaviors. Here, propose system will detect the intruder and inform the security within seconds. The Nvidia Jetson Nano board will be used to compute convolutional neural network algorithm for the facial recognition process. The basic idea will be to use this system where a database can be stored of the existing faces. The system will then take the data from the surveillance camera and run facial recognition algorithm on it. It will match all the faces with the ones already stored in the database and if it finds any face which is new, it will send an alert to the security personnel. This will help to increase the security of the place where there are many people gathered at a time, for example, schools, colleges, universities, etc.

Keywords Facial recognition · Biometric · Nvidia Jetson Nano · Surveillance

1 Introduction

Traditional video surveillance can only provide simple functions such as video capture and storage. It cannot automatically alarm for any intruders. In order to

M. U. Shaikh (✉) · D. Vora
Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India
e-mail: umraans@gmail.com

D. Vora
e-mail: deepali.vora@vit.edu.in

A. Anurag
CSE, Pune, India
e-mail: anuragabhishek@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_18

find any intrusion in real time, monitor personnel need to constantly observe the video. In this case, the monitor faces dozens of surveillance video images, which is easy to fatigue. It may not be able to respond in time to intrusion due to lack of concentration and loose key information in the video [1].

In addition, because a large amount of surveillance data, video needs to be stored for months or years, and it will result in a large storage cost. Therefore, the intelligent video surveillance system is urgently needed to assist the monitoring personnel to use the intelligent detection technology to process, analyze, and understand the video signal while retaining the original video key information and automatically detect the target and location information without manual intervention [1].

In the event of an intruder, an alarm is issued in time to effectively assist the monitoring personnel. The traditional moving target detection method can only detect frames with moving targets, but cannot understand the semantic content in the video. With the development of deep learning, advanced target detection, semantic understanding, and instance segmentation techniques are emerging, which realizes the semantic understanding of video content and improves the accuracy. In recent years, with the advancement of video analysis technology, intelligent video surveillance systems have developed rapidly, and a large number of representative technical achievements and applications have emerged [2]. With the help of embedded video analysis algorithms, the intelligent surveillance camera can identify the abnormal behavior of the monitoring site without manual intervention. Typical anomalous behaviors include entering and leaving an area, moving quickly, gathering, and so on.

The intelligent video analytics model (IVAM) also known as human object detection model will be used. The block diagram below shows how the model process flows.

Figure 1 shows the model for proposed surveillance system. The basic flows runs as, first, the data is fed as video input. From the video input, the image will be acquired, and each frame will be segmented. The IVAM will be applied on the

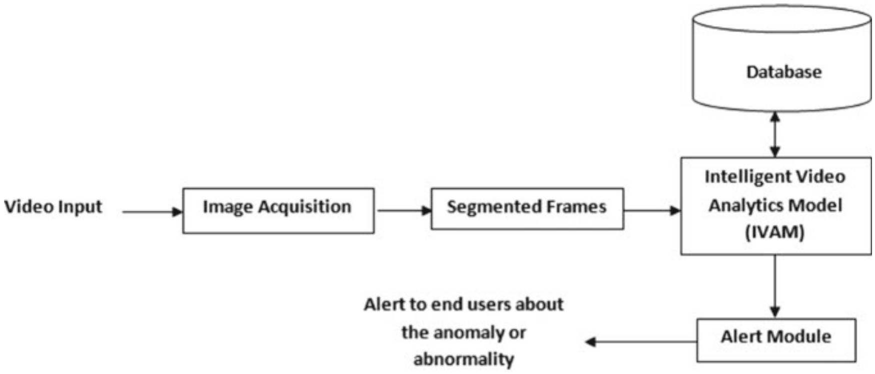

Database
Video Input
Image Acquisition
Segmented Frames
Intelligent Video Analytics Model (IVAM)
Alert to end users about the anomaly or abnormality
Alert Module

Fig. 1 IVAM block diagram

segmented frames. In this step, the feature extraction will be done. The data from the database will be matched to the output from IVAM. If it does not match, then the alert will be triggered to notify the security personnel. Once the intruder is detected, the image of the intruder will be snapshot along with the timeframe and will be saved in the database as evidence [3].

2 Literature Survey

Video monitoring and analysis for facial recognition problem has been studied by many researchers and different techniques/approaches have been suggested in Table 1.

3 Proposed System

To address the above issues in facial recognition, we propose a better surveillance system using Nvidia Jetson Nano instead of Raspberry Pi. Jetson Nano have much higher AI performance than Raspberry Pi. It has higher RAM and GPU which provides faster results compared to the Raspberry Pi. Jetson Nano board runs multiple neural networks simultaneously, and hence, it provides accurate results in much less time. The proposed system will be a much better surveillance system which will run on the Jetson Nano. The proposed system will be evaluated on the basis of efficiency, accuracy, and runtime. The OpenCV library will be used on the Jetson Nano to run the system. Figure 2 shows the process flow for the facial recognition system. The video input will be fed to system [19, 20].

- Face detection: First, the system will detect all the faces from the video feed. Here, the multi-task CNN algorithm will be used. This algorithm runs three neural networks. The first network will detect the face by scaling the image to a particular size. Jetson Nano supports DeepStream SDK and samples. These will help in face detection.
- Face alignment: The second neural network will align the face and add a padding to the faces which are partially outside the screen/bounding box. The third network will provide the final result, i.e., the aligned face.
- Feature extraction: FastICA algorithm to be used. It will add the geometric feature to the face for independent component analysis.
- Feature matching: Again the multi-task CNN algorithm will be used to match the features with the existing ones in the database.
- Results: Finally, if the features does not match, then the alarm is triggered to the security.

The proposed system uses Jetson Nano board by Nvidia which will provide better performance as compared to the traditional Raspberry Pi boards. Table 2 shows the

Table 1 Literature survey

Authors	Abstract	Algorithms	Inference from existing system
Lee and Shin [1]	Proposes a new process of ODTS by combining deep learning-based object detection network and object tracking algorithm, and it shows dynamic information of an object for a specific object class which can be obtained and utilized	R-CNN, object tracking algorithm	Adds CADA that discriminates every cycle based on dynamic information of the car objects It was possible to detect the accidents within 10 s
Mondal and Chatterjee [2]	This paper proposes face recognition involves a large number of challenges when it comes to deal with the visual analysis. Thus, this development or approach plays a significant role in security applications such as legal documents identification and identification of terrorists in public places	CNN algorithm	Runs perfectly on the standard benchmark dataset as well as the custom dataset prepared by us Performance has been better compared to the traditional approaches
Zhuang and Guan [3]	A novel method based on deep learning to solve the adverse impact imposed by illumination variation in the face recognition is proposed in this paper	Log-Gabor algorithm, LBP algorithm	Results have shown that the proposed method has superior performance comparing with some state-of-the-arts
Liu et al. [4]	This paper proposes a method based on mask R-CNN for intelligent monitoring of indoor surveillance video	R-CNN algorithm	Uses advanced target detection and instance segmentation technology, which effectively retains the key information in the original video

(continued)

Table 1 (continued)

Authors	Abstract	Algorithms	Inference from existing system
Ali et al. [5]	In this paper, an edge-based system for deep learning is proposed for efficient and large-scale video stream analytics. Using the infrastructure, an object recognition scenario was implemented	CNN algorithm	This shows 71% efficiency gain in the throughput of the system by employing a combination of edge, in-transit, and cloud resources when compared to a cloudy-only approach
Shan [6]	This paper proposes deep learning algorithms, such as CNN could provide high accuracy for great number of applications including ADAS and video surveillance analytics. Considering processing speed and energy efficiency, FPGA is a good hardware to construct customized CNN solution	CNN algorithm	System processes 16 channels of continuous input video with the resolution of 1080p. The FPGA used is Xilinx MPSoC ZU9 and the whole board including this FPGA only cost about 50 Watts with peak performance at 5.6 TOPS
Balasundaram and Chellappan [7]	The main objective of this paper is to design and implement a novel intelligent video analytical model as a human object detection method for surveillance video	Intelligent video analytics algorithm	The IVA model outperforms most of the existing approaches in terms of object detection and classification with less error IVA model has achieved 99.77% and 98.19% accuracy in correctly classifying the normal and abnormal frames

(continued)

Table 1 (continued)

Authors	Abstract	Algorithms	Inference from existing system
Sengur et al. [8]	Face recognition systems are now being used at industrial level throughout the world. Despite the great advances, counteracting face spoofing attacks has yet proven to be a challenging task. Therefore, in this paper, as a first step, we have investigated two well-adopted CNN models for face PAD without any fine-tuning	CNN algorithm	Investigation of two well-adopted CNN models for face PAD is without any fine-tuning. Experimental results on two publicly available databases are presented
Bailas et al. [9]	In the early age, smart city applications were relying on a cloud-centric approach. Thus, in this paper, it is explored that video analytics at the edge using a Dell edge gateway 5000. They can process more than one frame per second where most CCTVs only capture one image per second	Crowd density, ResNet50, crowd counting	The algorithm can process more than one frame per second where most CCTVs only capture one image per second which make them suitable for near real-time crowd monitoring The size of the frame highly impacts the performance It can increase by a factor of 5 the processing time at the edge
Qu et al. [10]	With the advent of the era of big data, deep learning theory has been rapidly developed and applied, especially in the field of image recognition. Based on the principle of CNN, a method of realizing face recognition on FPGA is presented in this paper	CNN algorithm	The speed of face recognition system is 400FPS, the recognition rate is 99.25%, and what is more, it has good robustness, which means it can complete the recognition function under most light conditions. Compared with the existing results, the indicators of the system have been significantly improved

(continued)

Table 1 (continued)

Authors	Abstract	Algorithms	Inference from existing system
Napiorkowska et al. [11]	Detection of objects in images has been long used in computer vision applications in fields such as surveillance or robotics. This paper shows how one of the networks developed for the ImageNet challenge can be applied to satellite imagery for object detection using three examples: roads, palm trees, and cars	VGG algorithm	Approach is good at finding objects that might have different colors and slightly varying shapes, which cannot be achieved as easily using more common techniques in remote sensing such as random forest or support vector machine
Yaseen et al. [12]	A system to perform video analytics is proposed using a dynamically tuned convolutional network. Videos are fetched from cloud storage and preprocessed, and a model for supporting classification is developed on these video streams using cloud-based infrastructure. A key focus in this paper is on tuning hyper-parameters associated with the deep learning algorithm used to construct the model	Video analytics algorithm	System is accurate with an accuracy of 0.97 as well as precise with a precision of 0.96, respectively. The system is also capable of coping with varying number of nodes and large volumes of data The time required to analyze the video data depicted an increasing trend with the increasing amount of video data to be analyzed in the cloud
Ran et al. [13]	In this work, they developed a measurement-driven frame-work, deep decision, that chooses where and which deep learning model to run based on application requirements such as accuracy, frame rate, energy, and network data usage	CNN algorithm	Deep decision can make smart decisions under variable network conditions. It does not include object tracking to reduce the frequency of running deep learning, generalizing the algorithm for a larger set of edge devices, and customizing the algorithm for different categories of input videos

(continued)

Table 1 (continued)

Authors	Abstract	Algorithms	Inference from existing system
Elmahmudi and Ugail [14]	In this paper, the use of deep learning approach for face recognition using partial face data is studied. Based on the popular CNN and using the VGGF model for extracting features from face, two different classifiers namely the cosine similarity and the linear support vector machine for classification were utilized	VGG face model, CNN algorithm	Experimental results validate that the cheek, nose, forehead, and mouth have low recognition rates. On the other hand, for top half of the face, right or left half and for three-fourth of the face, the recognition rates reach 100%. In addition, the cosine similarity measure greatly improves the performance of the classification when compared to the SVM
Kurban and Bilgic [15]	In face recognition systems, variables such as direction of light, facial expression, and reflection make identification difficult. With biometric fusion, both safe and high performance results can be achieved	VGG face model	The results show that the S gesture can be used for authentication with this method. These two biometrics are fused as a score level by sum rule to increase both safety and robustness
Tahboub et al. [16]	In this paper, a two-stage quality-adaptive convolutional neural network to address the problem of a changing video data-rate is proposed	CNN algorithm	Experimental results demonstrated that when adaptive data-rate streaming is used, our proposed quality-adaptive approach reduces the miss rate compared to the baseline detector
Sharma et al. [17]	Proposed algorithm is a face recognition algorithm from video using generalized mean deep learning neural network. The performance of the proposed algorithm is tested on two most commonly used databases, i.e., PaSC and YouTube databases	CNN algorithm	The results proved that the proposed algorithm is better in terms of identification accuracy

(continued)

Table 1 (continued)

Authors	Abstract	Algorithms	Inference from existing system
Burney and Syed [18]	In this paper, it is shown that the mid-level descriptors of the groups can be used to classify crowd videos, and also similar accuracy can be obtained if the whole crowd is considered as a single entity	CNN algorithm	Results shows the additional tasks of identifying groups, and computing their features and then combining them to define the crowd can be reduced to single step of computing the features of the crowd

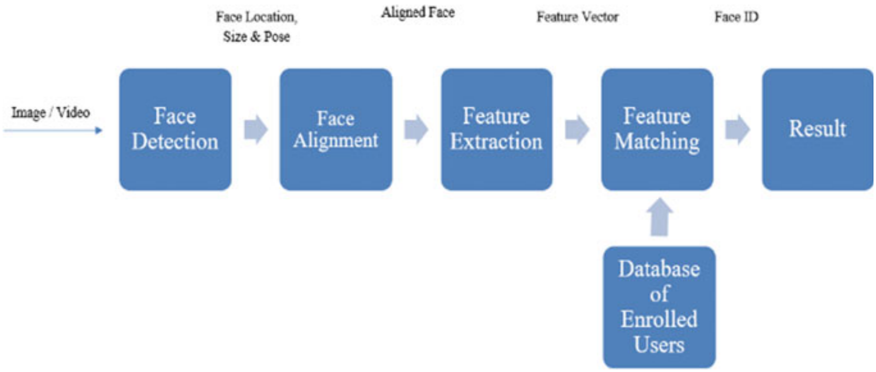


Fig. 2 Block diagram of proposed system

comparative study between the Jetson Nano board and Raspberry Pi board based on 13 parameters.

Table 2 shows that Jetson Nano has higher AI performance which is useful for applying deep learning algorithms and which will give a better edge at running the proposed surveillance system. The high-powered CPU, GPU, and RAM also help in faster computations of complex algorithms as well as it helps to run multiple neural networks simultaneously. The video output of the Jetson Nano has 4 K resolution thus providing clear image. The only constraint which goes against the proposed hardware is the price of the hardware but given the specifications and computation power it is worth the money.

4 Conclusion

The above-proposed surveillance system for the facial recognition promises a much efficient and highly accurate in giving the results. The use of the Jetson Nano will be critical in the performance of the entire system. Due to the use of Jetson Nano, the entire process will run much faster and multiple neural networks will run simultaneously. Moreover, the use of multi-task CNN algorithm for face detection, alignment, and feature matching will also provide much better results. One of the major advantages of the proposed system is that Jetson Nano uses the same CUDA and OpenCV libraries along with its own Jetpack SDK's. Along with these, Nvidia's TensorRT will also help for high-performance deep learning inference. So using the existing libraries, it will be much easier to design an efficient system which will work much faster and more accurately.

Table 2 Comparative study [21–24]

	Jetson Nano board	Raspberry Pi 3A+	Raspberry Pi 3B+
AI performance	472 GFLOPS	21.5 G FLOPs (est*)	21.4 GFLOPs (est*)
CPU	1.4 GHz 64-bit quad-core ARM Cortex-A57 MPCore	1.4 GHz 64-bit quad-core ARM Cortex-A53	1.4 GHz 64-bit quad-core ARM Cortex-A53
GPU	128-core Nvidia maxwell	Broadcom VideoCore IV	Broadcom VideoCore IV
RAM	4 GB LPDDR4	512 MB LPDDR2 SDRAM	1 GB LPDDR2 SDRAM
GPIO header	40-pin	40-pin	40-pin
Board dimensions	100 X 79 mm	65 X 56 mm	85 x 56 mm
Wireless	None	Dual-band 802.11ac wireless LAM, Bluetooth 4.2/BLE	Dual-band 802.11ac wireless LAN, Bluetooth 4.2
Ports	4x US8 3.0, wired		
Ethernet 10/100/1000 Mbps	1 USB 2.0	4 USB 2.0, wired Ethernet up to 330 Mbps	
Multimedia	2160p30 (H.264)	1080p30 (H.264)	1080p30 (H.264)
Video output	HDMI, display port (4K)	HDMI, display serial interface (DSI)	HDMI, display serial interface (DSI)
Camera serial interface	Yes	Yes	Yes
M.2 key E slot	Yes	No	No
Price	$99	~$25	~$35

References

1. Lee KB, Shin HS (2019) An application of a deep learning algorithm for automatic detection of unexpected accidents under bad CCTV monitoring conditions in tunnels. In: IEEE, 2019 international conference on deep learning and machine learning in emerging applications (Deep-ML), Istanbul, Turkey. https://doi.org/10.1109/Deep-ML.2019.00010
2. Mondal I, Chatterjee S (2019) Secure and hassle-free EVM through deep learning based face recognition. In: IEEE, 2019 international conference on ML, big data, cloud and parallel computing (COMITCon), Faridabad, India. https://doi.org/10.1109/COMITCon.2019.8862263
3. Zhuang L, Guan Y (2019) Deep learning for face recognition under complex illumination conditions based on log-gabor and LBP. In: 2019 IEEE 3rd information technology, networking, electronic and automation control conference (ITNEC), Chengdu, China. https://doi.org/10.1109/ITNEC.2019.8729021
4. Liu YX (2019) Intelligent monitoring of indoor surveillance video based on deep learning. In: IEEE, 2019 21st international conference on advanced communication technology (ICACT), Korea. https://doi.org/10.23919/ICACT.2019.8701964
5. Ali M et al (2018) Edge enhanced deep learning system for large scale video stream analytics. In: 2018 IEEE 2nd international conference on fog and edge computing (ICFEC), Washington DC, USA. https://doi.org/10.1109/CFEC.2018.8358733

6. Shan Y (2018) ADAS and video surveillance analytics system using deep learning algorithms on FPGA. In: IEEE, 2018 28th international conference on field programmable logic and applications (FPL), Dublin, Ireland. https://doi.org/10.1109/FPL.2018.00092
7. Balasundaram A, Chellappan C (2018) An intelligent video analytics model for abnormal event detection in online surveillance video. J Real Time Image Process: 1–16. https://doi.org/10.1007/s11554-018-0840-6
8. Sengur A et al (2018) Deep feature extraction for face liveness detection. In: IEEE, 2018 international conference on artificial intelligence and data processing (IDAP), Turkey. https://doi.org/10.1109/IDAP.2018.8620804
9. Bailas C, Marsden M, Zhang D (2018) Performance of video processing at the edge for crowd monitoring applications. In: IEEE, 2018 IEEE 4th world forum on internet of things (WF-IoT),Singapore. https://doi.org/10.1109/WF-IoT.2018.8355170
10. Qu X, Wei T, Peng C, Du P (2018) A fast recognition system based on deep learning. In: IEEE, 2018 11th international symposium on computational intelligence and design (ISCID), Hangzhou, China, China. https://doi.org/10.1109/ISCID.2018.00072
11. Napiorkowska M, Petit D, Marti P (2018) Three applications of deep learning algorithms for object detection in satellite imagery. In: IEEE, IGARSS 2018—2018 IEEE international geoscience and remote sensing symposium, Valencia, Spain. https://doi.org/10.1109/IGARSS.2018.8518102
12. Yaseen MU, Anjum A, Rana O, Antonopoulos N (2018) Deep learning hyper parameter optimization for video analytics in clouds. IEEE Trans Syst Man Cybern Syst: 253–264. https://doi.org/10.1109/TSMC.2018.2840341
13. Ran X, Chen H, Zhu X, Liu Z, Chen J (2018) DeepDecision: a mobile deep learning framework for edge video analytics. In: IEEE INFOCOM 2018—IEEE conference on computer communications, Honolulu, HI, USA. https://doi.org/10.1109/INFOCOM.2018.8485905
14. Elmahmudi A, Ugail H (2018) Experiments on deep face recognition using partial faces. In: IEEE, 2018 international conference on cyberworlds (CW), Singapore, Singapore. https://doi.org/10.1109/CW.2018.00071
15. Kurban OC, Bilgic A (2017) A multi-biometric recognition system based on deep features of face and gesture energy image. In: IEEE, 2017 IEEE international conference on innovations in intelligent systems and applications (INISTA), Gdynia, Poland. https://doi.org/10.1109/INISTA.2017.8001186
16. Tahboub K, Guera D, Reibman A, Delp E (2017) Quality adaptive deep learning for pedestrian detection. In: IEEE, 2017 IEEE international conference on image processing (ICIP), Beijing, China. https://doi.org/10.1109/ICIP.2017.8297071
17. Sharma P, Yadav RN, Arya KV (2016) Face recognition from video using generalized mean deep learning neural network. In: IEEE, 2016 4th international symposium on computational and business intelligence (ISCBI), Olten, Switzerland. https://doi.org/10.1109/ISCBI.2016.7743283
18. Burney A, Syed TQ (2016) Crowd video classification using CNN. In: IEEE, 2016 international conference on frontiers of information technology (FIT), Islamabad, Pakistan. https://doi.org/10.1109/FIT.2016.052
19. Test Data, https://motchallenge.net/data/MOT17/testdata, downloaded on 15 Nov at 6.15 pm
20. Face detection program using neural networks, https://towardsdatascience.com/how-does-a-face-detection-program-work-using-neural-networks-17896df8e6ff. Accessed on 17 Dec 2019 at 4.30 pm
21. NVIDIA Jetson nano specifications, https://www.cnx-software.com/2019/03/19/nvidia-jetson-nano-developer-kit/. Accessed on 13 Dec 2019 at 5.30 pm
22. Raspberry PI 3A+ Specifications, https://www.cyberciti.biz/hardware/raspberry-25-pi-3-model-a-released-complete-specs-and-pricing/. Accessed on 13 Dec 2019 at 5.45 pm
23. Raspberry PI 3B+ Specifications, https://www.cyberciti.biz/hardware/raspberry-pi-3-model-b-released-specs-pricing/. Accessed on 13 Dec 2019 at 5.40 pm
24. Jetpack SDK, https://developer.nvidia.com/embedded/jetson-nano-developer-kit. Accessed on 16 Dec 2019 at 7.45 pm

Clinical Risk Prediction of Acute Coronary Syndrome Using Deep Learning Approach

Swati Suchak and Uttara Gogate

Abstract Acute coronary syndrome (ACS) is a severe cardiovascular disease which is leading to death, and it is serious long-term disability globally. Prediction of ACS is important for diagnosing earlier to treat it. Previous ACS models based on small set of risk factors and predictive variables used to simplify the score calculation are numbered. This study has overcome the problem of existing system which developed stacked and regularized denoising predictive auto-encoder (SDAE) model to find clinical risks of ACS patients from huge collections of electronic health records (EHR) in biomedical engineering field. It determines patients at similar risk-level characteristics, low risk, high risk, and medium risk and preserves the results. This prediction approach is totally based on real-time dataset which we processed using SVM algorithm. Finally, SVM gives the accurate prediction of patients; if the prediction model shows the high risk, then doctor sends some precaution details to patients and his relatives for saving the life of patients. This approach is validated with more than 2000 real clinical dataset consisting of patient samples. The approach followed by us remains robust and also this model is very helpful for doctors and patients which can detect the cancer in an initial stage.

Keywords Acute coronary syndrome (ACS) · Deep learning · Electronic health records (EHR)

1 Introduction

Acute coronary syndrome (ACS) includes unstable angina and myocardial infarction (MI) which are the common types of coronary heart diseases (CHD). This disease occurs if the heart muscle does not receive sufficient oxygen in blood. Due to this,

S. Suchak (✉)
PG Scholar, Alamuri Ratnamala Institute of Engineering & Technology, Mumbai, India
e-mail: swatisuchak0702@gmail.com

U. Gogate
Associate Professor, Shivajirao S. Jondhale College of Engineering, Mumbai, India
e-mail: uttara.gogate16@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_19

patient may face problems like morality; ACS became an important health issue and its treatment cost and the prolonged chronic course mortality rates are increased [1]. Every year ACS is estimated to affect millions of people in china and USA [2]. Approximately, 1/3rd population are at risk of ACS disease [3]. In USA, more than 50,000–60,000 people are dying before taking them to hospital because of ACS attack. All the above-stated adverse effects have to be ended by diagnosing patients and prevent this disease [4]. Prevention of ACS may help both the patients and society. These are mentioned in points [5]. Clinical risk prediction has been done with tool so, that patient takes care of his disease and he can undergo proper disease treatment steps [6]. Clinical risk prediction helps clinicians to predict the chance of cardiac dysfunctions (such as myocardial infarction, death). This proposed tool helps in taking timely and appropriate intervention strategies to those at high risk of cardiac arrests, and to motivate patients to remain alert, following these strategies the patients can reduce high mortality. Extensive research work carried on clinical risk estimation of ACS. Earlier ACS risk-scoring tools like Global Registry of Acute Coronary Events (GRACE) and thrombolysis in myocardial infarction (TIMI) have been developed for monitoring population samples over long period. Here, we describe introduction about our work in which some important terms such as deep learning, clinical risk prediction, artificial neural network (ANN), MongoDB database that we have used in this research and plays very important role. In the proposed model, patient's information is extracted from admission records of hospital for clinical risk prediction models of ACS patients. Machine learning techniques using deep learning methods are new technology. They require properly organized datasets to provide correct answers to the questions we shoot them. A business depending on machine learning application needs to give time, resources and take risks. Artificial neural network (ANN) is capable of remembering the training set which consists of thousands of records and it also gives answer with 100% accuracy.

1.1 Clinical Risk Prediction

Clinical risk prediction models are used in health care for variety of applications as described in [7, 8]. Mostly they are targeted to identify patients with higher risk of ACS disease with levels like high, average, and low [9]. A model called Global Registry of Acute Coronary Events (GRACE), risk-scoring model, is a popular and well-accepted risk scores of ACS disease which was developed to predict clinical risk score on an individual patient each [7]. However, models along with this line have been estimated using a small set of patient features especially from highly stratified cohorts [1]. Bayesian networks and fuzzy inference systems have been proposed to find the potential of EHR data for risk prediction [2]. Karaolis et al. used C4.5 decision tree algorithm to retrieve essential risk factors of heart events from electronic health records [3]. Hybrid model was developed to identify risk factors of heart disease in patients. My proposed stacked denoising autoencoder model (SDAE) will adapt dynamic treatment information in EHR data to improve the performance of prediction for ACS disease. It helps in the clinical prediction of coronary heart diseases from a large volume of EHR data. Though it is giving answers to many questions, it may prove useless when new data is given. This problem is known as over-fitting (or over-training). To overcome these limits, deep learning algorithms are used.

2 Related Work

We are using SVM for the following reasons as in points [1–4, 7–9]. SVMs need less memory. SVM is less prone to errors. They perform text classification in her systems and pattern recognition in echo-cardiograph imaging [9]. They help to stratify clinical violation risk and physicians make decisions. SVM is feasible for large and complex data which is nonlinear, such as "omic" data [7]. SVM accuracy enhances processing time. The wrong choice in kernel functions may lead to errors. SVM works even with small datasets [1]. The decision tree algorithm encounters over-fitting because of the small datasets. It is used with a series of yes/no questions. SVM classifies data into categories [2]. In addition, it can be used in clinical decision making. Random forest algorithm is an extension to decision tree algorithm. In this algorithm, decision trees are combined together and every decision tree is trained independently [3]. Random forest algorithms compute angiography, readmission, high frequency risk, and survival prediction models [4]. The Naive Bayes classifier is probabilistic classifier obtained from Bayes theorem. It executes small training datasets which is be used in text classification problems in clinical violation risk factor identification and decision-making systems. Fuzzy logic is used in areas such as prediction of early coronary artery disease and mortality. K-nearest neighbor is nonparametric methods. It works on small training datasets. Fuzzy is used in ECG interpretation problems. K-nearest neighbor takes more space and time when using large datasets. There are many data mining algorithms like decision trees, fuzzy inference systems, Bayesian networks, etc., and have been said to explore the uses of EHR data for clinical risk prediction. When considering an example, Tay et al. gave a neural-inspired deep learning ANN algorithm for risk prediction by using EHR data. There exists a Bayesian analyses which predict human clinical adverse events which very well helped in drug development programs. In existing work, a genetic fuzzy system is designed for risk prediction in unstable angina. It also evaluated on a dataset which is collected from a Chinese hospital in EHR patients. We previously developed a probabilistic topic modeling. The said model finds risk stratification by exploring the potential of electronic health recording unsupervised fashion.

3 Problem Definition

We have studied literature survey and found the exact research gap of our work, and every author suggested acute coronary syndrome (ACS) as a common and severe cardiovascular disease, and it is a leading cause of death and the principal cause of serious long-term disability globally. Many authors processed this EHR dataset and have done a lot of work but we found that no one can decide the range of syndrome. Basically, here we have focused on deep learning approach with a greater number of layers, and it can potentially extract abstract and invariant features for better performance of patient classification. The ability of inference on a large volume of heterogeneous EHR data is particularly suitable for our aim. Therefore, this paper proposes a novel approach for clinical risk prediction of ACS based on deep learning. Among various deep learning models, the stacked denoising autoencoder (SDAE) has particular advantages such as rapid inference and the ability to reconstruct features. After considering all reconstruction learning phases, softmax regression layer is appended on the top of the resulting reconstructed feature representation layer. This can help in solving clinical risk prediction. Our proposed model can learn more patient feature representations and improve the performance of clinical risk prediction. The main objective of this research is stated below. The main objective of this work is to compare different deep learning algorithms in order to detect risk factors (Levels) of coronary heart disease in a real-world, high dimensional Electronic Health Record data sets with heterogeneous data. The goal of the work is to find different algorithms and find algorithm that is suited for large datasets working. In order for the chosen algorithms to scale different levels of risk factors in ACS with an increasing size of the patient sample size, find an algorithm that achieves accuracy with less preliminary processing of the data.

4 Proposed System

The proposed framework contains modules like Administrator, Clinician, and Researcher. The Administrator role is to make registration that is to add or remove users and also the role of Administrator is to provide the analysis of overall process. This model shows the complete process in graphical format. The Clinician's module role is to find risk factors related to acute coronary syndrome symptoms and define level of risk factor. The complete process can removed the noising, and disease regarded information is collected from samples of various patients and it is stored in MongoDB database. Researcher module role is to provide the parameters for the analysis and it has to find risk levels based on predictions by applying softmax regression algorithm to the provided data. Process flow of proposed framework can be seen verified in the below diagram which is given in proposed framework use in vector graphs-1 (Fig. 1).

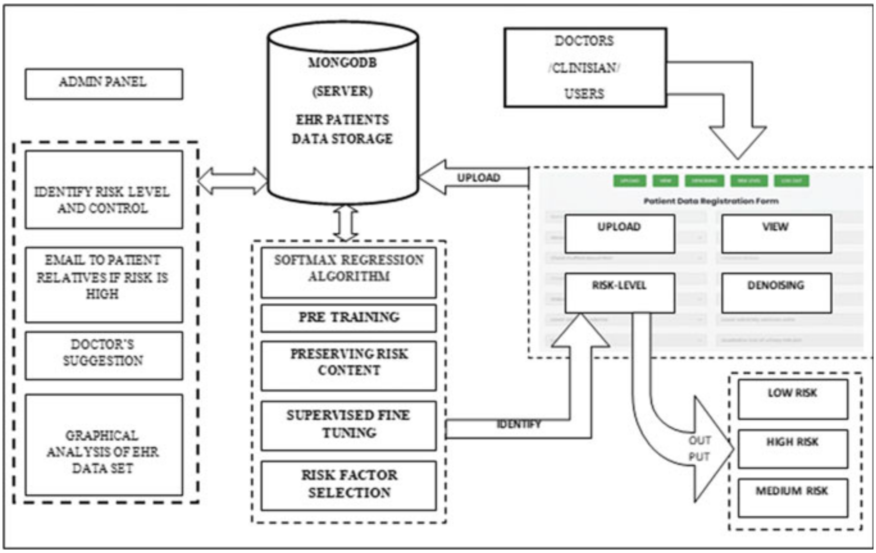

ADMIN PANEL
IDENTIFY RISK LEVEL AND CONTROL
EMAIL TO PATIENT RELATIVES IF RISK IS HIGH
DOCTOR'S SUGGESTION
GRAPHICAL ANALYSIS OF EHR DATA SET
MONGODB (SERVER) EHR PATIENTS DATA STORAGE
DOCTORS /CLINISIAN/ USERS
UPLOAD
Patient Data Registration Form
UPLOAD
VIEW
RISK-LEVEL
DENOISING
SOFTMAX REGRESSION ALGORITHM
PRE TRAINING
PRESERVING RISK CONTENT
SUPERVISED FINE TUNING
RISK FACTOR SELECTION
IDENTIFY
OUT PUT
LOW RISK
HIGH RISK
MEDIUM RISK

Fig. 1 Overall frame work of ACS system

In this framework, we discuss the training and validation of ACS clinical risk prediction model by using our proposed approach. This model is a symmetrical neural network which is mainly used for learning all the features of a dataset by using techniques of unsupervised learning so as to build a deep learning architecture by using K hidden layers and is trained using greedy layerwise learning in unsupervised mode.

The above framework has been used for classification problems, and the features reconstructed by syndrome processed are in unsupervised manner. While taking other side, risk information is contained in the training dataset [10]. This dataset is included into learning so as to make the reconstructed features that are beneficial for the future tasks of clinical risk prediction. We add two specific constraints like intra-risk-level affinity and inter-risk-level repulsion. These important constraints help this framework to preserve clinical risk information of training on patient samples. These important constraints use the reconstructed feature representations of patients with the same risk level and also on different risk levels.

4.1 Supervised Fine-Tuning

After the completion of pretraining, we added a softmax regression layer. We placed it on the top of the reconstructed layer to construct a deep artificial neural network which is named as stacked denoising autoencoder using softmax regression model. Denoising performs clinical risk prediction task. Here, we fine-tuned and remove it by using algorithm techniques of backpropagation. This backpropagation minimizes the cross-entropy loss by taking help of softmax regression layer.

4.2 Risk Factor Selection

In our proposed model, risk factor selection strategy identifies risk factors of ACS patients with different risk levels. We assume that patient features with less reconstruction errors are more reconstructible. Patient with more reconstructible features are likely to bear the below characteristics. With patient sample within a specific risk level, there is small reconstruction error between the original and reconstructed data as it learns regularized supervised model. We can also argue that different risk-level patients behave differently according to their risk factors and produce a big reconstructed error. We expect a big reconstruction error in the discriminative learning. This happens as there is mismatch of input patient sample and the testing model used for implementation.

4.3 Deep Learning Approach

Deep learning is an artificial intelligence concept applying set of algorithms in machine learning that model very high-level abstractions in the data by using architectures which are composed of multiple nonlinear transformations [8]. Stacking the nonlinear transformations is the basic idea used for deep learning algorithms [9]. The more layers the data goes through within the deep learning architecture, the more tough to use nonlinear transformations construction [7]. Stacked denoising autoencoder (SDAE) is the most investigated deep learning architectures. It is a symmetrical neural network, which is used for learning the features from dataset in unsupervised fashion [1, 11]. Autoencoder in acute syndrome is trained to reconstruct input that is clean "repaired" from the version that is corrupted [2]. Chen et al. proposed a deep learning model for phenotyping from electronic health records, and the model is predictive modeling of chronic diseases [3]. Tran et al. proposed a computational framework to harness EHR with a type of deep neural network. Many deep learning models promise results in learning good representations in an unsupervised manner. This model performs supervised tasks, such as clinical risk prediction. On the contrary, in the proposed paper, deep learning models using classes, it can reconstruct more discriminative features by using supervised task. Thus, in the proposed model, a regularized syndrome for clinical risk prediction of ACS has been developed in our research.

4.4 MongoDB

It is big data management system which is used to manage and process document-based unstructured data. It stores the document and continually creates replica for it so it provides high performance, high availability, and easy scalability. For the features of replication, it is more popular for storing, creating, and retrieving documents as it increases the availability and accessibility. It is a smallest part of Hadoop system which is used to store the structured, semi-structured, and unstructured documents. In this research, we have used the MongoDB to store the EHR dataset which is basically document (semi-structured)-based dataset which includes various form of dataset; so in MongoDB, there are different documents to store different types of data and each document has unique system generated key hence called document-oriented storage. It also provides auto-sharing which is a method to distribute data across machines. Aggregation feature provides multitasking as it can batch process data and get single result after doing some sequence of tasks on group of data. MongoDB is used for this research work as it is supporting NoSQL which is multimodel, easily scalable, distributed, and flexible.

5 Algorithms

In this section, we describe different process and algorithm which define the risk level and communication process with MongoDB and stored the process output in it.

Algorithm 1: Risk Identification

```
1. Input: Read the input parameter from documents.
2. Remove the unwanted noise from documents.
3. Preprocess the input parameter:
       If the parameter range is inside the defined range or valid value then:
       Training required using learning regression.
       else:
       Post the documents for testing phase.
4. Process all training documents using softmax function for clustering process (call
   softmax function).
5. Output of softmax function is risk level: low, high, medium in training phase.
6. Calculate the risk prediction of current patient (training Phase).
7. Trained machine predicts the class of risk using deep learning.
8. Finally, we identify the risk category and stored into the processed documents.
```

Algorithm 2: Softmax

```
1. Process all documents on individual cluster.
2. Merge all cluster as per the specific range value
3. Calculate the total number of cluster i.e. Tc
4. Check the documents range as clusterwise

      If within high range then:
      Stored in high-risk documents.
      If within medium range then:
      Stored in medium-risk documents.
      Else
      Stored in low-risk documents.

5. Sort this risk category as per cluster and merge all risk level and train the machine
   using deep learning category.
```

Risk factor

We propose a risk factor selection strategy to identify informative risk factors for ACS patients within different risk levels [12]. We assume that patient features with lower reconstruction errors are more reconstructible, and more reconstructible patient features are more likely to bear the underlying characteristics. In this regard, given a patient sample within a specific risk level, we expect a small reconstruction error between the original data and the reconstructed one can be obtained by the learned regularized standard model. On the other hand, we argue that all different risk-level patient samples behave differently with respect to their risk factors so as to produce a large reconstructed error. Thus, we expect a large reconstruction error during the discriminative learning due to the mismatch of an input patient sample and the model used for testing.

6 Experimental Study

We have taken the generalized dataset of 2000 patient record of all different fields. Every field of EHR dataset is very important for processing the records. Approval and suggestions were taken from cancer specialist doctors and institutional review department to carry this research work.

6.1 Informative Risk Factor Selection

The proposed model is used to find informative risk factors from electronic health records dataset. Top 10 risk factors for the patient group with same risk level were taken and confirmed their validity with the clinical experts, and patient's features were divided into two classes 0 or 1 (saying patient case has this particular mentioned feature or not) in the model learning process. Creatinine, smoking status, etc., are validated inside the clinical cohort study. To apply this proposed model in clinical practice, we should consider the simpler model implementation which investigates the simplicity and accuracy. Clinicians are interested to know risk factors that collect to obtain a predictive performance. We apply our risk factor selection strategy to find risk factors with regard to each risk level. We used top 20 in the union of these risk factors to proposed model for our ACS risk prediction model.

6.2 Software Details

Used software and algorithms in this research are operating system and Windows 7. Application server used in research is Tomcat Server 8.0. Frontend details are JavaScripting. Database server is used is MongoDB. Database connectivity used is Robomongo-0.8.5-i386. As the electronic health records are huge volumes of data and related with machine learning technique, Java is used for frontend design as the data is more secure and robust. Since GUI design is possible in Java, it is opted for my research work.

7 Result Discussion

In this graphical analysis, we processed the all documents using MongoDB and machine learning algorithm. Our algorithm shows the results in three categories like low, high, and medium risk (Fig. 2).

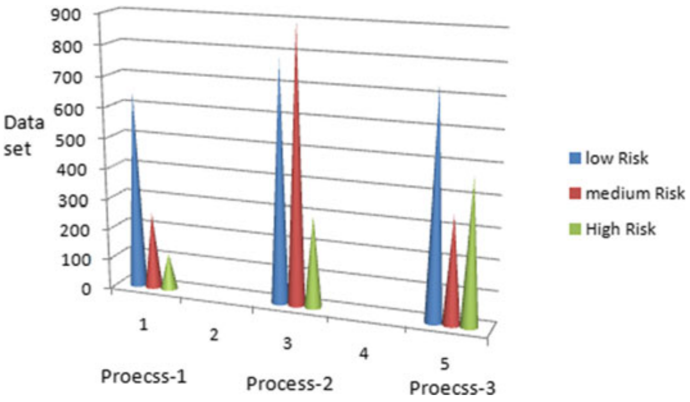


Fig. 2 Bar chart analysis

Above chart shows number of risk patients categorized in different risk levels. Depending upon the input values of patients' different test records, system assigned various risk levels. Maximum patients who get admitted as uneasy or restless feeling are coming in low-risk category. Some patients are not having any complaints initially, but their various tests show that they are at high risk and need to be treated quickly.

8 Conclusion

This work proposes learning approach to discuss the clinical risk identification problem of acute coronary syndrome (ACS) from electronic health record data. With proposed approach, we are able to utilize a large volume of heterogeneous electronic health records of patients to determine different ACS risk level, i.e., low, medium, and high. Our model is an accurate and robust clinical risk prediction model which gives the 90% accuracy to predict the level of cancer patients using the different dataset. In this work, we have used support vector machine (SVM) algorithm to achieve the best result and it also directly informs the patient's relatives about the high-risk level of the patient so they can take suggestions from expert doctors. The experiments in the research were conducted on a clinical dataset and the results show that the proposed model is able to achieve good performance in clinical risk prediction as compared to algorithms like state-of-the-art classification algorithms.

References

1. Huang Z et al (2015) A probabilistic topic model for clinical risk stratification from electronic health records. J Biomed Inform 58:28–36
2. Matheny M et al (2011) Systematic review of cardiovascular disease risk assessment tools. Technical Report. Agency for Healthcare Research and Quality, US
3. Karaolis MA et al (2010) Assessment of the risk factor of coronary heart events based on data mining with decision trees. IEEE Trans Inf Technol Biomed 14(3):559–567
4. Bandyopadhyay S et al (2015) Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data. Data Min Knowl Disc 29(4):1033–1069
5. Paredes S et al (2015) The CardioRisk project: improvement of cardio vascular risk assessment. J Comput Sci 9:39–44
6. Mega JL et al (2012) Rivaroxaban in patients with a recent acute coronary syndrome. The New Eng J Med 366(1):9–19
7. Wilson PW et al (1998) Prediction of coronary heart disease using risk factor categories. Circulation 97:1837–1847
8. Goff DC et al (2014) 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 129:S49–S73
9. Boersma E et al (2000) Predictors of outcome in patients with acute coronary syndromes without persistent ST-segment elevation. Results z from an international trial of 9461 patients. Circulation 101(22):2557–2567

10. Murray CJL, Lope AD (1997) Global mortality, disability, and the contribution of risk factors: global burden of disease study. Lancet 349(9063):1436–1442
11. Antman EM et al (2000) The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. J Am Med Assoc 284(7):835–842
12. Goodman SG et al (2009) The expanded global registry of acute coronary events: baseline characteristics, management practices, and hospital outcomes of patients with acute coronary syndromes. Am Heart J 158(2):193–201

Securing IoT Devices Generated Data Using Homomorphic Encryption

Anita Caudhari and Rajesh Bansode

Abstract Cloud computing is emerging trends. Cloud service provider provides services like network, application, storage, etc. People use cloud storage for storing and retrieving their data for future use. Internet IoT devices are memory and power constrained. All data generated by these devices are outsourced on cloud, but sharing private data to cloud is risky. Though it is in encrypted format, processing data on cloud is import task. Our research focuses on processing on encrypted data. GSW scheme is used to perform computation of an encrypted data, collect data from IoT devices, and encrypt it using GSW algorithm. Computation is performed on encrypted data like addition and multiplication on cloud server.

Keywords HE · Cloud · Operation

1 Introduction

Homomorphic encryption method carried out on encrypted data is ciphertext instead of plaintext. The main thing in homomorphic encryption is after performing computation on encrypted data, generated result should match with the computation performed on plaintext data. If we perform simple design of homomorphic addition where plaintext is 4 and 7 and generated result is 11, and when we do encryption of 4 and 7 by using simple method, just multiply plaintext by 3 we will get encrypted result 12 and 21; at decryption side, when we use method ciphertext divide by 3 we will get original result. IoT devices delegate encrypted data to cloud, if some computations are required on data, user has to download data and perform computation. This problem can be solved by using homomorphic encryption on cloud side [1, 2]. For preserving privacy of data, data encryption should be performed, before

A. Caudhari (✉) · R. Bansode
Mumbai University, Mumbai, India
e-mail: anitac@sjcem.edu.in

R. Bansode
e-mail: rajesh.bansode1977@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_20

uploading on cloud. This might bound the value of the information, but current trends in cryptography permit probing on encrypted information and execution of different actions on encrypted data, without revealing secret information [3]. An encryption system which permits random actions on ciphertexts is known as a fully homomorphic encryption (FHE) scheme. The initial FHE system was constructed by gentry, and several succeeding systems have quickly developed more practical, with enhanced performance and parameters. Subsequent FHE systems have called somewhat homomorphic encryption (SHE) system as an fundamental structure and use a method called bootstrapping to spread in FHE scheme [4]. Computation on encrypted data is ongoing research and different authors propose different algorithms to reduce computational cost. This type of model can be used in healthcare application to secure patient data [5].

2 Literature Survey

Public-key encryption with keyword exploration is based on homomorphic encryption in multiuser situation. In this method, DGHV homomorphic encryption applied cloud server to generate reversed encryption directory structure. DGHV does not use query trap door and permits several users to achieve encrypted keyword search over encrypted data. If number of queried keywords is 300, it takes 9 s [3]. The encoded information can be considered as a good solution to forget over these impediments on computing. Researchers have proposed another encryption form, homomorphic encryption (HE), that the third-party has the capacity to complete tasks on encrypted information. The homomorphic property makes HE schemes helpful in a extensive scope of protection saving applications by the homomorphic property, for example, electronic voting and advanced human services. The homomorphic encryption method was presented in 1978.

From that, numerous HE schemes have been proposed. HE plans can be isolated into two classes: somewhat and fully homomorphic encryption. The first categories give one homomorphic property [6]. A few homomorphism properties are offered by the full homomorphic. Since re-appropriated scrambled information under homomorphic schemes can be put away for a long-lasting, more often not by the equivalent key, in CC conditions, the cloud customers frequently access to cloud services utilizing asset-restricted gadgets. These plans should be advanced as far as security level and running time do work successfully. Author proposed, FPGA homomorphic encryption co-processor [7].

Cloud computing is used to convey on-request facilities (e.g., capacity, applications, networks, and servers) using the Internet. This is because of a few benefits given by cloud providers containing low costs, high administrations, power calculations flexibility, and adaptability. Regardless, failures over the privacy of sensitive information are as yet the main drawback constraining the adoption of cloud computing services [8]. Translating algorithm that can run on encrypted data translates basic operators like bitwise, arithmetic, and relational operators. Author discussed decision making, loop handling data structures like stack, queue and in that how these algorithms can be applied on encrypted data. Processed each algorithm and measured time required to process encrypted data [9, 10].

Hybrid homomorphic encryption model combines public key encryption and somewhat homomorphic encryption for reducing storage requirement. Here, message is encrypted using PKE and computation is carried out using FHE or SHE. It uses public key size 3 TB, multiplicative depth of 20, comparative analysis of proposed algorithm with SHE [11]. Homomorphic proxy re-encryption scheme permits different users to share data they outsourced homomorphically encrypted using separate public key, uses Damgard--Jurik cryptosystem, and evaluated the performance of proposed model in terms of time computation and encrypted data volume. Delegator will take 0.004 s, proxy, i.e., cloud 120 s, delegator 30 s for 92 × 122 pixel [12].

Secret keys implemented a lattice-based key-policy attribute-based encryption (KP-ABE) scheme. The specific KP-ABE plan can be utilized straightly for attribute-based access control (ABAC) applications. Graphics processing units (GPUs) are the state-of-the-art implementation and demonstrate that the homomorphic open key and ciphertext evaluation activities, which command the execution time of the KP-ABE plot, can be performed in a sensibly short measure of time [13].

Privacy-preserving data processing (PPDP) framework with the help of a homomorphic re-encryption scheme (HRES) expands fractional HE from a single client framework to a multiclient by offering ciphertext re-encryption. PPDP system can maintain seven essential activities over ciphertexts, which incorporate addition, subtraction, multiplication, sign acquisition, comparison, equivalent test, and variance [14]. Implemented optimized RLWE scheme is in Bayesian spam filter, decision tree, secure multiple keyword search.

Alhassan et al. [11] developed model for securing healthcare data maintaining patient data privacy which is important. Focused on NTRU-based implementation on GSW-HE, results show improvement of 58× in CPU performance. Performing operation on blood pressure data, person goodness of fit test health [15] to secure patient data. Proposed secure privacy preserving data aggregation scheme, based on bilinear pairing.

It is described that an integrated individual health data framework permits secure capacity and prepared the medical information in the cloud by utilizing an exhaustive homomorphic encryption model to protect information security. Medical cyberphysical system is based on different algorithms that are proposed to secure patient data. ECIECS, KP-ABE, CP-ABE, PAILIER, and FHE algorithm are used in securing healthcare data but main limitation is computation time [16]. Provided solution to protect users data directly into the cloud while preserving the user privacy uses cardiac risk factor algorithm on encrypted data to execute this algorithm, and it takes 3.2 s. HE is implemented on healthcare data perform computation on encrypted heart rate, provided average heart rate, LQTS, Min, and Max heart rate [17].

3 Methodology

Homomorphism If ($F1$, $*$) and ($F2$, $\otimes$) are two groups, then a function f: $F1 \to F2$ is a group homomorphism if $f(a * b) = f(a) \otimes f(b)$ for all x, $y \in F1$ Examples: $f(a) = e\ a$, $f(a) = \log(a)$, … (Fig. 1).

In our research, collecting data from sensors form sensors data that will be forwarded to IoT gateway locally using communication technology. Data encryption will be performed on IoT gateway using GSW encryption scheme and it will be forwarded to cloud. So, all data stored in cloud will be encrypted format. Authenticated user can fire encrypted query to cloud, once query received, cloud performs computation based on query on encrypted data. Then, encrypted results are forwarded to authenticated user. Once user receives result from cloud, it performs decryption by using his/her secret key.

GSW Algorithm

Ciphertext will be derived from plaintext in matrix fashion. Addition and multiplication will be performed on ciphertexts. Consider Plaintext space $M = X$q.

Consider following steps in GSW:

1. Key generation (public key and private key)
2. Encryption
3. Decryption
4. Perform multiplication on ciphertext
5. Perform addition on ciphertext.

Key Generation (1^n)

Select integer $h = h(n)$, modulus $q = q(n)$ and $\lambda = [\log q]$ and $g = h.\lambda$, secret vector is $\boldsymbol{v} \leftarrow \boldsymbol{\alpha x}_{\boldsymbol{q}}^{\boldsymbol{h}-1}$ generates secret key Sk = (S||1).

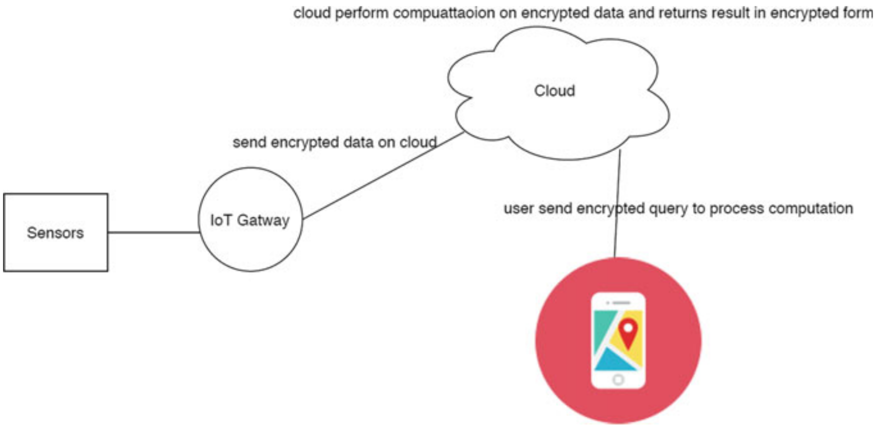

cloud perform compuattaoion on encrypted data and returns result in encrypted form
Cloud
send encrypted data on cloud
Sensors
IoT Gatway
user send encrypted query to process computation

Fig. 1 Proposed architecture

Generates public key $p = (h, q)$.

Encryption (Sk, ∈ M) generates ciphetext

Encryption process takes input as secret key and message σ, random matrix $D \leftarrow \boldsymbol{\alpha x_q^{(h-1)*j}}$ and error vector err $\leftarrow Z^j$ for error distribution.

Generates

$$\text{Ciphertext} = \begin{pmatrix} -D \\ S^{iA+\mathrm{err}^i} \end{pmatrix} + \alpha.B \in X_q^{h*j} \tag{1}$$

where B is block diagonal matrix.

Decryption (Sk, ciphertext) → $^{\mathbf{I}}$

Here, input is ciphertext and Sk is secret key.

$$\mathrm{Sk}^i * \text{Ciphertext} = \bar{S}\begin{pmatrix} -D \\ S^{iD+\mathrm{err}^i} \end{pmatrix} + \sigma.B = \mathrm{err}^i + \sigma.S^i \tag{2}$$

Homomorphic Addition

Performing addition operation on ciphertext.

1. Ciphertext G1 for plaintext $\sigma 1$.
2. Ciphertext G2 for plaintext $\sigma 2$.

$$\mathrm{G3} = \mathrm{G1} + \mathrm{G2} \tag{3}$$

$$s^i\mathrm{G3} = S^i(\mathrm{G1} + \mathrm{G2}) = \mathrm{err}_1^i + \mathrm{err}_2^i + (\sigma 1 + \sigma 2)S^i\ .\ B \tag{4}$$

Homomorphic Multiplication

$$\mathrm{G1} = \begin{pmatrix} -D1 \\ S^{iD1+\mathrm{err}1^i} \end{pmatrix} + \sigma 1.B \tag{5}$$

$$\mathrm{G2} = \begin{pmatrix} -D2 \\ S^{iD2+\mathrm{err}2^i} \end{pmatrix} + \sigma 2.B \tag{6}$$

$$\mathrm{G} = \mathrm{G1.B}^{-1}.\mathrm{G2} \tag{7}$$

$$\mathrm{G1}.B^{-1}(\mathrm{G2}) = \begin{pmatrix} -D^* \\ S^{iD^*+\mathrm{err}1^{*i}} \end{pmatrix} + \sigma 1.\sigma 2.B \tag{8}$$

In our research work, we had implemented GSW algorithm for addition and collected data sets from IoT devices. Sensors sense data and forward it to IoT gateways. Gateways send data to cloud after successful authentication mechanism. Cloud servers maintain encrypted data, if users send query to cloud, our cloud server performs computation on encrypted data collected from IoT devices. After computation, data is forwarded to user. User decrypts that data and takes decision.

4 Experimental Results

A homomorphic encryption is the transformation of data into ciphertext that can be examined and functioned on encrypted form. Homomorphic encryption shows a significant part in cloud computing, permitting patients to store encoded PHR files in a public cloud and take advantage of the cloud provider's analytic services. The scheme prevents rogue insiders from violating privacy and avoids unintentional leakage of private data. Homomorphic encryption schemes are utilized to achieve processes on encoded information without knowing the private key (i.e., without decryption), and the client is the only holder of the secret key. Once the result of the process is decrypted, it is the equivalent as it had carried out the calculation on the raw data.

In this section, we will discuss results of homomorphic encryption on data set and collected from IoT devices and performed GSW encryption on it (Table 1; Fig. 2).

If we increase security parameter, there will be exponential increase in time for performing operation.

Algorithm provides same result as of plaintext computation and ciphertext computation. If performing computation on plaintext $p1 + p2 = p3$, the enc $(p1) +$ enc $(p2) = p3$. If result is successful for all range of numbers, then it is noise. If noise level is increased, then there is chances to of getting wrong computation result.

Table 1 Time taken for encryption and decryption based on record size (in second)

Record size	Encryption time (first field)	Encryption time (second field)	Decryption time (addition)
50	2.40	2.43	2.46
100	111.97	111.98	112.00
200	233.42	233.43	233.45
300	382.32	382.33	382.37
500	648.47	648.49	648.50

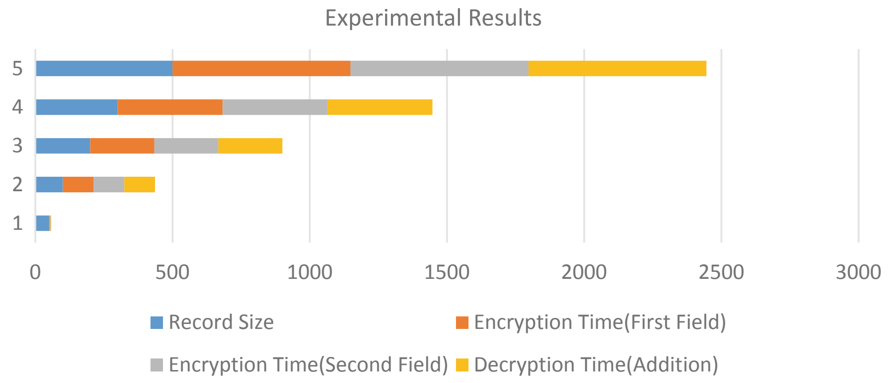


Fig. 2 computation time based on record size

5 Conclusion

In this research, we had collected data from IoT devices and after encryption uploaded it on to cloud server, user sends query to cloud for performing operation on encrypted data. Computation will be performed once query will be fired on cloud. Here, we are using only addition and multiplication operation. Computation time varies based on data set and security parameters. If noise parameter will be more, operation performed on ciphertext will give incorrect result. In future, focus will be on performing multiplication and comparative operations on ciphertext.

References

1. Abbas A, Asku H (2018) A survey on homomorphic encryption schemes: theory and implementation. ACM Comput Surv 51(4), Article 79
2. Çetin GS, Chen H, Laine K, Lauter K, Rindal P, Xia Y (2017) Private queries on encrypted genomic data. BMC Med Genom 10(2):45
3. Wang B, Zhan Y, Zhang Z (2018) Cryptanalysis of a symmetric fully homomorphic encryption scheme. IEEE Trans Inf Forensics Secur 13(6):1460–1467
4. Cousins DB, Rohloff K, Sumorok D (2017) Designing an FPGA-accelerated homomorphic encryption co-processor. IEEE Trans Emerg Topics Comput 5(2):193–206
5. Wu DN, Gan QQ, Wang XM (2018) Verifiable public key encryption with keyword search based on homomorphic encryption in multi-user setting. IEEE Access 6:42445–42453
6. Ayantika C, Indranil S (2015) Translating algorithm to handle fully homomorphic encrypted data on cloud. IEEE Trans Cloud Comput
7. Jung Hee C, Jinsu K (2015) A hybrid scheme of public key encryption & some what homomorphic encryption. IEEE transaction on information forensics and security
8. Reda B, Gouenou C (2017) Proxy re-encryption based on homomorphic encryption. ACM
9. Ding W, Yan Z, Deng RH (2017) Encrypted data processing with homomorphic re-encryption. Inf Sci 409:35–55
10. Alhassan K, Gulak G (2015) SHIELD: scalable homomorphic implementation for encrypted data classifiers. IEEE transactions on computers

11. Alhassan K, Gulak G (2017) SecureMed: secure medical computation using GPU—accelerated homomorphic encryption scheme. IEEE J Biomed Health Inf
12. Ara A, Al-Rodhaan M, Tian Y, Al-Dhelaan A (2017) A secure privacy-preserving data aggregation scheme based on bilinear Elgamal cryptosystem for remote health monitoring systems. IEEE Access 5:12601–12617
13. Dai W, Doröz Y, Polyakov Y, Rohloff K, Sajjadpour H, Savaş E, Sunar B (2018) Implementation and evaluation of a lattice-based key-policy ABE scheme. IEEE Trans Inf Forens Secur 13(5):1169–1184
14. Bocu R, Costache C (2018) A homomorphic encryption-based system for securely managing personal health metrics data. IBM J Res Dev 62(1):1
15. Kocabas O, Soyata T, Aktas MK (2016) Emerging security mechanisms for medical cyber physical systems. IEEE/ACM Trans Comput Biol Bioinf 13(3)
16. Sergiu C, Gianpiero C (2016) Practical privacy preserving medical diagnosis using homomorphic encryption. In: IEEE 9th international conference on cloud computing (2016)
17. Ovunc K, Tolga S (2015) Utilising homomorphic encryption to implement secure and private medical cloud computing. In: IEEE international conference on cloud computing

Secure Data Aggregation and Data Transmission Using HMAC Protocol in Cluster Base Wireless Sensor Network

Sunny Sall and Rajesh Bansode

Abstract Wireless sensor network (WSN) is facing many issues due to implementation in vulnerable environments. Various researchers have already defined systems for data transmission with WSN, still such system has high data loss issues, maximum packet delay as well as packet overhead during the data transmission. Different parameters have been considered to improve such losses like cluster network generation, data aggregation, secure data encryption, and data transmission using distribution approach that produces effective outcome and eliminates such issues. Moreover implementation with cluster network and selection of a Cluster Head (CH) based on trust, which produces much effective results and provides flexibility to system. In this research, we propose a secure data transmission in cluster network and investigate the Quality of Service (QoS) parameters using various experimental analysis. Initially, we create different clusters with collaboration of multisensor nodes while each node consists of individual battery power as energy. We calculate the trust of each node and define a CH based on the highest energy, Data Aggregation (DA) and Broadcast Tree Construction (BTC) are the two different techniques that have been used to eliminate network lifetime or cut generation in the network during data transmission. In partial experiment, analysis system shows improved QoS parameters like throughput, delay, packet overhead, etc., respectively. It also enhances the network life because of proposed energy conservation approach.

Keywords Cluster network · WSN · Data aggregation · Broadcast tree construction · QoS

S. Sall (✉)
Department of Computer Engineering, St. John College of Engineering and Management, Palghar, Maharashtra, India
e-mail: sunny_sall@yahoo.co.in

R. Bansode
Department of Information Technology, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India
e-mail: rajesh.bansode@thakureducation.org

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_21

1 Introduction

In this era, the dagger development of mobile computing expedients that primarily embody laptops, personal digital assistants (PDAs), as well as hand-held digital devices, is driven by a revolutionary change in the computing world. Security approaches to eliminate different network attacks like wormhole attack [1] in untrusted environment are to identify intruder or threats from large traffic defined in [2]. Computing will not simply place trust within the power provided by private computers, and the idea of gift computing often arises and becomes one within the applied science society at every research hotspots [3]. Throughout this environment, a path behalf of the two hosts might contain steps through one or extra nodes inside the painter. An important drawback in associate degree passing mobile ad hoc network is finding and maintaining routes since host quality can cause topology changes [4]. Several routing algorithms for MANETs are proposed inside the writings which they differ inside the painter. New routes square measure set up and existing ones square measure are modified. Basically, MANET networks' square portion is extra susceptible to suffer from the hateful performances. In this paper, we proposed wormhole attack detection and prevention approach using secure mechanism of detection of such malicious behavior. In this paper, we proposed various attack detection and prevention approaches using secure mechanism of detection in malicious behavior [5] and investigated the proposed experimental analysis.

Figure 1 shows basic WSN view in sensor network, which is basically similar to wireless ad hoc networks in the intelligence that they rely on wireless connectivity and unstructured establishment of networks so that sensor data can be ecstatic wirelessly.

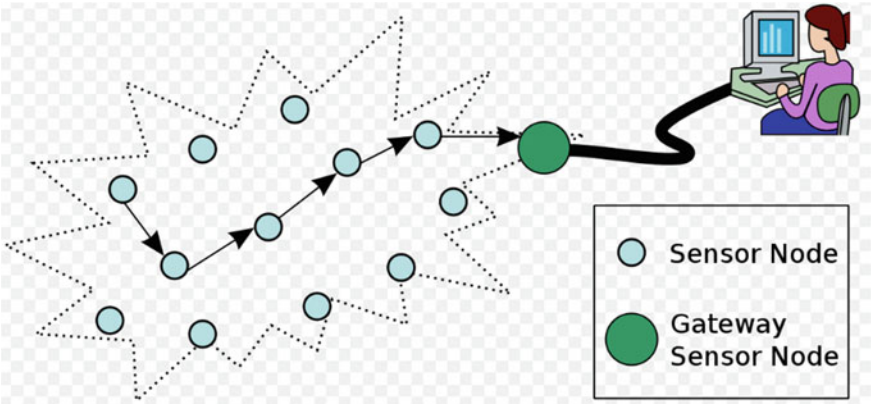


Fig. 1 Representation of a wireless sensor network [6]

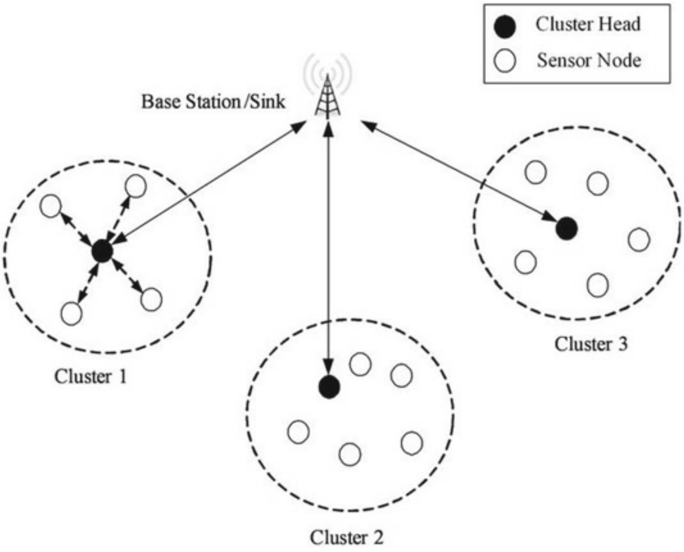


Fig. 2 Proposed system cluster scenario

2 Related Work

Gohil et al. [1] proposed a true connection that verifies whether there is an immediate use of truelink to connect a node to a neighbor next door. The main drawback is that truelink only operates on IEEE 802.11 systems which are backward compatible with a computer code upgrade. A trip time technique (RTT) is evolving to overcome more device optimization issues. The RTT is that the time taken to send RREQ to an endpoint supply node and to receive RREP. A node is expected to calculate the RTT between its adjacent nodes and itself. Throughout RTT, the malicious nodes are rated higher than other nodes. This approach should assess the output of its actual and misconducting neighbors. This detection technique is efficient only in the case of hidden attacks.

Upadhyay and Shukla [2] describe numerous methods that are planned to be used to identify wormhole attacks using a leash method for the packets. Packet leash is the technique that guards against wormhole attack. The leashes are either spatially or temporally categorized. All nodes inside the network should have data from their own position and stable synchronized clock in spatial leashes. This requires its own current location when a sender sends the data packet and UTC accepts wormhole nodes. Directional data are exchanged between source and destination during this technique. The method designed by destination node is also able to find the wormhole, the received signals from malicious nodes gives directional information of the attacker. The relation to the wormhole is observed if the supply signals and intermediate nodes are distinct.

Sharma et al. [3] describes, detection of wormhole attack in trustworthy communication network, this work also carried out a secure communication mechanism between source and sink node. Different calculations have been done to evaluate if the node belongs from trustworthy communication. Another such assault is a wormhole attack, in which two or many nodes are often flustered together to gain knowledge calculation and communication. A wormhole compromised network is illustrated during this post, and the research model can be used for secure communication within the network being targeted. The network model is designed to identify the secure communication node, so that connectivity is maximized. Additionally, in most mobile network-based attack, the writers address a way of generating the safe path. The model eventually provided the optimized parameter for reconciliation communication. Results show the improved performance of model in terms of the communication throughput and reduced the loss.

Zamani and Soltanaghaei [4] propose identification of malicious node into the network, the detection has been done based on various confidence values of respective nodes. This approach is similar to Internal Intrusion Detection System (IIDS) [3]. Authors propose a way to defend against victimizing wormhole attack combination of parameters such as energy, connection range, and node buffer length. A node's confidence value is calculated backed by these parameters. This node confidence value is then compared with the threshold value of network faith. This comparison was supported and one can find out if the selected node is either malicious or valid or is not. The planned methodology consists of two phases: first, the analysis of the network parameter and the calculation of the threshold, and second, the implementation of the protection on the routing protocol that prevails. Performance evaluation has been done for the system using AODV protocol and calculated the system efficiency with similar protocols like DSDV, DSR, SAODV etc. The performance of the proposed confidence-based approach, based primarily on defense, is cost-effective and reasonable against comparative findings wormhole attacks in MANET.

Kumar and Shekhawat [5] has done different techniques for detection of wormhole attack as well as prevention from similar attacks in large network. A projected defensive A system of wormhole attacks is based primarily on the hash-based compression function (HCF), which is really mistreatment of any secure hash, chips an RREQ packet hash field, and the intended approach looks terribly optimistic compared to other solutions proposed in literature.

Jain and Verma [7] system describe various safety hazards which affect different levels of safety. Because of its simplicity and self-routing nature, MANET is responsive to various threats, and various attacks can breach separate layers within the network. There are numerous attacks and each attack has its own impact on different layers since some can only influence a single layer of the network while others may reach other layers, i.e., depending on the nature of the intrusion on how it reacts. Intrusion is capable of crippling the communication network entirely all the attacks.

Patel et al. [8] system present AODV with the implementation of the MANET as well as the detection and prevention of wormhole attack (WADP) technique for this revamped AODV, and the routing protocol is modified to detect and prevent real-world wormhole attacks. So that malicious nodes and false positive authentication of limit nodes detected within the network function are eliminated. Simulation tests also indicate that node authentication does not eradicate false positives although it helps to chart the actual location of the hole jointly and could be a relatively double test for hole attack detection. This algorithmic rule does not use any special hardware for police work wormhole attack.

Patel et al. [9] take initiative to eliminate or minimize the impact of a hole attack by offering a solution that might experience a void within the initial level of route setup. This resolution is based on the hop count research method, i.e., the hop count is used as a parameter for characteristic shapes involving tunnel space. Hop counts are evaluated to detect hostile nodes. Simulation of the planned work is completed in many node and traffic situations in the presence of a wormhole attack. In the simulation results, the predicted strategy demonstrates superior performance as PDR and attendance decreases however, in fact, "normal end-to-end gap" rises. It has been noticed within the analyzed state of affairs that the MAODV works superior to AODV. Edited is ideal for wormhole attack identification and bar detection. This strengthens the insecure circumstances of the packet distribution level, with a borderline reduction in turnout and a reasonable raise in end-to-end delay.

System [10] throws light on a weak wormhole attack, during which two or more malicious nodes type a tunnel as a relay mechanism into the packets themselves. Such an attack may result in the packets sent being selectively diverted, fabricated, and changed. During this paper, AN identity is planned on the side of the clusters to shield the network from hole attacks based primarily on the signature theme. The designed theme need not transmit any credentials between nodes, thereby overhead calculations. Cluster during which cluster heads area unit selected in such a way that they cannot be harmful, specifically focused architecture is employed. This style works in three steps. The effects of the simulation indicate the improved performance of the theme in terms of throughput, packet delivery ratio and end-to-end delay.

According to [11], the proposed protocol is based on neighborhood overhearing and frequent analysis of key, and different tables and expected procedure information were found to be safe and a few region assault units are checked on it. Wormhole assault is observed by nodes being overheard. The results show that M-AODV has been strengthened in terms of the packet transmission magnitude relationship and the latency has therefore been lowered further, however the overhead amount has been raised. M-AODV jointly improves network reliability and stability. Thus, the planned protocol is proven to be protected in simulations alongside wormhole and blackhole attacks.

Hu et al. [15] fix problems with WSN because every node is created to handle all the information from a forward-looking and network management-friendly, autonomous system to all levels, from the appliance layer to the appliance layer. This works very well, especially thanks to the algorithms developed with the short-shortcut WSN, they do not create simplicity and completeness when trying to implement widely and widely distributed different and low-energy WSNs at the same time.

Gante et al. [16], for example, announce good organization of SDWSNs to boost potency and overcome the simple task fundamental to WSNs. The control design for WSNs with allied integrated controllers draws on the idea of a base station. The controller sets transmission rules for area units stored in flow tables from role knowledge gained through the application-layer context methodology of the design.

Olivier et al. [17] projected gradable design referred to as software package outlined clustered detector. It is believed that multiple base stations networks (SDCSN) are used as controllers which additionally play the role of cluster head. Clusters of large nodes are divided into clusters and each one has the head of a cluster. The detector nodes in each cluster are monitored and organized by the cluster head and the information collected in each cluster is passed to the cluster head.

De Oliveira et al. [18] proposes an approach to design a multiple controllers in Software Defined Network (SDN) called as Tiny-SDN. This approach can perform balancing of heavy data traffic or network overhead in large cluster nodes. This solves issues like in-band scheduling, long contact times, and limited power supplies.

Tootoonchian et al. OpenTM [19] proposed a technique on implementation of openFlow protocol in SDN-based WSNs, and demonstrates the performance evaluation using various QoS parameters. Freshly but, work has been conferred on a network mensuration design supported SDN for observation of WSN data like routing path per-packet, the ratio of each link and hence the delay in each ointment-hop configuration.

According to Jain and Paul [20] illustrates state of the art node virtualization using SDN in cloud environment, different node virtualization mechanisms have been demonstrated, how it works when Virtual Machine (VM) has been hitted. It is vital to say here that SDN is not concerning SDNs have been shown To be able to change network management and encourage creativity by network programmability rather than network performance enhancement.

This approach demonstrates the application framework of SDN approach in Wireless Sensor and Actor Network (WSAN), the basic objective of system improves the efficiency and scalability of system. The similar functions have been implemented like SDN to improve the QoS parameters. A number of solutions and process steps have been provided by Zhou et al. to stop the problems associated with the SDN controller nodes and offer WSN efforts [21].

3 Proposed System Design

Figure 2 describes the cluster creation process and transmission between source and destination node. At the end of each TS network, nodes verify sensed data and broadcast messages to nodes within given cluster distance (CD) for cluster creation. Cluster creation uses the relay node (RN) and CD to group the sensor in same cluster. Upon accepting the broadcasted message, each node verifies the value of RN. If its value is within RN, it stores in its memory and compares CD with each node's distance. If the distance between nodes is same or less to CD and sensed value is within given RN, then those group of nodes forms a cluster. The nodes NID which are related will not broadcast message for cluster creation. Nodes which are not participating in cluster creation process are based on *RN* and *CD*.

4 Cluster Head Selection

Input: Cluster set with nodes.

Output: CH selection with remaining sensor node.

Step 1: select all nodes as initial population.

Step 2: Select evaluation set.

Step 3: Apply crossover on similar power nodes.

Step 4: Apply mutation on each sensor node.

Step 5: Apply fitness on all nodes power.

Step 6: select best node using route let wheel selection.

Step 7: Check GA evaluations.

Step 8: Select final max energy node as CH node.

5 Data Transmission Using BTC

A dynamic cluster will be designed when the goal reaches several cluster boundaries. A difficult role problem is how the device identifies the situation, particularly in a fully distributed environment, when the goal reaches the boundaries. We use boundary nodes to solve this problem in a fully distributed way.

The following assumptions are taken in order to design the proposed algorithm.

In this work, we have carried out an algorithm which is wormhole attack identification that has been through in a cluster-based network creation manner to eliminate the wormhole attacks. Basically, AODV routing protocol is utilized as the fundamental network topology. A multi-layer method is used to determine if a node is participating in a wormhole attack and introduces the layered approach to decrease the capacity of dispensation on respective cluster heads. Due to safety point of view, this also decreases the risk of a cluster head that should be hacked by attacker.

The complete network is separated in clusters sets illustrated in Fig. 1. Sometimes clusters might be corresponded or separate. Every cluster contains a single cluster head as well as number of cluster member nodes. Member nodes forward data only to the cluster head when any nodes want to send data to cluster head (CH). The CH is responsible for forwarding the collective data to all its other cluster members. The CH is selected enthusiastically and preserves the routing information.

```
Input: Primary source node Sender_node, Destination node Dest_node, Group of
nearest nodes Neigh_node [], node id as N_id, node energy N_eng;
Output: From source to destination way based on the given algorithm.
Step 1: initially system select the Sender_node and Dest_node on dynamically
Step 2: select the packet or file f for info broadcast.
Step 3: if (file or data ! =null)
Step 4: read each byte bytes form file or data when reach null
Step 5: send data, initialize cost_filed_1, cost_filed_2, parent_filed_1,
parent_fileld_2
Step 6: while (nd[i] when reach NULL)
        cost_filed_1=node[i]_eng
        parent_filed_1= node[i]_id
        cost_filed_2 =node[i+1].eng
        parent_filed_2= node[i+1] _id
Step 7: if (cost_filed_1> cost_filed_2)
        cost_filed_2=null
        parent_filed_2=null
        Else
        parent_filed_1= parent_filed_2
        cost_filed_1= cost_filed_2;
        parent_filed_2=null
        cost_filed_2=null
Step 8: end of while loop
Step 9: reiteration till when extent at the sink node
```

6 Results and Discussion

After this section, we present experiment analysis using log file, once simulation is finished it will create trace file at background, which contains all information of node communication as well as other log information. We have created a database of five text files which contains reading of 5 ms each till 25 ms as our simulation time is 25 ms. After that, we read the text file in a program created for the trace in Netbeans IDE 8.2. We got readings of various events in Netbeans from which we have plotted the graph of various parameters such as drop rate (DR), throughput, and packet delivery ratio (PDR) calculated according to Eqs. (1), (2) and (3), respectively. The evaluation has been done with various WSNs as well as cluster network existing protocol [12–14] (Tables 1 and 2).

1. **Drop Rate**:
   Drop rate is basically defined as the number of packet lost per number of packet sent. The smallest amount value of drop rate states superior performance of the protocol.

$$\text{Drop rate} = \sum_{i=0}^{n} \left(\frac{\text{packet received } [i \ldots n]}{\text{sent packet } [i \ldots n]} \right) \tag{1}$$

Table 1 Simulation parameters have been used which are described in below table

Parameter	Values
Simulator	NS-allinone 2.35
Simulation time	25 s
Channel type	Wireless channel
Propagation model	Two ray ground
Standard	MAC/802.11
Simulation size	1000 * 1500
Max packet length	1000
Ad hoc routing	AODV
Traffic	CBR

Table 2 Shows the basic difference between the proposed and existing WSN

Parameters	WSN [13, 14]	Proposed (cluster base with AODV)
Data aggregation	No	Yes
Data security	Yes (selective)	Yes
Energy conservation	No	Yes
Packet loss	High	Low
End-to-end delay	High	Low
Packet overhead	High	Low

2. **Throughput**:
   It is basically defined as the total number of packets made available throughout the simulation phase. This is a combination of the total number of packets sent by TCP and the total number of packets submitted. The higher throughput value means rising protocol performance.

$$\text{Throughput} = \left(\frac{\sum_{i=0}^{k} \text{received packet [TCP]}}{\sum_{i=0}^{l} \text{sent packet [TCP]}} \right) \tag{2}$$

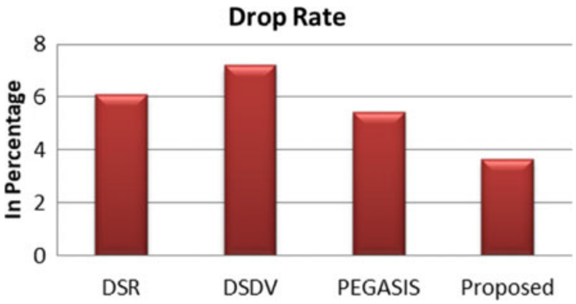
3. **Packet delivery ratio (PDR)**:

The PDR is defined as the ratio of the data packet numbers to the number of packets produced through the network. The higher value of the packet distribution ratios represents greater efficiency of the protocol.

$$\text{PDR} = \sum_{i=0}^{n} \left(\frac{\text{packet received [TCP]}}{\text{sent packet [TCP]}} \right) * 100 \tag{3}$$

This Fig. 3 will provide the drop rate overall simulation during the communication with other protocols. The above graph was calculated based on various experiment

Fig. 3 Drop rate of proposed versus existing



analyses in NS2 environment. All defined protocols have been used with different number of nodes in cluster network. The graph shows minimum packet drop rates of proposed AODV than other protocols.

This Fig. 4 will provide the throughput of system during the communication with other protocols. Throughput is the most vital parameter that measures the QoS of any network. Moreover, in the first experiment .tr files have been utilized to evaluate the throughput for all protocols. Figure 4, depicts proposed approach produces the highest throughput than all three existing protocols.

This Fig. 5 will provide how the system will increase the actual time percentage of simulation due to proposed energy conservation protocol. Network lifetime has been calculated based on the energy conservation protocol, conservation technique is

Fig. 4 Throughput of proposed versus existing

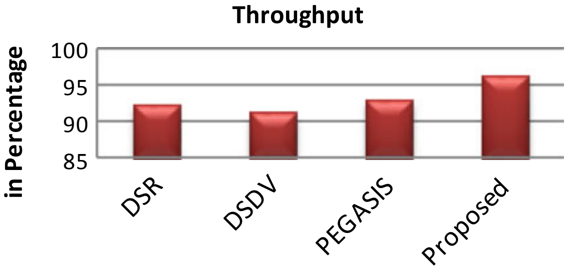
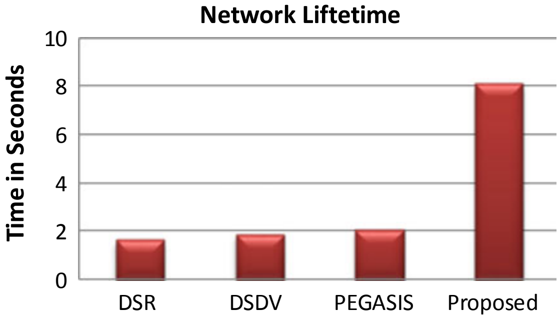


Fig. 5 Network lifetime of proposed versus existing



to eliminates unnecessary energy utilization by sensors during the data transmission by receiver as well as internal nodes respectively. In the proposed system, BTC is used for best path selection as well as energy conservation protocol.

7 Conclusion

This proposes an aggregation of the HMAC and data in WSN. The CHs are initially chosen based on the connectivity of the node, which acts as a data aggregator. The proposed techniques provide the ad hoc network with greater security and also prevent different types of network attacks. It supports in the direction to escalation of Packet Delivery Ratio (PDR), and decrease the packet losses, network overhead with shortest path finding technique using Broadcast Tree Construction (BTC) which cultivates network efficiency. In the future, for faster detection of wormhole nodes, it will also need to modify the table entries in the receiver node and also improve the security of the wireless ad hoc network. By means of set up, such effective approach prevents various kin fog network attacks with new defined algorithm. It also detects and prevents wormhole attack throughout data communication. The proposed techniques offer greater security for various types of ad hoc networks and also deter network assaults. This facilitates the elevation of the packet distribution ratio (PDR) and reduces the overhead network by encouraging the enactment of the respective routing protocol. To get the detection of wormhole nodes faster, the table entries at the receiver node need to be improved in future plan as well. And the security of ad hoc wireless networks is also improved. By applying such an effective approach, the kin fog network will deter various attacks with new defined algorithm.

References

1. Gohil Y, Sakhreliya S, Menaria S (2013) A review on: detection and prevention of wormhole attacks in MANET. Int J Sci Res Publ 3(2)
2. Upadhyay V, Shukla R (2013) An assessment of wormhole attack over mobile ad-hoc network as serious threats. Int J Adv Netw Appl 5(1)
3. Sharma D, Kumar V, Kumar R (2015) Prevention of wormhole attack using identity based signature scheme in MANET. Comput Intel Data Min 2. Volume 411 of the series Advances in intelligent systems and computing, pp 475–485. Springer, 10 Dec 2015
4. Zamani E, Soltanaghaei M (2016) The improved overhearing backup AODV protocol in MANET. J Comput Netw Commun 2016, Article ID 6463157, 8 pp
5. Kumar A, Shekhawat SS (2015) A parameter estimation based model for wormhole preventive route optimization. IJCSMC 4(8):80–85
6. https://en.wikipedia.org/wiki/Wireless_sensor_network
7. Jain AK, Verma R (2015) Trust-based solution for wormhole attacks in mobile ad hoc networks. Glob J Multidiscip Stud 4(12)
8. Patel R, Patel A, Patel N (2015) Defending against wormhole attack in MANET. In: Fifth international conference on communication systems and network technologies. IEEE

9. Patel D, Trivedi P, Potdar MB (2015) A brief analysis on detection and avoidance techniques of wormhole attack in MANET. Int J Comput Appl 117(16)
10. Biswas J, Gupta A, Singh D (2014) WADP: a wormhole attack detection and prevention technique in MANET using modified AODV routing protocol. In: 9th international conference on industrial and information systems (ICIIS), 2014. IEEE
11. Raghuwanshi K, Saxena A, Manoria M (2015) An enhanced integrated solution for identification and elimination of wormhole attack in MANET. Int J Comput Appl 110(7)
12. Uddin MA, Islam MM, Khanom R, Mosaddek M. Efficient and secure cluster based routing protocol for wireless sensor network
13. Ouafaa I, Mustapha E, Salah-Ddine K, Said EH (2016) Performance analysis of SLEACH, LEACH and DSDV protocols for wireless sensor networks (WSN). J Theor Appl Inf Technol 94(2). 2016 Dec 31
14. Upadhyay R, Bhatt UR, Tripathi H (2016) DDOS attack aware DSR routing protocol in WSN. Proc Comput Sci 78(C):68–74
15. Hu F, Hao Q, Bao K (2014) A survey on software-defined network and OpenFlow: from concept to implementation. IEEE Commun Surv Tutor 16:2181–2206
16. Gante AD, Aslan M, Matrawy A (2014) Smart wireless sensor network management based on software-defined networking. In: Proceedings of the 2014 27th Biennial Symposium on Communications (QBSC), Kingston, ON, Canada, 1–4 June 2014, pp 71–75
17. Olivier F, Carlos G, Florent N (2015) SDN based architecture for clustered WSN. In: Proceedings of the 2015 9th international conference on innovative mobile and internet services in ubiquitous computing (IMIS), Blumenau, Brazil, pp 342–347, 8–10 July 2015
18. De Oliveira BT, Margi CB, Gabriel LB (2015) TinySDN: enabling multiple controllers for software-defined wireless sensor networks. IEEE Lat Am Trans 13:3690–3696
19. Tootoonchian A, Ghobadi M, Ganjali Y (2010) OpenTM: traffic matrix estimator for OpenFlow networks. In: Proceedings of the international conference on passive and active network measurement, Zurich, Switzerland, 7–9 April 2010. Springer, Berlin, pp 201–210
20. Jain R, Paul S (2013) Network virtualization and software defined networking for cloud computing: a survey. IEEE Commun Mag 51:24–31
21. Zhou J, Jiang H, Wu J, Wu L, Zhu C, Li W (2016) SDN-based application framework for wireless sensor and actor networks. IEEE Access 4:1583–1594

Performance Evaluation of Adversarial Examples on Deep Neural Network Architectures

Aruna Animish Pavate and Rajesh Bansode

Abstract Tremendous advancement in machine learning over the recent years leads to the use of deep neural networks in many applications from weather prediction to safety critical applications like disease diagnosis. Latest work revealed that, Deep Neural Networks are possibly being attacked using the perturbed input like images/text/audio, also referred to as adversarial examples. Even a small change in considering single pixel can cause neural network models to make mistakes in predicting the output. This has resulted in increased safety concern of deployment of safety critical applications. It is observed that the adversarial examples are transferred from one neural network model to another neural network model with considering adversary knowledge either black box which corresponds to a real-life assumption with the adversary having nearly no knowledge of the framework to be attacked, or white box or gray box. Adversarial examples can be categorized by various characteristics. This paper provides a good inclination of adversarial instances in the context of computer vision with details on various types of adversarial attacks on individual types of neural network architectures and also elaborate the different metrics applied to validate the system performance. We conclude that adversarial learning is a factual threat to application of machine learning not only in the physical world but also during training the model and testing the model. There are some certain counter measures that exist but none of them can act as an individualistic solution for all the challenges. It remains an extensive challenge for the machine learning community to deal with robustness.

Keywords Deep neural network · Adversarial attack · Machine learning · Computer vision

A. A. Pavate (✉) · R. Bansode
Department of Information Technology, Thakur College of Engineering, Mumbai University, Mumbai, India
e-mail: arunaapavate@gmail.com

R. Bansode
e-mail: rajesh.bansode1977@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_22

1 Introduction

Machine learning system gained remarkable success in almost all the applications. Machine learning is used in many applications whose failure could result in loss of life including self driving cars [1, 2], robots [3, 4], etc., as well as in other applications like natural language processing such as caption generation [5], speech recognition [6–9], malware classification [10, 11], anomaly detection [12], and many more. Most machine learning algorithms are considered to map from an input sample domain to an output sample domain by detecting a piece of sample pairs of inputs and outputs from these domains. Lately, machine learning techniques are known to be easily attacked by adversarial attacks. In computer vision algorithms, object detection, object segmentation, and object classification are known to be vulnerable to adversarial attacks, by adding small perturbations to the input. In recent times, the number of approaches have been implemented and examined to produce adversarial examples for different neural networks architectures and updating the parameters of the model to apply optimization techniques to find the optimal solution [13–19]. Szegedy et al. were premier to introduce the concept of adversarial samples, applied over image classification, and misled the neural network with high confidence [13]. Deep neural network is noticed as blackbox as neural network works well with limited knowledge [20, 21].

In this work, we have explored the different attack approaches for creating adversarial examples. Most of the adversarial instances are created across the computer vision domain. This work emphasis on one pixel attack as a one of the important adversarial attacks that demands very few adversarial information to deceive the various network architectures because of the immanent properties of the evolutionary algorithm and generates the low cost adversarial attacks to assess the robustness of the classifier based on Differential Evolution (DE). Using the available framework, implemented an experimental attack on image classification models using CIFAR-10, CIFAR-100 data sets.

1.1 Discussion About Basic Terminologies

In this section, we highlight some major terminologies used during the work that are related to the generation of adversarial samples on deep neural networks in the computer vision field.

Informal Framework

General framework for adding perturbation to the input sample that helps to generate the adversarial examples. Consider a classifier $K(X) = \hat{Y} \varepsilon \{1 \ldots N\}$. Let $L(x)$ be the labels that belong to the class and $M(X)$ be the logits.

$$\arg\ \max\ L_i(X)_{i \varepsilon \{1 \ldots N\}} = \hat{Y} \tag{1}$$

$$\sum_{i=1}^{N} L_i(X) = 1 \tag{2}$$

$$L(X) = \mathrm{Softmax}(M(X)) \tag{3}$$

Adversarial example

Definition An adversarial example is an instance to the model such that perturbed input sample x' addressed at $L(X') \neq L(X)$ and the distance between $\|X' - X\|$ being narrow, i.e., sometimes unnoticeable to human eyes. Figure 1 represents the generation of adversarial samples.

Threat Models

Practically, the main intent of an adversarial instance is to alter the input data sample as small as possible of a classifier in such a way that is not correctly classified. The attacker's knowledge can be considered as 1. Black-box advisory knowledge 2. White-box advisory knowledge 3. Grey-box advisory knowledge.

Black-box advisory knowledge: In this, it is considered as an adversary aware about the input data and observes the result of the network. A typical strategy of the attack is to make use of a proxy model for crafting an adversarial attack and

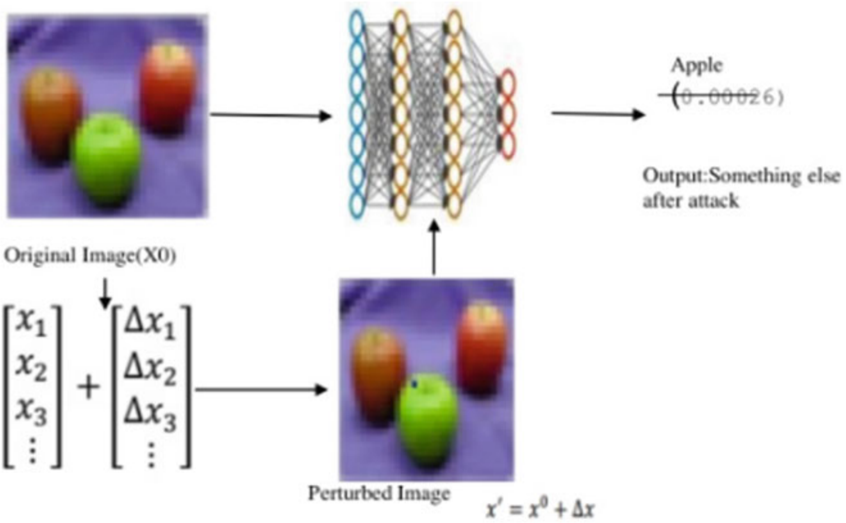

Apple
~~(0.00026)~~
Output:Something else after attack
Original Image(X0)
$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix} + \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \vdots \end{bmatrix}$
Perturbed Image
$x' = x^0 + \Delta x$

Fig. 1 An illustration of adversarial examples developed on Lenet. The predicted label belongs to something else than the actual class after the attack. The prior confidence was 0.00026. The actual class is an Apple. Here, perturbed image x' is the smallest modification of an image by changing a singular pixel. x^0 is an original image

adversarial samples tested on different architectures [22]. Another method is zero-order optimization to generate adversarial samples without accessing model gradients [23]

White-box advisory knowledge: Here, the adversary is aware about the neural network model, network architecture type, number of hidden layers, activation functions, and hyper-parameters [15]. As per the research, maximum of the adversarial attacks are white-box attacks where advisory knows all the details, but these attacks can be converted into black-box models due to the transfer of examples property of the adversarial samples [22].

Grey-box advisory knowledge: Any variation between the white-box advisory knowledge and black-box advisory knowledge is considered as grey-box attack where adversary has a path to access the architecture, training [24].

Attacks Specificity

In most of the machine learning algorithms, the type of threat can be either targeted attack or untargeted attack.

Targeted attack: manipulate deep neural networks to misclassify the input to a specific class. Targeted attacks mostly applied to multiclass classification problems. For example, Here, model misclassified an ‘Apple’ sample to ‘bowl’ class.

Non-targeted Attack: The output of the model can be represented arbitrary, i.e., output does not assign to specific class. These types of attacks are easy to craft and apply. For example, model misclassified an ‘Apple’ to as any other class

Attack Frequency: Attack can be one time or iterative. One time attack is generated only a single time whereas iterative attack can take many attempts to generate adversarial examples. For some rigorous algorithms (reinforcement learning), one time attacks are more suitable. Iterative attacks need many queries to be posted to succeed the attack, and so the computational cost is more to generate them as compared to single time attack.

Attack Measures: There are many measures useful for evaluating the performance in terms of efficiency and success of the proposed system. The successive measures utilized here to evaluate system effectiveness are as follows:

1. *Confidence*: This measure specifies the intermediate probability of the labels of the expected class output to the successfully altered class label of an input sample from expected to target.
2. *Success Rate*: This measure specifies that the observed probability of the original image that can be successfully changed to another targeted attack and non-targeted attack by modifying the original image
3. *Average Distortion*: Average distortion for a single pixel is computed by considering average modification of the channels (R, G, B) to evaluate the cost of the attack. If the value of the average distortion is eminent, such that it can be easily visible to the human eyes then the cost is high.

2 Related Work

This section highlights different approaches for generating adversarial attacks. In the recent studies, there are many approaches in the below described defeated by countermeasures. The presence of these approaches needs to be studied which improves the transferability and thus robustness of the neural network models.

Szegedy et al. proposed box-constrained L-BFGS adversarial examples. Author discovered that mapping of input to output in deep neural networks interrupted to some extent. It is possible to cause the network to present an image belonging to another class by making certain perceptible perturbation which can be possible by exaggerating the network's prediction error. Also it is possible that the same perturbation can cause other different networks, trained using different dataset to analyze the same image belonging to a different class. The proposed concept is applied on different networks like AlexNet, QuocNet, FC [15].

The Fast Gradient Sign Method (FGSM) [16] is an efficient solution tried for both targeted and non-targeted label attack. Fast gradient sign method computes the ramp cost function corresponding to the input image of the deep neural networks.

Moosavi et al. proposed an accurate and simple method which computes the interval by adding small perturbation to the image sample to provide a robust classifier. Deepfool computes the nearest interval against the input image sample to the outcome boundary of the perturbed input. The attack is applied iteratively to avoid the problem of nonlinearity in high-dimensional space. Deepfool proved on both binary and multiclass classifiers [17].

Papernot et al. presented Jacobian-Based Saliency Map Attack (JSMA) that is estimated as an expensive method as it runs much slower than the fast gradient sign method where Jacobian matrix computed for the input sample. The adversarial success rate, 97% achieved by applying small perturbation of 4.02% input features per sample [18].

Tom B. Brown et al. introduced a targeted universal attack using image patch and tried over the real-world problem without knowing the knowledge of the type of classifier, scene-independent, angle of the camera, etc. Trained patch retrieved using Expectation over Transformation (EOT) framework [25] where the prediction is applied through locations, arbitrary images, and transformed images. The attack is evaluated across different Imagenet models like resnet50, inceptionv3, xception, VGG16, and VGG19. Here, attack is examined and shown that small perturbations are not sufficient, large, local can also be possible to break classifiers confidence 99% [26].

Universal perturbations were developed in [27] for targeted attacks by changing each pixel in the input image with high probability but the attacks were not tried in the physical world. This attack is an iterative attack that presents observations on the geometry of the decision boundaries of neural networks.

Chen et al. [28] proposed an elastic Net attack by introducing set of adversarial examples with small L1 bias and proved the similar performance like the other advanced methods in various attack frameworks like FGM, C & W attack, IFGM. The developed attack enhanced transferability of attack and the capacity of adversarial training for the deep neural network.

Brendel et al. [29] proposed real-world stable attack by considering adversary knowledge as blackbox. The implementation is based on boundary between adversarial and non-adversarial samples, and needs final class prediction. This attack is simple, does not require a substitute model, flexible in terms of adversarial criteria, and requires limited hyper parameter tuning. Author claimed that the decision boundary attack is more robust than gradient-related attacks, score-related attacks, and transfer-related attacks.

Nicholas et al. proved that providing defensive distillation does not improve the durability of the classifier by developing new attacks for L_0, L_2 and L_∞ which are much more powerful than the previous one. These attacks are successful with 100% transfer success rate between the secured model and unsecured model. These attacks were evaluated on trinity different datasets: handwritten digit recognition (MNIST), CIFAR-10, and digit recognition task [30].

3 Methodology

In computer vision research, adversarial examples are original images that are input to the deep neural network that have been specially crafted to missclassify the input by the model. It is proved that deep neural networks can be easily misled by an attacker by simply modifying the color of a single pixel [31] and this approach is referred to as “one pixel attack”. In many cases, attackers can even cause the network model to generate the output of what they want, i.e., targeted attack. We demonstrate this work into three different directions (1) Appling one pixel attack on CIFAR 100 dataset (2) Generation of the low-cost adversarial attacks to measure the robustness of the classifier based on differential evolution (DE) algorithm (3) Analysis of one pixel attack on different network models. Our implementation is based on [31] uses basic convolutional neural network as shown in Fig. 2.

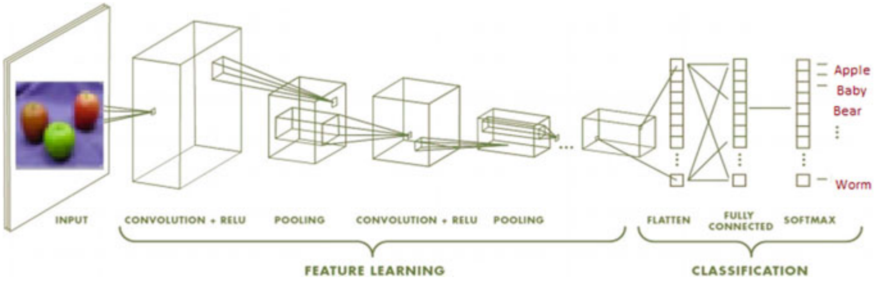


Fig. 2 Basic framework of convolutional neural network applied on [CIFAR 100 dataset [32]. The result of the convolutional neural network belongs to one of the class labels among 100 labels

One Pixel Attack

Su et al. [31] proposed the approach that causes neural networks to misclassify an input sample by just crafting a single pixel in the input sample. The attack is developed using genetic algorithms without crossover operation and is referred to as differential evolution (DE). The attack is generated for both targeted and untargeted attacks, and image specific which perturbs a single pixel to make a classifier to generate false results. Our implementation is based on one pixel attack on convolutional neural networks with different architectures.

Algorithm 1: One_Pixel in Pseudocode

```
procedure perturb_image( (x,y,R,G,B), img):
  for x,img in zip((x,y,R,G,B), imgs): do
perturbation pixels: pixels = np.split(x, len(x) // 5)
    for pixel in pixels:do
    At each pixel's x,y position, assign its rgb value:
     x_pos,  y_pos, *rgb = pixel
     img[x_pos, y_pos] = rgb
     return imgs
```

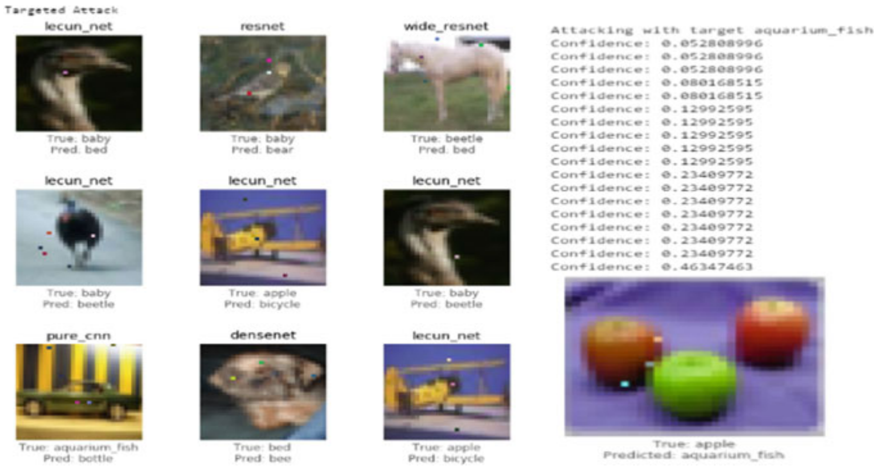
Input samples are encoded into vectors of five elements (x, y, R, G, B). Here, x, y elements indicate the position of an image and the next three values represent the channel for the image. Here, in this case, R, G, B, i.e., color value. Attacks are developed using differential evolution.

Differential evolution:

To find out the pixel which helps in providing a successful attack, the problem is considered as an optimization problem. Here, the problem is either minimization of confidence assuming non-targeted attack and maximization problem in case of targeted attack. Differential evolution is a type of stochastic algorithm useful for global optimization problems.

Algorithm 1: Differential Evolution

```
Initialize population with random positions in search space: popsize = 400
Until termination: adequate fitness reached do
For each solution in the population do
$Xs = x0 + 0.5(x1 - x2)$
Calculate fitness of each child if fitness [Parent] < fitness[child]:do
Replace: fitness [Parents] = fitness[child]
Repeat 100x, reating the updated solution as the parent solution
best = trial_denorm
Yield best, fitness [best_idx]: Best fitness is the “winner”
```



```
Attacking with target aquarium_fish
Confidence: 0.052808996
Confidence: 0.052808996
Confidence: 0.052808996
Confidence: 0.080168515
Confidence: 0.080168515
Confidence: 0.12992595
Confidence: 0.12992595
Confidence: 0.12992595
Confidence: 0.12992595
Confidence: 0.12992595
Confidence: 0.23409772
Confidence: 0.23409772
Confidence: 0.23409772
Confidence: 0.23409772
Confidence: 0.23409772
Confidence: 0.23409772
Confidence: 0.23409772
Confidence: 0.46347463
```

Fig. 3 Targeted Attack: One pixel attack implemented on trained on CIFAR-100 different types of DNNs: The right-side image represents targeted attack for single sample with confidence 0.4634. The predicted output is aquarium_fish whereas the image belongs to Apple class

Targeted Attack: An intense fact of the targeted attack is to step down the confidence of the correct class and maximize the average of the probabilities of the other classes. This type of attack is very hard to implement as it constrains the classifier to predict the wrong class than that of the actual one. There is no guarantee of success of attack; after several iterations, the confidence of the classifier increases. Figure 3: shows that Letnet model to classify an image of an apple to an image of aquarium_fish

Untargeted Attack

An expressive objective of the untargeted attack is to categorize an input sample to as given target class. To design a successful non-targeted attack, the input sample is perturbed as to maximize the probability of the selected class. For CIFAR 10 and CIFAR100, we evaluated randomly drawn 100 samples from the validation set as shown in Figs. 4 and 5.

4 Result and Analysis

We demonstrate the effectiveness of one pixel attack in collation with an original CIFAR-10 test dataset [31]. We have used 100 arbitrary images as input samples for evaluating results not belonging to a specific class(non-targeted attack) and 100 random images for selected labels (targeted attack). It has been observed that the target DNNs have higher accuracy and confidence. All settings are kept the same as in the implementation of attack on CIFAR-10 dataset. Attack on different models conveyed in Table 1 for both the datasets. The limitation of this attack is it works finest

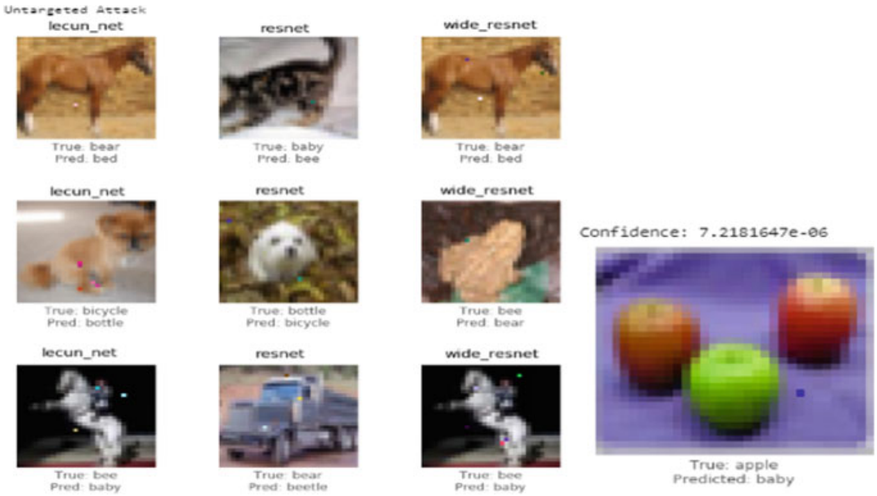


Fig. 4 Untargeted attack: One pixel attack implemented on trained on CIFAR-100 different types of models of DNNs: The right image indicates non-targeted attack for single sample with confidence 7.21. True class is an apple whereas the predicted class belongs to baby

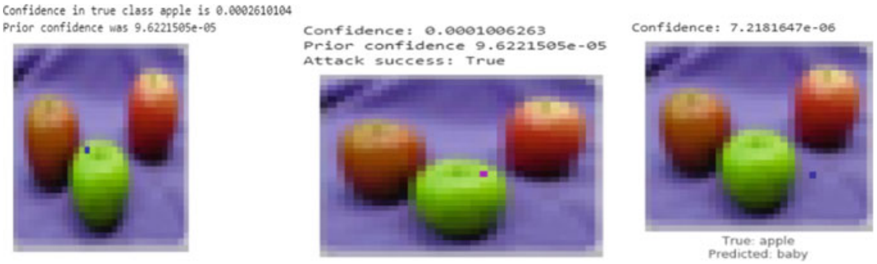

Confidence in true class apple is 0.0002610104
Prior confidence was 9.6221505e-05
Confidence: 0.0001006263
Prior confidence 9.6221505e-05
Attack success: True
Confidence: 7.2181647e-06
True: apple
Predicted: baby

Fig. 5 Shows confidence of the model changes after adding small perturbation into the image. The middle image shows the success of attack on a single image for non-targeted attack prior confidence 9.6221505e-05 & the attack successful

on low-resolution images. After implementing one pixel attack on various models on 100 different randomly selected samples, it successfully generates 100 adversarial images and calculated the success ratio. Multiple experiments run by varying the number of pixels (here 1,3, 5) and the results were analyzed. The attack showed on CIFAR 10 works with an average accuracy of 88.08% and on CIFAR 100 with an average accuracy of 91%.

Table 1 Attacks on different CNN models with 1, 3, 5 pixels perturbation

Model	Parameters	Pixels	CIFAR 10 [31]			CIFAR 100		
			Test accuracy	Attack success (untargeted)	Attack success (targeted)	Test accuracy	Attack success (untargeted)	Attack success (targeted)
LeNet	62 K	1	74.9%	63.0	34.4	96%	65.0	40.4
		3		92.0	64.4		92.0	64.4
		5		93.0	64.4		95.0	70.4
Pure CNN	1.4 M	1	88.8%	13.0	6.67	81%	15.0	7.67
		3		58.0	13.3		50.0	18.3
		5		63.0	18.9		68.0	18.9
Network in network	970 K	1	90.8%	34.0	10.0	95%	38.0	19.0
		3		73.0	24.4		77.0	27.4
		5		73.0	31.1		77.0	34.1
ResNet	470 K	1	92.3%	34.0	14.4	100%	39.0	34.4
		3		79.0	21.1		87.0	25.1
		5		79.0	22.2		78.0	25.2
Dense Net	850 K	1	94.7%	31.0	4.44	95.6%	33.0	4.44
		3		71.0	23.3		75.0.0	34.3
		5		69.0	28.9		76.0	34.9
Wide ResNet	11 M	1	95.3%	19.0	1.11	96.3%	20.0	21.11
		3		58.0	18.9		45.0	28.9
		5		65	22.2		55.3	23.2
CapsNet	12 M	1	79.8%	19.0	0.00	78.3%	22.0	3.00
		3		39.0	4.44		19.0	4.44

(continued)

Table 1 (continued)

Model	Parameters	Pixels	CIFAR 10 [31]			CIFAR 100		
			Test accuracy	Attack success (untargeted)	Attack success (targeted)	Test accuracy	Attack success (untargeted)	Attack success (targeted)
		5		36.0	4.44		35.0	5.44

5 Conclusion

The final purpose of this research is to propose an attack over the convolutional neural network. The applied attack shows that CNN is sensitive to adversarial attacks generated using stochastic differential evolution algorithms of optimization problems. It has been observed that the convolutional neural network is susceptible to low cost and low dimension attack. The characteristics of neural networks have a major impact on the attack both in terms of the average distortion and success rate. The accuracies of the CNN models are not robustly related with the attempt of implementing a successful attack on an input sample. Perhaps, it is most powerful to the type of attacks. It has been observed that the CapsNet has the minimum attack success rate as compared to all other models both in case of CIFAR10 and CIFAR 100 dataset however the model is still vulnerable to attacks.

References

1. Bojarski M, Del Testa D, Dworakowski D, Firner B, Flepp B, Goyal P, Jackel LD, Monfort M, Muller U, Zhang J, et al (2016) End to end learning for self-driving cars. Arxiv Preprint: Arxiv:1604.07316
2. Bourzac K (2016) Bringing big neural networks to self-driving Cars, Smartphones, and Drones. https://Spectrum.Ieee.Org/Computing/Embedded-systems/Bringing-big-neural-networks-to-selfdriving-cars-smartphones-and-drones
3. Janglova D (2005) Neural networks in mobile robot motion. Cutting Edge Robotics 1(1):243
4. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533
5. Andor D, Alberti C, Weiss D, Severyn A, Presta A, Ganchev K, Petrov S, Collins M (2016) Globally normalized transition-based neural networks. Arxiv Preprint: Arxiv:1603.06042
6. Graves A, Mohamed Ar, Hinton G (2013) Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (2013), IEEE, pp 6645–6649
7. Hinton G, Deng L, Yu D, Dahl G, Rahman Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition. Signal Proces Mag
8. CarliniN,Mishra P,Vaidya T, Zhang Y, Sherr M, Shields C, Wagner D, Zhou W (2016) Hidden voice commands. In: Usenix security symposium, pp 513–530
9. Zhang G, Yan C, Ji X, Zhang T, Zhang T, Xu W (2017) Dolphinatack: inaudible voice commands. Arxiv Preprint: Arxiv:1708.09537
10. Dahl GE, Stokes JW, Deng L, Yu D (2013) Large-scale malware classification using random projections and neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing (2013), IEEE, pp. 3422–3426
11. Yuan Z, Lu Y, Wang Z, Xue Y (2014) Droid-sec: deep learning in android malware detection. ACM SIGCOMM Comput Commun Rev 44: 371–372
12. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv (Csur) 41(3):15
13. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R (2013) Intriguing properties of neural networks. Arxiv Preprint: Arxiv:1708.09537

14. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R (2013) Intriguing properties of neural networks. Arxiv Preprint: Arxiv:1312.6199
15. Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. Arxiv Preprint: Arxiv:1412.6572
16. Kurakin A, Goodfellow I, Bengio S (2016) Adversarial examples. In: The physical world. ICLR'2017 Workshop
17. Moosavi-dezfooli S, Fawzi A, Deepfool PF (2016) A simple and accurate method to fool deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2574–2582
18. Papernot N, Mcdaniel P, Jha S, Fredrikson M, Celik ZB, Swami A (2016) The limitations of deep learning in adversarial settings. In: The 1st IEEE European symposium on security and privacy (Euros&P), Saarbrucken, Germany, pp 372–387
19. Madry A, Makelov A, Schmidt L, Tsipras D, Vladu A (2017) Towards deep learning models resistant to adversarial attacks. Arxiv Preprint: Arxiv:1706.06083
20. Knight W (2017) The dark secret at the heart of AI. MIT Technology Review
21. Castelvecchi D (2016) Can we open the black box of AI? Nat News 538(7623):20
22. Zhao Z, Dua D, Singh S (2017) Generating natural adversarial examples. Arxiv Preprint: Arxiv:1710.11342
23. Chen PY, Zhang H, Sharma Y, Yi J, Hsieh C-J (2017) Zoo: zeroth order optimization based black-box attacks to deep neural networks without training substitute models. Arxiv Preprint: Arxiv:1708.03999
24. Vivek BS, Mopuri KR, Venkatesh Babu R (2018) Gray-box adversarial training, 6 Aug 2018. Arxiv Preprint: Arxiv:1808.01753v1 [Cs.Cv]
25. Athalye A, Engstrom L, Ilyas A, Kwok K (2017) Synthesizing robust adversarial examples. Arxiv Preprint: Arxiv:1707.07397
26. Brown TB (2018) Dandelion Mané, Aurko Roy, Martín Abadi, Justin Gilmer, Adversarial Patch, 17 May 2018. Arxiv Preprint: Arxiv:1707.07397 [Cs.Cv]
27. Moosavi-dezfooli S-M, Fawzi A, Fawzi O, Frossard P (2016) Universal adversarial perturbations. Arxiv Preprint: Arxiv:1707.07397
28. Chen PY, Sharma Y, Zhang H, Yi J, Hsieh C-J (2018) EAD: Elastic-net attacks to deep neural networks via adversarial examples, 10 Feb 2018. Arxiv:1709.04114v3 [Stat.Ml]
29. Brendel W, Rauber J, Bethge M (2018) decision-based adversarial attacks: reliable attacks against black-box machine learning models, 16 Feb 2018. Arxiv Preprint: Arxiv:1712.04248v2 [Stat.Ml]
30. Carlini N, Wagner D (2017) Towards evaluating the robustness of neural networks. University Of California, Berkeley, 22 Mar 2017. Arxiv Preprint: Arxiv:1608.04644v2 [Cs.Cr]
31. Su J, Vargas DV,Kouichi S (2017) One pixel attack for fooling deep neural networks. Arxiv Preprint: Arxiv:1710.08864
32. https://www.cs.toronto.edu/~kriz/cifar-100-python.tar.gz. Last accessed 2 Nov 2020

Study on Analysis of Gene Expression Dataset and Identification of Differentially Expressed Genes

Saroj Ghadle, Rakesh Tripathi, Sanjay Kumar, and Vandana Munde

Abstract The analysis on the dataset of complex genetic diseases such as cancer, cardiovascular, respiratory diseases, AIDS in Homo sapiens at early stage of diagnosis is important. It is a challenging task in today's biomedical and other research area domains. In this paper, the analysis of gene expression dataset is discussed. The gene expression datasets are high-dimensional datasets and give more scope for us to analyze the diseases due to change in gene expression levels in the cell of human body. In this paper, gene expression datasets E-GEOD-6575 and of tumor are analyzed using heat map and classification technique. The gene expression dataset of RNA-seq tumor of 801 samples based on classification is discussed. Each sample of tumor dataset is associated with set of gene expression values. The gene expression values in the dataset of 801 patients are distributed among the five different types of classes of tumors. The main objective of this paper is to understand the significance of change in gene expression values in the dataset that are the cause of diseases. The analysis will be helpful in prognosis of diseases and will help the physicians for prediction and deriving conclusions. The analysis is also helpful to identify the existence of differentially expressed genes.

Keywords Classification · Differentially expressed genes · GEO dataset · Gene expression · Heat map · National Center for Biotechnology Information · NCBI

1 Introduction

In an organism, each cell carries same number and types of genes, and the cell carries gene expression. The gene expression is changing with time due to changes

S. Ghadle (✉) · R. Tripathi · S. Kumar
National Institute of Technology, Raipur, Chhattisgarh, India
e-mail: saroj.shambharkar@gmail.com

V. Munde
TCET College, Mumbai, Maharashtra, India

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_23

in cell activities, cell development stages morphological, physiological activities like sporulation or changes in gene expression may happen in response to environment.

There are two technologies, microarray and RNA-seq technology, used to generate the gene expression data.

The paper's objective is to understand the importance of analysis of gene expression dataset. There are various techniques and methods used to visualize, interpret, and understand the gene expression data. The objective behind the visualization, interpretation, and understanding gene expression data is to get the knowledge about the condition of a human being. The analysis is helpful for the physician for proper diagnosis of the disease.

The National Center for Biotechnology Information, NCBI Web site provides us the gene expression dataset. The required gene expression datasets can be searched through Gene Expression Omnibus (GEO). From NCBI, the Gene Expression Omnibus, GEO dataset E-GEOD-6575, gene expression dataset is imported from NCBI. The domain experts have uploaded their data after conducting experiments on various diseases cases, and they are freely downloaded by the user. These datasets are authentic and regularly updated.

1.1 Gene Expression Dataset

The gene expression tumor dataset is downloaded from UCI machine learning repository. This dataset consists of 801 samples of five different types of tumor with 20,531 attributes. These attributes are associated with gene expression values. Before the analysis of gene expression dataset, the dataset has to go through the preprocessing depending of dataset considered by us. The gene expression dataset sometimes may get affected by the certain environment conditions. And because of change in environment conditions, the noise, missing values, and inconsistency are found in the dataset, and such dataset requires the preprocessing.

As we mentioned, the analysis is performed on the datasets of tumor and E-GEOD-6575 of gene expression in blood of children with autism. The dataset of gene expressions fetch from the Web site National Center for Biotechnology Information, NCBI, and UCI is used. The gene expression dataset of cancer (tumor) and gene expression dataset with accession number E-GEOD-6575 are considered and studied. The paper focused on analysis of change in gene expression attribute values is associated with each sample.

2 Literature Survey

Earlier in past decades, linkage analysis was carried out for the purpose of identification of novel disease genes. The problem or disadvantage of linkage analysis is it is time-consuming and expensive method. In the research papers, it is mentioned that the diagnosis of autism spectrum disease is costly and time-consuming process using linkage analysis, so to overcome these problems the machine learning approach can be used.

In the literature, the authors used a random forest classifier, and they achieved 80% accuracy for predicting new autism spectrum disorder (ASD) genes and their results are better than previous work. In this paper, for identification of biological markers the machine learning methods were used, and they used gene ontology to improve the identification of biological markers in complex diseases like ASD. The training performed by machine learning classifier is done using gene ontology. And performance can further be improved by incorporating functional similarities. There are algorithms like random walk restart (RWR) which are widely used for prediction of disease genes. The gene expression helps the physician to know the disease transition in human body due to change in gene expression profile.

In the paper "Prediction of gene function by genome-Scale Expression Analysis: Prostate cancer-Associated genes," the authors' research objective was identifying the diseases associated with genes and discovered for cancer using the guilt by association method. In this paper, they worked on the expression pattern consist of 40,000 human genes taken from 522 cDNA libraries.

From the research papers, it is understood that the disease similarity network can be extracted from a disease similarity matrix, using text mining techniques performed by taking OMIM records. But the accuracy of this approach is less and it is improved if we use phenotype ontology database and also improves prediction performance.

The study on interaction between the protein in protein--protein interactions (PPI) network provides us the biological insights going with the cells of an organism. So, it is important to have an automated method or high-throughput tool which can efficiently predict the PPIs. In view of efficient prediction from such network of interaction, from the literature, it is found that there is a predictor, called iPPI-PseAAC (CGR). It was developed by authors to incorporate the information of "chaos game representation" into pseudo-amino acid composition. This predictor has used the machine learning algorithm called random forest. The advantage of key sequence-order or sequence-pattern information obtained is more effectively incorporated during the treatment of the protein pair samples. The cross-validations applied on their datasets have given higher success rate than the existing approach [1].

The challenges identified from literature survey for the study toward our research are given below:

(i) In post-genomic scenario, the challenge is to identify and predict "disease genes" from the large amount of genetic-related information.
(ii) Identify biological marker from complex diseases is difficult as these diseases present highly heterogeneous genotype.
(iii) Identify and predict human disease-resistant genes from large genomic data.
(iv) Identify and predict disease and non-disease genes from genomic dataset or predict true disease genes and essential genes from genetic dataset.

To solve the above-mentioned challenges, there is a need of some techniques, computational methods, approaches, or tools. The question comes from the study for complex diseases how we can analyze and identify the gene expressions are changed or not. And what is the purpose of analyzing gene expression datasets. From the literature, it is observed that the researchers employed various techniques to study and analyze the genes-related information; some of them used machine learning approach and deep learning approaches and concluded they are effective, powerful computational technique to understand and analyze the genomic datasets.

2.1 Visualization and Interpretation of Gene Expression Dataset

The gene expression dataset E-GEOD-6575 is analyzed using heat map which is combined with clustering. The most common method to visualize the gene expression dataset is heat map. The heat map is a grid-like structure containing rows and columns, the rows indicating genes and columns indicating samples. The color and intensity of each grid indicates the state of gene expression. From the color and intensity, we can analyze that there is change in gene expression.

The heat map of E-GEOD-6575 gene expression dataset is obtained using R-programming shown in the figure. The heat map shown is containing the group of gene expressions patterns based on the similarity. Therefore, the most similar gene expression patterns were grouped together.

Figure 1 showing values using different colors in each sample. The colors yellow and red are used to analyze the attribute values of E-GEOD-6575 gene expression. In Fig. 1, there are three samples (represented using 3 columns) containing probe sets 204252_at, 211803_at, 1804_s_at as attributes or considered as samples. The rows of heat map consist of gene expression values of genomic series GSE6575 dataset. In middle column (sample specified by probe set 211803_at), we can observe the similarity is more and they used same color. There is less change in gene expression values more are in red color and less values in yellow colors. It means less change in gene expression values in middle sample. But if we observe the sample specified by probe set 204252_at and 1804_s_at, there is more variation in values for same set of values (values specified by rows) and less variation in sample 1804_s_at as compared to 204252_at. If the change in expression values is more, we conclude them as differentially expressed. So, among three samples, we conclude more changes or differentially expressed genes in first leftmost sample (204252_at).

In this way, heat map is giving good visualization and analysis of values stored E-GEOD-6575 gene expression dataset.

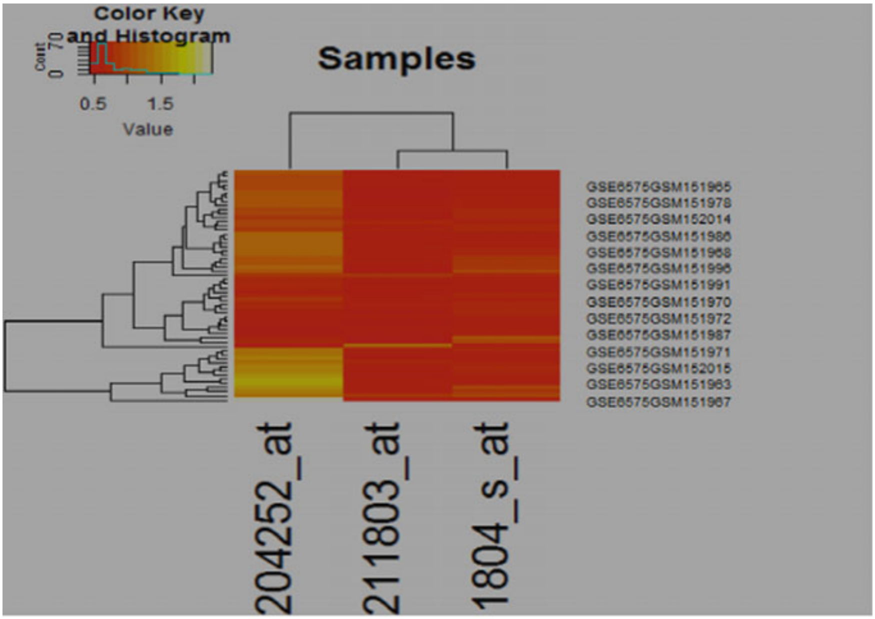


Fig. 1 A figure showing the visualization of gene expression dataset GEOD-6575 using heat map

2.2 *Classification of RNA Seq. Gene Expression Dataset of Cancer*

The gene expression dataset for cancer RNA-seq dataset is fetched from the UCI repository (machine learning). The size of the dataset is 201.354 MB. The 801 samples were taken, and each sample contains the change in gene expression value. Each sample consists of numerical values of gene expressions. There are total 20,531 attributes from gene_0 to gene_20530. This dataset also giving us different ways to analyze the classification of each sample based on values associated with each sample.

Figure 2 is showing the classification of 801 patient samples RNA-seq PANCAN dataset. The dataset referred by us is used by the authors Weintenstein, John N. et al. in their paper “The cancer genome atlas pan-cancer project" published in Nature genetics 45.10 (2013).

The patients are classified into five classes of different types of tumors. The patient is classified into different categories of tumors and found more cases under category of BRCA in the dataset of 801 samples of tumor patients. The classification tells us the 801 samples containing 300 patients suffered from BRCA, 146 with KIRC, 141 of LAUD, 136 with PRAD, and only 78 with COAD tumor. This helps a physician to analyze the patients with BRCA tumor, what makes changes in gene expression and how it can be reduced. From the samples of 801 patients, we conclude from

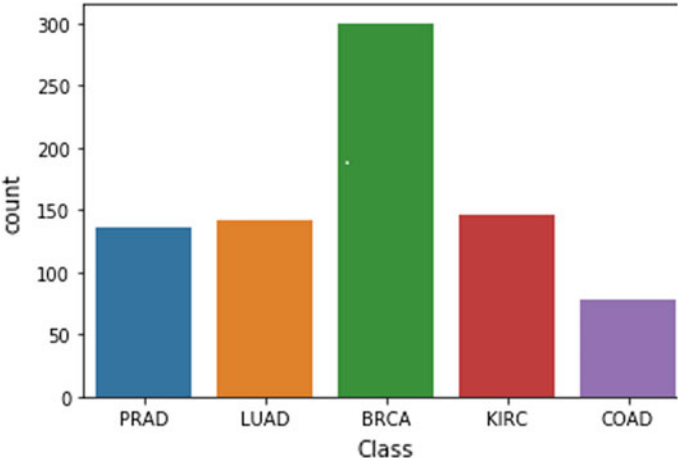


Fig. 2 A figure showing the classification of RNA-seq gene expression dataset of cancer

the classification more changes in gene expression values in samples of BRCA and conclude more differentially expressed genes in it.

This classification done for the patients whose changed expression values changed and the change in gene expression shows the patients are suffered by cancer. From the classification results, we say the gene expression dataset analysis gives us proper decision on the available sample of a patient.

3 Conclusion

In this paper, the main focus is on understanding the gene expression dataset and change in values of gene expression. The change in gene expression value (not as in normal state of genes) inside the cells is identified as differentially expressed genes. It is studied from the various approaches toward gene expression dataset analysis that the identification of differentially expressed genes will help in analyzing the state of a patient. From analysis of gene expression dataset, the differentially expressed genes can be identified, we can also identify up-regulated and down-regulated genes as a future work from genetic disease datasets available on NCBI. The value of log2FC (log to the base 2 of fold change) greater than threshold say 1 the genes are up-regulated otherwise down-regulated.

From the literature, it has been observed the identification of the differentially expressed genes requires the comparison of gene expression levels or read counts at different conditions of an organism say human. The difference between gene expression level at normal condition and abnormal condition has to be calculated to diagnose the disease. If the difference is found, then we conclude the gene expression levels have been changed and the patient condition is not normal.

In future work, we will work on analysis of gene expression dataset on other diseases using different machine learning algorithms and other approaches to identify differentially expressed genes and significant genes.

References

1. Jia J, Li X, Qiu W, Xiao X, Chou K-C (2019) iPPI-PseAAC(CGR): identify protein-protein interactions by incorporating chaos game representation into PseAAC. J Theor Biol 460:195–203. 7 Jan 2019. Epub 2018 Oct 9(2019). https://doi.org/10.1016/j.jtbi.2018.10.021

A Modified EDDKA Routing Approach to Enhance Quality of Service (QoS)-Enabled Data Dissemination in VANETs

Amol S. Dudhe and Salim Y. Amdani

Abstract VANETs have gained great focus among the researchers since last few decades. The areas of great interest are the types of routing and quality of service (QoS). The main challenge is to find the ways to counter the continuously changing VANETs topologies and its high speed nature, and then determining which routing protocols are best suited for a particular transmission type, and which ones provide more consistent and stable routing performances. Routing in networking is one of the key factors in determining how effective and efficient data packets can be transmitted within a network environment. As the topology is consistently ever-changing, a reliable communication with needed QoS is of utmost importance. QoS with respect to routing protocol is a domain which demands improvement. Great efforts have and are still being made to improve the routing capabilities of the protocols in existence to enhance QoS mechanisms in routing protocols. The simulation results indicate that the proposed approach can lead to improvements in terms of QoS metrics like throughput, packet delivery ratio, end-to-end delay, energy consumption, and residual energy.

Keywords VANETs · QoS · Throughput · Packet delivery ratio · End-to-end delay · Energy consumption and residual energy

1 Introduction

Traffic safety is a major challenge recognized by the major players in the automotive industry and by many governments, according to which annually thousands of road accidents are reported in any country. Traffic accidents are most of the day's results of the driver's failure to access quickly and properly the driving conditions.

A. S. Dudhe (✉)
Department of Information Technology, BNCOE, Pusad 445204, India
e-mail: amol_dudhe@rediffmail.com

S. Y. Amdani
Department of Computer Science and Engineering, BNCOE, Pusad 445204, India
e-mail: Salim_amdani@rediffmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_24

Normally drivers have imperfect information about road situations, speed, and position of vehicles around them and typically are compelled to form decisions like breaking and lane changing without the benefit of whole data. "The need for communication when the deployment of any fixed infrastructure is impossible and therefore the advancement of computer and wireless communication technologies, led to the event of Mobile Ad-hoc Networks (MANETs)" [1]. Vehicular ad hoc networks as a subset of mobile ad hoc networks which provide data exchange via vehicle-to-vehicle (V2V), vehicle-to-roadside (V2R), and vehicle-to-infrastructure (V2I) communications and a car which takes part in such a network are equipped with a WLAN and cellular communication device [2]. VANETs are additionally defined as a wireless communication technology which is additionally ready to enhance driving safety and velocity by exchanging real-time transportation information, and "it should upon implementation, collect and distribute safety information to massively reduce the amount of accidents by warning drivers about the danger before they really face it" [3]. In addition, VANETs also are ready to minimize incidents and improve traffic conditions by providing vehicles, drivers, and passengers with information about the road condition. VANET has its own unique characteristics in comparison with other sorts of MANETs, and the unique characteristics of VANET include: predictable mobility, lack of powers constraints, variable network density, rapid changes in network topology, high computational ability, and large-scale networking [4].

Safety services information such as traffic accidents and road congestion which are sensitive to reliable and real-time communication should be broadcasted immediately. Data transmission in such environment is critical and has got to be distributed in multiple paths to enhance the end-to-end delay. Some stale routes are generated in the routing table which leads to unnecessary routing overhead causing frequent link failures as well as route discoveries. Therefore, the discovered route between couple of vehicles should be as stable as possible to satisfy QoS requirements [5]. The intermittent nature and short-lived of these algorithms make the created clusters to provide scalability with lower communication overhead [5]. Rapid change in topology, owing to time varying vehicle densities and other factors, both external and internal, makes preserving a route very difficult and this in turn incurs high routing overheads as well as low throughput [6]. A cluster on-demand minimum spanning tree with Prim's algorithm has been proposed in [7] for VANET. In this approach, the vehicles have been clustered by accounting the intra-cluster QoS. An extended Kruskal's algorithm has been proposed in [8] to support QoS.

Thus, we aim to develop a QoS-enabled data dissemination using Kruskal's algorithm to provide efficient data dissemination (QoS-EDDKA) and quality of service in VANET. This approach constructs the clusters aid of minimum spanning trees in every road segments by considering the intra-cluster QoS. Each spanning tree will have a cluster head that is responsible to collect or disseminate the data from the leaf nodes and to other coordinator nodes and vice versa.

2 Related Work

A bee communications-inspired QoS routing scheme to improve the network throughput and to reduce packet loss in vehicular ad hoc network (VANET) environment has been proposed in [9]. Their scheme was inspired by the natural bee communications paradigm. The routing scheme is inspired by the bee's foraging behavior in the nature. Control messages are exchanged within all nodes in the cluster to ensure a reliable and stable connection. However, this may contribute to overhead problem in routing.

In [10], the authors present cross-layer-based vehicular routing model (CL-RVR), to facilitate reliable routing in VANETs, which would cater to QoS requirements for desired applications by combining the parameters from physical and network layers. Their simulation results show that their proposed scheme can improve the packet delivery rate and reduce link failure ratio in routing in VANETs.

A cluster-based routing protocol for urban vehicular ad hoc networks (VANETs) using optimized link state routing (OLSR) was presented in [11]. In their model OLSR, multiPoint relays (MPRs) used for routing are selected at each node using neighbors' reachability resulting in a high percentage of MPRs.

In [12], the authors present a link prediction algorithm as an addition to the AQRV protocol that selects the best routing path in urban vehicle environments to predict the duration of availability of the present path. This methodology expects to anticipate a connection breakdown before its incidence and directing the data packets through a substitute route. A QoS-enabled data dissemination using Kruskal's algorithm provides efficient data dissemination and quality of service in hierarchical VANET [13].

3 System Model

The proposed system model consists of n number of vehicles and road-side unit (RSU). Every vehicle can communicate with other vehicles and also with RSU. QoS provisioning often needs negotiation between host and network, calls admission control, resource reservation, and priority scheduling of packets [14]. A communication link between the nodes will exist if they are present in their communication range. The vehicles present within the radius of the intersection point are not allowed to form the clusters. Figure 1 shows the clustering of vehicles in VANETs with the proposed technique. The vehicles which are away from that region can form and join the clusters. It has been assumed that if the road contains any of the road-side unit then it will act as the cluster head for that region. Vehicle clustering represents a management scheme in which the adjacent vehicles are gathered into a group known as a cluster. Each cluster has one active node that plays the role of a CH. CH is selected to control and manage the cluster activities. All the other nodes in each cluster are entitled cluster members. These members are usually belonging to one

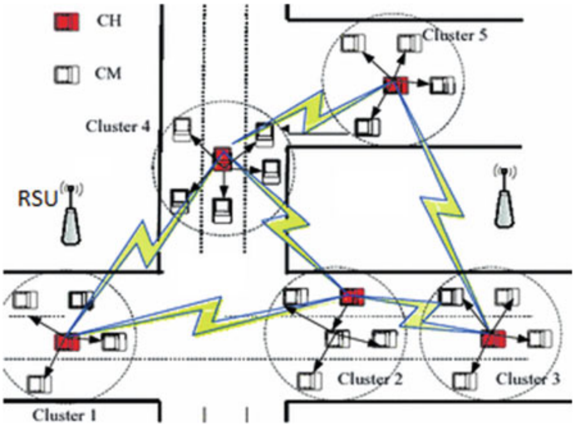


Fig. 1 Clustering of vehicles in VANETs

cluster or in some cases it belonged to multiple clusters. The cluster head (CH) can communicate with the cluster member (CM) as well as with the other cluster heads.

The number of clusters can be formed on a road segment and can be determined by using following equation

$$\text{No of Clusters} = \frac{(1 - 3 * TR)}{3 * TR}$$

Here, *l* is the length of the road segment measured in meters; *TR* is the inter-cluster communication range of the nodes. According to this, the road has been segmented and the cluster head has been selected in each segmented road. Figure 2 shows the segmentation of road in the proposed approach. The clusters will be formed for the nodes which are traveling on the same direction.

The node stability has been computed based on the time to live (TTL) and the acceleration of the vehicle

$$\text{TTL} = \frac{D_\text{n} - \text{rs}}{S}$$

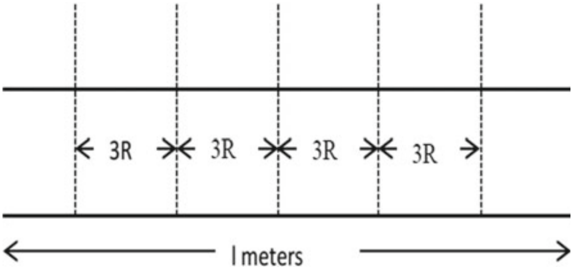


Fig. 2 Segmentation of road

where D is the distance between the node and the road segment, S is the speed of the node. The stability of the node can be given as follows,

$$\text{NS} = \frac{\text{TTL}}{\alpha}$$

$$\alpha = \frac{\text{Size of the queue} - \text{length of the queue}}{\text{Size of the queue}}$$

where size of the queue is the present size of queue at node i, length of the queue is the queue length between the nodes i and j.

The probability density function (pdf) has been calculated for every node relied upon the NS value.

$$\text{Pdf} = \frac{1}{\sigma\sqrt{2\Pi}} \text{e}^{\left(-\frac{1}{2}\right)\left[\frac{S-\mu}{\sigma}\right]^2}$$

Here, μ is the mean of the NS value gotten from the nodes in the road segment, σ is the standard deviation valued to the particular mean.

The signal-to-interference ratio between a node i and node j can be given as follows

$$y_{i,j} = \frac{\text{pt}_{i,j}}{\sum_{K=1}^{N} \text{pr}_{j,k}}$$

Here, N is the number of channels available between the node I and node j, $\text{pt}_{i,j}$ is the transmitted power between the two nodes i and j. $\text{pr}_{j,k}$ is the measured received power at the node j on the channel k.

The weight for each link based on NS can be given as follows

$$W_{i \to j} = \alpha * \text{NS}_{i,j} + \beta * y_{i,j}$$

Initially, the VANET topology was created by SUMO 30.0.0 and MOVE.jar in the Ubuntu 14.04 LTS. The vehicle (CM) wants to send the message communicated to the corresponding cluster head for transmitting the message by sending group information request. In order to provide data security within the message digital signature was attached by using ECDSA algorithm accordingly it generates the private and public key and also it encrypts the data. The road-side unit (RSU) holds both private and public key information. Every vehicle which sends group information request message was encrypted by private key of his own with certificate attached as shown in Table 1.

This message is signed by public key of RSU and it is forwarded to trusted authority (TA). TA decrypts this message and checks certificate of vehicle from its database. If it is valid then it sends verification message to RSU. TA assigns pseudo

Table 1 Frame format for group information request message

Vehicle ID	Message ID	Message	Timestamp	Certificate

Table 2 Trusted authority table

Vehicle ID	Pseudo ID	Temporary public/private key pair	Timestamp

Table 3 RSU vehicle information table

Vehicle pseudo ID	Pseudo ID	Temporary public/private key pair	GL/GM

Table 4 Cluster head table

Vehicle pseudo ID	Temporary public/private key pair	Pairwise key

ID to vehicles and temporary public and private key pair. TA sends this pseudo ID, temporary public, and private key pair, and this message is encrypted by permanent public key of vehicle as shown in Table 2.

In this way, all the vehicles get authenticated and RSU form cluster of vehicles, and also it selects group leader (GL) or cluster head (CH) and assigns group ID. The format of it is as shown in Table 3.

Cluster head (CH) or group leader (GL) maintains table of vehicles in group and which has vehicles' pseudo ID and their temporary public and private key pair as shown in Table 4. Cluster head then broadcast this message to all vehicles encrypted by group key.

4 Algorithmic Steps

```
Start:
        Algorithm packet Scheduler (Graph T){
    // Problem description : Hierarchical Vanet Packet Scheduler
    For every vertex V in V[G]:
        Set priority queue Q which composed of all the edges of G
        Remove all the Successive tree links for node(Vertex V) to first trust node.
        R <- Total no of removes links
        Best candidate <- b
        Run BFS Algorithm from v in a radius
        If nodes other than b on disconnected main tree:
            Candidates <- nearest Candidate nodes into tree list
            Best weight =infinity
        For each node g in the candidate list:
            Hg  <- Hop count from g to the sink
            Dg <- Degree of node g
            If (Dg+Hg)< Best weight:
                Best weight=Dg+Hg
                Best candidate <-g
        Connect node V to Best candidate in shortest path.
```

In our work, we run BFS algorithm from v in a radius to form the minimum spanning tree in which best weight considered as infinity, nearest candidate with minimum weight is added to tree list. For each node g in the candidate list Hop count (Hg) from g to the sink and Degree of node g (Dg) is calculated. If sum of hop count (Hg) and degree of node g (Dg) are less, then the best weight is modified as Dg + Hg and node g is considered as best candidate. Connect node V to best candidate in shortest path.

5 Simulation Results

6 Conclusion

Through the simulation results obtained from NS2 as shown in Figs. 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12, we can conclude that by using proposed EQDKA algorithm QoS has been improved with respect to throughput, packet delivery ratio, end-to-end delay, energy consumption, and residual energy.

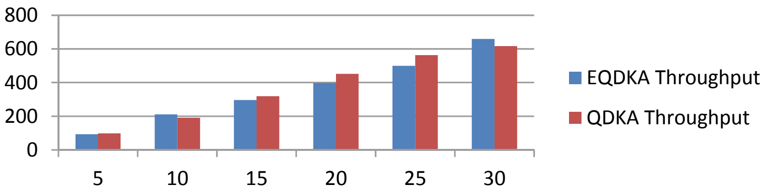


Fig. 3 Comparative analysis of QDKA and proposed EQDKA algorithm on throughput with respect to time

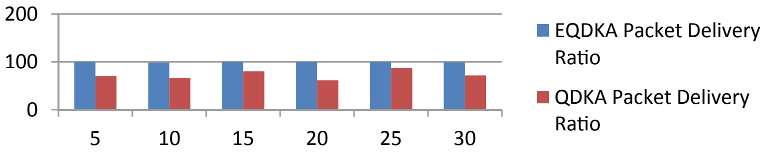


Fig. 4 Comparative analysis of QDKA and proposed EQDKA algorithm on packet delivery ratio with respect to time

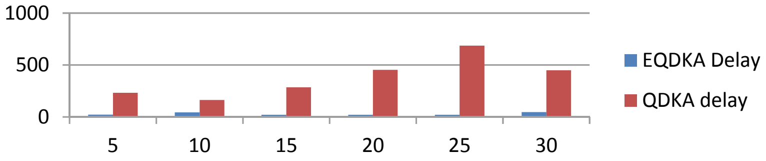


Fig. 5 Comparative analysis of QDKA and proposed EQDKA algorithm on end-to-end delay with respect to time

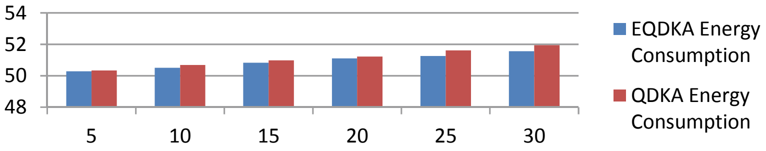


Fig. 6 Comparative analysis of QDKA and proposed EQDKA algorithm on energy consumption with respect to time

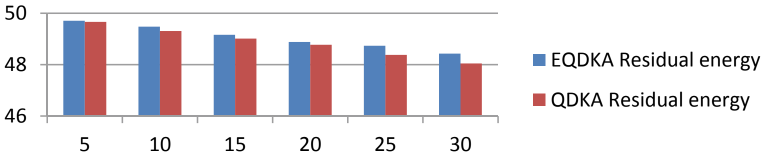


Fig. 7 Comparative analysis of QDKA and proposed EQDKA algorithm on residual energy with respect to time

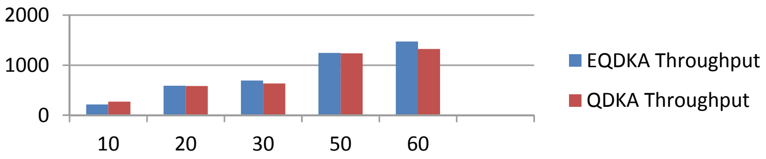


Fig. 8 Comparative analysis of QDKA and proposed EQDKA algorithm on throughput with respect to number of nodes

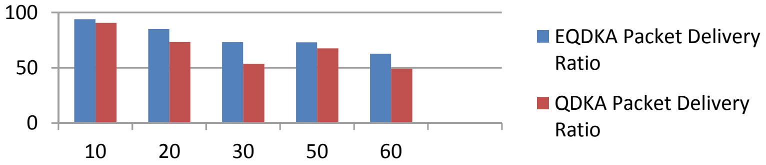


Fig. 9 Comparative analysis of QDKA and proposed EQDKA algorithm on packet delivery ratio with respect to number of nodes

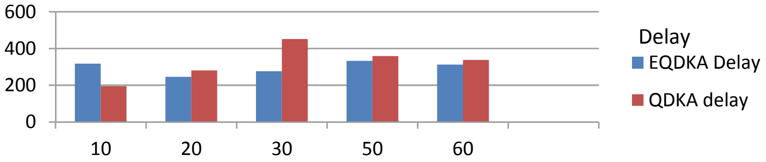


Fig. 10 Comparative analysis of QDKA and proposed EQDKA algorithm on end-to-end delay with respect to number of nodes

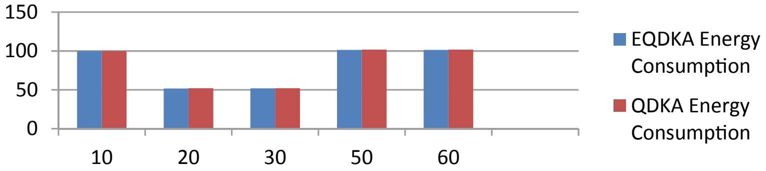


Fig. 11 Comparative analysis of QDKA and proposed EQDKA algorithm on energy consumption with respect to number of nodes

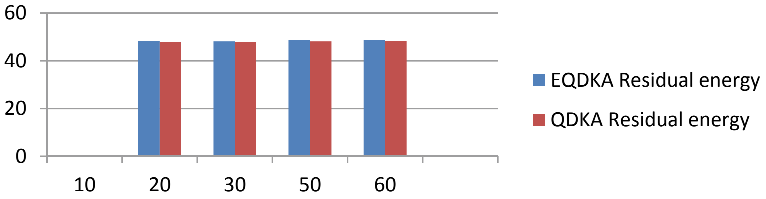


Fig. 12 Comparative analysis of QDKA and proposed EQDKA algorithm on residual energy with respect to number of nodes

References

1. Khairi D, Berqia A (2015) Survey on QoS and security in vehicular ad-hoc networks. Int J Adv Res Comput Sci Softw Eng 5(5):42–52
2. Shrivistava P, Ashai S, Jaroli A (2012) Gohil, vehicle to road-side-unit communication using Wimax. Int J Eng Res Appl 2(4):1653–1655
3. Adibi S, Erfani S (2015) Mobile Ad-hoc networks with QoS and RSVP provisioning. In: CCECE, Saskatoon, Canada (2015)
4. Al-Sultan S, Al-Doori M, Al-Bayatti A, Zedan H (2014) A comprehensive survey on vehicular ad hoc network. J Netw Comput Appl 37:380–392
5. Múazu AA, Fageeri SO (2017) Enhanced bandwidth reservation guarantees for QoS routing in vehicular network. In: 2017 international conference on communication, control, computing and electronics engineering (ICCCCEE), Khartoum Sudan, pp 12–17 (2017)
6. Bernsen J, Manivannan D (2009) Unicast routing protocols for vehicular ad hoc networks: a critical comparison and classification. Pervasive Mob Comput 5:1–18
7. Kponyo JJ, Kuang Y, Zhang E, Domenic K (2013) VANET cluster-on-demand minimum spanning tree (MST) prim clustering algorithm. In: International conference on computational problem-solving (ICCP), Jiuzhai, pp 101–104
8. Aissa M, Mnaouer AB, Murray R, Belghith TA (2011) New strategies and extensions in Kruskal's algorithm in multicast routing. Int J Bus Data Commun Netw 7(4):32–51
9. Jung LT et al (2014) Bee inspired QoS routing in VANET. In: 2014 4th world congress on information and communication technologies (WICT 2014), Bandar Hilir, Malaysia, pp 176–181 (2014)
10. Gawas MA et al (2018) Cross layer approach for neighbor node selection in VANET routing. In: 2018 11th international symposium on communication systems, networks & digital signal processing (CSNDSP), IEEE, pp 121–127
11. Kadadha M et al (2018) A cluster-based QoS-OLSR protocol for urban vehicular ad hoc networks. 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC). Limassol, Cyprus, pp 554–559
12. Naresh M, Raje A, Varsha K (2019) Link prediction algorithm for efficient routing in VANETs. In: Proceedings of the third international conference on computing methodologies and communication (ICCMC 2019), pp 1156–1161
13. Mukunthan B et al (2017) A Novel Krushkal's algorithm approach to ensure quality of service enabled data dissemination in hierarchical vehicular ad-hoc network. IRACST—Int J Comput Netw Wirel Commun (IJCNWC) 7(4):6–12
14. Saharan S, Kumar R (2015) QoS provisioning in VANETs using mobile agent. Int J Comput Sci Commun 1(1):199–202

A Novel and Secure Approach for Quantum Key Distribution in a Cloud Computing Environment

Rajanikanth Aluvalu, Krishna Keerthi Chennam, V. Uma Maheswari, and M. A. Jabbar

Abstract The vast majority of the clients worry about the security, privacy, and verification in the cloud computing framework. Cloud foundation requires better approaches for giving security. Public key infrastructure (PKI) is important critical arrangement. PKI depends on asymmetric key cryptography (AKC). The issue is accidentally losing the secure key might be hopeless. The recommended method is to make quantum key distribution (QKD) reasonable for a multi-party framework with quantum user authenticate plan utilizing a secured finite key. We use network control and acknowledgement of threat authentication plan to accomplish objective. The principle highlight of the proposed plan utilizes multi-party key and lower utilization of quantum transmission. In addition, the loss and other factors are taken care of utilizing this multi-party key and correspondence with various groups. The results shows the high security of quantum cryptography where malicious or eves dropper unable to retrieve the information in the future.

Keywords Public key infrastructure (PKI) · Quantum computing · Quantum key distribution (QKD) · Asymmetric key cryptography

1 Introduction

The vast majority of the clients concur that privacy, uprightness, and authentication are the important worries in cloud framework. Numerous confirmation plans have been present in quite a while. The majority of the scientist concurs that cloud foundation requires better approaches for giving security. As of now, arrangement of PKI is a major critical arrangement. PKI including changing key utilizes declarations by means of open channel to validate clients in the cloud foundation. Anyway there are

R. Aluvalu (✉) · V. Uma Maheswari · M. A. Jabbar
Vardhaman College of Engineering, Hyderabad, India
e-mail: rajanikanth.aluvalu@gmail.com

K. K. Chennam
Muffkham Jah College of Engineering and Technology, Hyderabad, India

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_25

few issues relating to the PKI verification where people in general key cryptography just gives computational security in light of the fact that PKI depends on AKC. It is presented to far-reaching security dangers, for example, snooping, man in the middle attack (MITM), and so on. This implies that the MITM could undoubtedly regulate an individual's secret key which leaks the data to others. Different issue is that the private key may lost which might be hopeless sometimes, where the messages which are received cannot be decoded any longer if the secret key is lost. This needs the requirements of new confirmation system with a protected channel that includes different clients that guarantee the safe communication.

QKD is one of the endeavoring and substantial items accessible in the market. For the most part, QKD is limited which tends toward to the multiple party framework. Building up a multi-party framework is a difficult investigation. In any case, QKD is helpless against accessible and loss aspect. Moreover, QKD has insufficient authentication by the user. This outcomes QKD framework intended to less outlook. The recommended method is to create QKD reasonable for a multi-party framework with quantum user authenticate plan utilizing a secured finite key. We use network control and acknowledgement of threat authentication plan to accomplish our objective. Multi-party key is the most important essential viewpoints in organization security system. The principle highlight of the recommended plan utilizes multi-party key and lowers utilization of quantum transmission. In addition, the loss and other aspects take care of utilizing multi-party keys and correspondence with different groups, by focusing investigation of size regarding key with secured finite key and security scrutiny with error rate.

Section II discusses about the preliminaries like quantum cryptography, cloud computing, and cloud security. Proposed methodology is discussed in section III. Illustration of conclusion is discussed in section IV.

2 Preliminaries

2.1 Authentication Scheme

Authentication is very popular concept with fundamental, and it is a procedure of deciding if a person or thing, whom can do it or what it is announced to be. Nowadays, communication security depends on the ambiguous computing security. Every transmission task transmitting the information or data with security group of three is primary to investigate. It required high security, honesty, and accessibility. Authentication is truly required in sincerity. It needs to guarantee that information did not revise because of malicious. Now, we understand the importance of authentication.

With fundamental cryptographic work's authentication with a significant undertaking, before correspondence that ensures that the client recognition and the original data are legitimate where the malicious user possibly change the original data of the real user, the distribution of key plans and encryption plans are effectively undermined.

Table 1 represents the current secured plans, authentication plans which can be written as digital challenges. All through research is recommending digital cryptography challenge on quantum. The convention given below discussed in Table 1 is fundamentally classical approach which is widely used.

Table 1 Digital challenge-response authentication schemes [1]

Protocol	Characteristics	Mechanism	Advantages	Limitation
Challenge-handshake authentication protocol (CHAP)	Authenticates a user or network host to an authenticating entity	Shared secret key, one-way hash function, three-way handshake	Protection against: replay attack	Distribution of secret key
CRAM-MD5	SMTP mail agent authentication	Hash function, concatenation, fresh random challenge	Resist to replay attack	Lack of mutual authentication storage of password, vulnerable to dictionary attack
Kerberos	Network authentication protocol over insecure channel	Issue tickets Trusted third party	Mutual authentication Resist to replay attack and eavesdropping	Single point of failure Vulnerable to man-in-middle attack Time constraints
Otway–Rees protocol	Network authentication protocol over insecure channel	Usage of nonce, session identifier, server	Resist to replay attack and eavesdropping	Vulnerable to intercept and resend attack
Needham–Schroeder protocol	Network authentication protocol over insecure channel	Shared secret key, server	Establish session key and mutual authentication	Key distribution
Wide mouth frog protocol	Network authentication protocol over insecure channel	Global clock, server, shared secret key, BAN logic	Resist replaying attack and eavesdropping Detection of modification	Key distribution and required trusted server
CAPTCHA, reCAPTCHA	Network user identification non-cryptographic scheme	Images	Widely used in webmail	Availability

2.2 Cloud Computing

This is the primary concept in the information and communication technologies (ICT) segment. The promotion cycle is alluded to cloud as the “most advertised idea in IT.” Cloud computing is an inclining position in Google since 2009 preceded with activity.

Step by step, the institution is anxious to execute cloud computing with ICT framework. The servers in cloud we discussed tend to physical or virtual machines. Cloud computing is another wording which is utilized from researcher and the industries. Cloud computing portrays idea where software, resources used, work as a virtual platform crosswise over a wide range of host machines, associated with the Internet or by an internal network by organizations. By the filed or framework client’s perspective, the cloud gives, by means of virtualization, a single platform where it can work [2]. National Institute of Standards and Technology (NIST) classified the cloud using with three different services.

Software as a service (SaaS) provides software applications readily available to use for subscribers. Microsoft online services that aided with unique models of Microsoft exchange and Microsoft share-point are henceforth as platform as a provider (PaaS). Making use of PaaS, cloud carrier provider develops the fundamental foundation, consisting of database and operating system that enables institutions to bring together and run packages utilizing languages and tools through the CSP. It can be observed in the Microsoft windows azure platform. The last service kind is infrastructure as a service (IaaS), wherein a CSP shows the combination access to basic IT industry on which the affiliation can bring its very own packages and records in a virtualized states that had been created utilizing tools and languages does not helped by the CSP. Times of this are Amazon’s EC2 and Rackspace’s cloud servers. This cloud version approaches openness and is constructed from five fundamental characteristics. The services grant, and there are additionally four deployment models exist in cloud computing. Basically, each model is shown in Table 2 (Fig. 1).

Bounded by SaaS, PaaS, and IaaS, infrastructure service apparently, the important CSP, previously tendering a major assortment of items and propelled abilities: computerized adaptability, pay-per-use, and on-request planning are probably great useful.

Virtualization is median for any cloud computing environment model. Serviced organizations, or private endeavor cloud providers, use virtualization innovation to acknowledge competency and adaptability provided by the cloud computing [4]. This

Table 2 Models in cloud computing [3]

No.	Model	Description
1	Private cloud	Operated by the individual industry
2	Community cloud	Operated with associated industries which have same services
3	Public cloud	Entire industry data is public for various users to modify, view, and add
4	Hybrid cloud	Communicating the private and public clouds

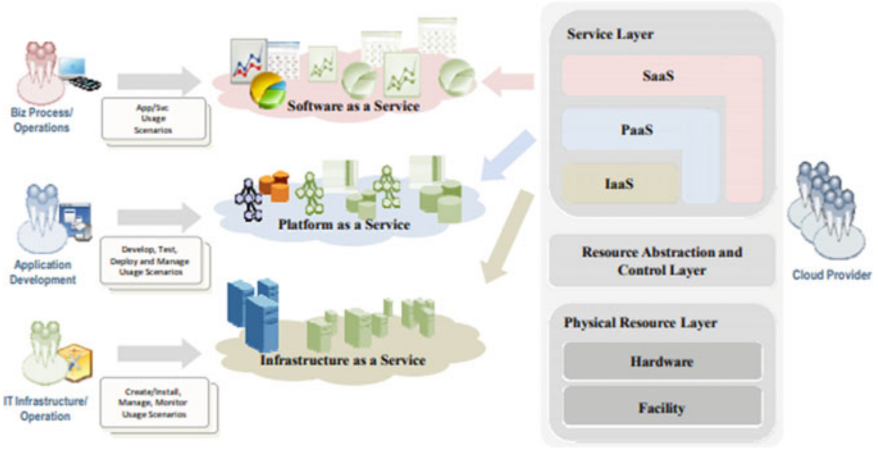


Fig. 1 Cloud architecture

protects the applications from security hazards and threats by web-based hackers to the hypervisor by cloud computing virtualization’s security mechanism and assure access to the multi-tenancy of systems which are virtual by the single host.

Virtualization likewise associates the unpredictability of proof conservancy, given the layered absorption where future data can reside here. It involves with much more noteworthy and progressively explicit aptitude which must be guzzled and used to guarantee that no trade-off or modification is influenced to computerized proof in a virtual machine condition. For instance, the host working framework in single layer about equipment and furthermore with the machine screen can be called as a hypervisor.

All conversations from the digital device and hardware are ideally communicated by the host’s main working device. The guest machine which is virtual maintains on another level of abstraction. It shows the collaboration with the outdoor place in hardware and various services possible by means of the help of the virtual system display/hypervisor. Based on this, view of VMware workstations on this view factor assures the effective correspondence with the host-running machine to empower its visitor platform to utilize resources.

2.3 Cloud Security

Leakage and malicious data is the biggest threat from the third-party cloud when data is shared in the cloud computing. By the way of referenced in the past area, cloud computing is a blend of different processing substances, universally isolated, yet electronically associated. As the computation is transferring toward corporate server rooms, this shows greater concepts along with safety, for instance, virtualization security, distributed computing, application protection, identity management, and authentication. In any case, high user authentication is significant for cloud computing to guarantee that is legitimate client who have access from the server.

Some authentication method shows all together for focusing the gaps. It depends maximum on client server. The initial isolated user authentication plans show the recommended method by Lamport [5], where the storage of server uses' password-hashed value. In Lamport's plan, password is utilized to check the authenticity of users, yet in the event that this secret key table is undermined, taken, or adjusted by an enemy, and at that point, the framework could be mostly or totally undermined [6–9] proposed a smartcard-based authentication model for the secured password, and a large number of the plans are appeared by [7, 10, 11]. Shoup-Rubin [12] recommended by Bellare-Rogaway model [13] depends on three party key distribution protocols and smartcard is utilized to store the long-term secret keys. By the plan, it is utilized to prevent the advisories, and it is accepted that smartcard is never undermined. Fundamentally, the plan shows a single factor class as multiple factor plans can be damaged by negotiating from the variables as it were. Liao et al. [14] attempted to merge various passwords- and smartcard-based effects and recommended two factors smartcard and secret word confirmation conspire. Cloud computing is a variation of customer server design, where many customers utilize a similar framework at a large scale. Therefore, it needs more grounded authentication than traditional customer server between systems administration framework with public key and authentication model by the band for cloud computing [15]. Be that as it may, the plan transmits information in a plaintext structure which can be effectively caught by advisories.

Mishra, Kumar, and Mukhopadhyay recommended method for combing free authentication model for cloud computing [16] updates the mutual authentication for client server, where plans are not address access control for cloud computing clients.

Alluding with the past research [17], which present a trusted cloud data server that is third party, is entrusted with guaranteeing explicit privacy qualities inside a cloud domain. The recommended method of arrangement shows on cryptography, and explicitly PKI is working together with single sign on (SSO) and lightweight directory access protocol (LDAP), to guarantee the confirmation, respectability, and secrecy of included information with correspondences. The arrangement introduces a flat degree of administration accessible to every single embroiled substance that understands a security work, inside which fundamental trust is kept up.

PKI which include changing of key, the use of certificates, and revocation listing must be with skills to identify the customers within cloud platform. All the users know about the importance of confidentiality, integrity, and authentication that is important issues in this cloud-based platform. By the way, the positive problem with the PKI authentication wherein the generating the public key cryptography simplest gives extreme protection due to the fact PKI is dependent on AKC, which is defined to considerable safety threats together with eavesdropping, guy inside the center attack, masquerade, and so on. The MITM may want to effortlessly show the persons private key. This can also margin to statistics leaking. The different trouble is the loss of secret key which cannot be re-collected back. It is not possible to decrypt again the retrieved data, if we lost the non-public key. Based on this method, the desires of authentication approach related to more than one users ensure the safety conversation across the kingdom. As we can see from above literature, the present authentication schemes still have a room for improvement. Thus, from there, we located a gap to decorate the existing schemes. Therefore, a great feasible solution to deal with this issue is by integrating the multi-party quantum key distribution (MQKD) protocol with the PKI. Integrated have the deployment of a high-secured finite key scheme to authenticate the cloud model regarding multi-user verbal exchange.

2.4 Quantum Cryptography

A long time ago, combo of quantum mechanics with cryptography was developed popularity and named as quantum cryptography. Advanced cryptography is used maximum in networks and is based on computational complexity. Many parts of the classical encryption algorithms in use today would be susceptible to being solved by a quantum algorithm in significantly less time than it would take for a classical computer. But, QC can offer without safety conditions mainly by means of its property of no-cloning theorem and Heisenberg's uncertainty method. Quantum-based total protection schemes can classify into main divisions known as single photon and entangled photon. A quantum-complicated state is a correlated kingdom among particles such that end result of measurement on one particle impacts the position of other particle, and this is naturally separated from the measured particle. Quantum cryptography utilizes the original properties of quantum mechanics together with great position, difficulty, and so on. With these characteristics, some facts can be secretly shared among customers through a quantum channel. The facts can be a key or a message. Quantum cryptography concerning QKD protocols is used to share a key, and quantum direct communication (QDC) protocols are hired to transfer or share a message [13].

QKD is a positive research with different protocols, methods, and functions. The logic with QKD ends up an on-call for studies due to the fact a hazard inclusive of acting or middle men attack shows the QKD susceptible. Being the authentication domain in quantum cryptography is the toughest session based on its degree of complexity. Based on this, quantum cryptography is most effective that shows with clear up the distribution of keys problem and now in reality transfers the required statistics. The advantage of the cryptosystem relies upon with all troubles with an eavesdropper which identifies the data. The advantage of quantum computing is that it will become very easy to find any cryptosystem. Classical cryptography is not a secure method by now and in future too. Securing reality and data are the most important than other due to the fact that the not secure data is processed which may have grave consequences on each the economy and final security. Quantum cryptography is predicated on the laws of quantum mechanics which gives safety device at the same time as the traditional device is based on the computational trouble of the encryption methods used to provide a safety device [18]. Based on this, research shows that QKD is a resource to the system in authenticate the communication in a cloud infrastructure. The final purpose of quantum key distribution protocols is to provide the reliable events, Alice and Bob with random, correlated, and personal classical statistics, the important thing. Along with this, quantum channel at their disposal is to be assumed completely under the management of the adversary.In this manner that anything in quantum kingdom which Alice or Bob sends via the channel, the output may be absolutely arbitrary, and the most effective restriction is consistent with quantum mechanics. Based on the quantum channel, the dependable events can make use of a public, classical channel, which is thought to be authentic, and by suggesting, it cannot change or fake data.

The distribution key is identified based on the events, and we must pay attention which came about implements extra than legitimate events. In clouds, we can see that it depends on a quantity or clients, and here, we will introduce using of MQKD. MQKD is a key distribution protocol where identical key is shipped to exceptional parties primarily based on quantum mechanism [19]. MQKD can be referred as a key distribution protocol establishes a commonplace key among a number of customers. By gaining the optimal solution and ease in MQKD, a popular cryptographic like authentication is necessary. Authentication is vital mission to ease the verbal exchange among customers. User identification and the foundation of information are needed to be actual, due to the fact that if a malicious user masquerades as a valid person, the important thing distribution schemes and encryption schemes might be without difficulty compromised.

As an earlier significant research, Matsumoto proposed a first convention without the utilization of entrapment to accomplish MQKD which empowers three gatherings that without a moment's delay on a mutual regular arbitrary piece strings within the sight of busybodies [20]. The fundamental distinction with suggested method convention and Matsumo's convention permits quantities of gatherings to share a typical secret key after the foundation about the personal key among the gatherings. Beside, our convention uses one-way open communication (post preparing) to share a last secret key. Here, Matsumoto's convention requires three-way posts preparing successfully. Every one of the gatherings is required to take part in the computation. In contrast, our suggested method convention requires just the sender to transfer an open message to the gatherings. For whatever length of time, people in general items are verified and unedited by Eve, and then, our proposed convention demonstrates unrestricted security. In addition, we utilize straightforward post-processing system to share a typical secret key among the parties.

2.5 Quantum Cryptography for Big Data Security

To improve the maximum security and database maintenance with minimum complexities in big data, the quantum cryptography is used. The big data security requires recollecting the symmetric key for a cipher text based on the design of cipher which is easy for the big data. A complication consistently improves when the big block ciphers are used, but the steps can reduce dynamically. Block ciphers are used for Grover's algorithm for the efficient key search which is another feasible approach of quantum cryptography to reach the big data security. Grover's algorithm is used for authentication server and the user for communication purpose.

Groves 'algorithm with block ciphers designed is very high for the conventional key management in the symmetric key for eventual data centers. Let the block cipher size of key is assumed as n. Eq. (1) is used to diminish the number of steps and the complications when the attack is started.

$$\text{Complexity} = O\left(2^{n}\right) \tag{1}$$

As per the Lin at el. [21–23], to identify the security and privacy with different areas, this can be used for cloud fundamentals and big data security. When user give information of big data, then the big data security issues will identify which requires the control of big data transmission among the authentication models in the data centers and users.

3 Proposed Methodology

3.1 Quantum Cryptography for Access Control

This is designed based on the solution of cryptographic methods by [20, 24, 25]. Regular methods are encrypting all the objects and then send the decryption keys to the legitimate agents. Assume that we are the owner of $\{data1, \ldots, datak\}$. We first use *key*1,..., *keyn* to encrypt those data. So, we have *Enckey*1 (*data*1), …, *Enckeyn* (*datan*). The *keyi* is given to the agent where they are eligible to access *datail*. Distributing key has major role in providing access.

Quantum Key Distribution
Three-pass protocol method is the following.

- Alice is appointed as a gate keeper to the computer café.
- Bob knows about that try to give some object to Alice.
- Bob contains a big strong safe which can keep the object safe and have various locks, and Alice cannot receive any object because Alice does not have any key for any of those locks.

Now, the question arises that how can Bob secretly and safely transfer the object to Alice?

Now, they may use the following three-pass protocol:

1. Bob keeps the object safely into one box and makes sure that he locked that box and sends it to Alice.
2. Alice identifies the box with object and again locks with her own one and sends to Bob.
3. Bob can unlock the Box and send to Alice again. At the end, Alice unlocks and receives the object which is kept in a box.

We can use regular cryptography by the simple exclusive-OR operation $\oplus$. This is not secure observed by eavesdropping. The safest protocol under eavesdropping is quantum cryptographic protocol because the quantum no-cloning theorem prevents the copy of non-trivial quantum states [26].

We use qubit $|0\_$ and $|1\_$ to encode 0 and 1, respectively. Our key space for encryption and decryption contains

4 X-gates $\{X(0), X(\pi 2), X(\pi), X(3\pi 2)\}$. The encryption of a qubit $|i\_$ with key k is defined as $Enck\ (i) = k|i\_$, and the decryption of a qubit $|i\_$ is $Deck\ (i) = k|i\_$, where $k \in \{X(0), X(\pi 2), X(\pi), X(3\pi 2)\}$. We let $(X(m), X(2\pi - m))$ be a pair of encryption/decryption keys.

Figure 2 shows our quantum three-pass protocol for a owner (agent 1) to send his key to a user (agent 2). At the beginning of the protocol, agent 1 encrypts the string element-wise and sends the resulting string to agent 2. Then, agent 2 encrypts the cipher text and sends the result back to agent 1. Agent 1 then decrypts the string and sends it to agent 2. Now, agent 2 decrypts the string and gets the key.

The correctness of our protocol is guaranteed by the rules (S) and $(I\ D)$ of the ZX-calculus. The (S) rule ensures that the sequential application of X-gates can be summed up, and while the $(I\ D)$ rule says the application of an $X(2n\pi)$, gate is the same as the identity operator.

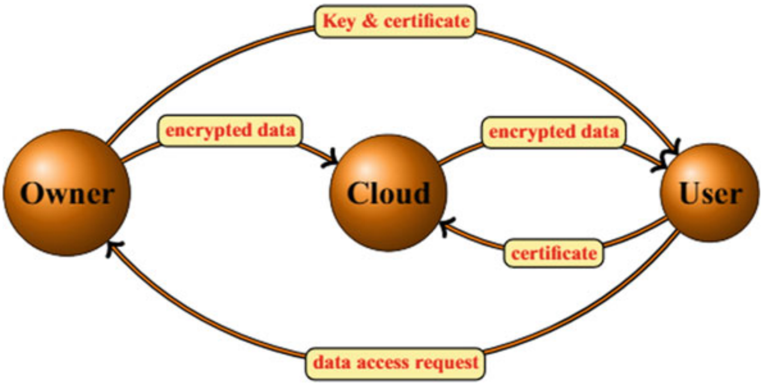


Fig. 2 Data access in cloud computing

4 Security of QKD

To identify the real-time situations in QKD, identify the eves dropper probability. By analyzing of QKD with multi-key part, the information and measurement results are showed in Figs. 3 and 4.

Figure 3 shows the eves dropper probability by identifying the records, where the probability of security is close to the 100% based on the eves dropper or malicious person. Figure 4 shows nearly close to the 30% transmitted photons. Based on these results, we can conclude that based on the quantum, communication is identified. The probability of eves dropper is detected with noise interference or not.

Fig. 3 QKD multi-part framework

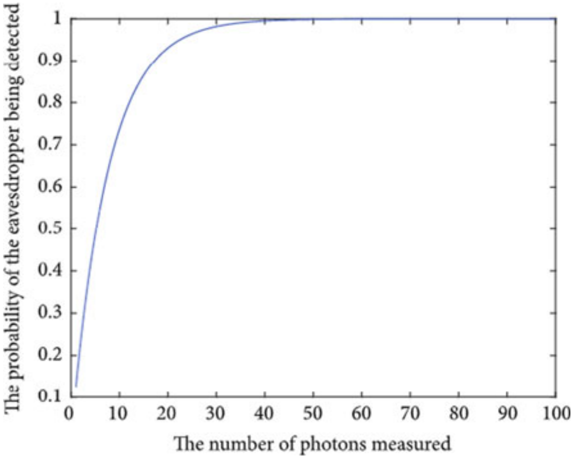
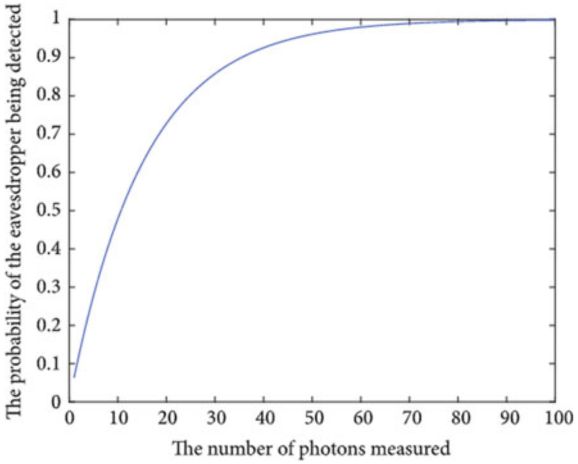


Fig. 4 QKD multi-part framework with noise channel



5 Conclusion

This paper discusses about quantum cryptography for access control in cloud computing. As an earlier significant research, Matsumoto proposed a first convention without the utilization of entrapment to accomplish MQKD which empowers three gatherings that concur without a moment's delay on a mutual regular arbitrary piece strings within the sight of busybodies. The fundamental distinction between our proposed convention and Matsumo's convention is that our convention permits quantities of gatherings to share a typical secret key after the foundation of secret key among the gatherings. In addition, we utilize straightforward post processing system to share a typical secret key among the parties. Our proposed system will enhance the security and privacy in cloud computing. This method can be widely used in many organizations which adopt cloud computing.

References

1. Mell P, Grance T (2011) The NIST definition of cloud computing. National Institute of Standards and Technology, U.S. Department of Commerce, Special Publication 800-145
2. Armbrust M, Fox A, Griffith R, Joseph AD, Katz RH, Konwinski A, Lee G, Patterson DA, Rabkin A, Stoica I, Zaharia M (2009) Above the clouds: a Berkeley view of cloud computing, Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley. https://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html
3. Baghel S, Theng D (2016) Multilevel security model for cloud third-party authentication. In: Shetty N, Prasad N, Nalini N (eds) Emerging research in computing, information, communication and applications. Springer, Singapore
4. AlZain M, Pardede E, Soh B, Thom J (2012) Cloud computing security: from single to multi-clouds. In: 2012 45th Hawaii international conference on system science (HICSS), pp 5490–5499, Jan 2012
5. Athena J, Sumathy V (2017) Survey on public key cryptography scheme for securing data in cloud computing. Circ Syst 8:77–92
6. Prajapati P, Patel N, Macwan R, Kachhiya N, Shah P (2014) Comparative analysis of DES, AES, RSA encryption algorithms. Int J Eng Manage Res 4(1):132–134
7. Mahajan P, Sachdeva A (2013) A study of encryption algorithms AES, DES and RSA for security. Glob J Comput Sci Technol 13(15)
8. Kumar P, Rana SB (2016) Development of modified AES algorithm for data security. Optik Int J Light Electron Opt 127(04):2341–2345
9. Khan SS, Tuteja RR (2016) Cloud security using multilevel encryption algorithms. Int J Adv Res Comput Commun Eng 5(1)
10. Setiadi I, Kistijantoro AI, Miyaji A (2015) Elliptic curve cryptography: algorithms and implementation analysis over coordinate systems. In: 2015 2nd international conference on advanced informatics: concepts, theory and applications, Chonburi, pp 1–6, 19–22 Aug 2015. https://doi.org/10.1109/icaicta.2015.7335349
11. Watson PJ (2012) Cloud Comp. Springer, Berlin, Heidelberg 1: 15. https://doi.org/10.1186/2192-113X-1-15, https://doi.org/10.1186/2192-113X-1-15, online ISSN 2192-113X
12. Dillon T, Wu C, Chang E (2010) Cloud computing: issues and challenges. In: Proceedings of the 24th IEEE international conference on advanced information networking and applications, pp 27–33

13. Kaushik S, Gandhi C (2016) An improved encryption and signature verification ecc scheme for cloud computing. In: Lobiyal D, Mohapatra D, Nagar A, Sahoo M (eds) Proceedings of the international conference on signal, networks, computing, and systems. Lecture Notes in electrical engineering, vol 396. Springer, New Delhi
14. Laurikainen R, Laitinen J, Lehtovuori P, Nurminen JK (2012) Improving the efficiency of deploying virtual machines in a cloud environment. In: Proceedings 2012 international conference on cloud and service, computing, pp 232–239
15. Metz C (1999) AAA protocols: authentication, authorization, and accounting for the Internet. IEEE Internet Comput 3(6):75–79
16. Ahuja SP, Komathukattil D (2012) A survey of the state of cloud security. Netw Commun Technol 1(2):66–75. https://doi.org/10.5539/nct.v1n2p66
17. Bessani A et al (2013) DepSky: dependable and secure storage in a cloud-of-clouds. ACM Trans Storage TOS 9(4):12
18. Cloud Security Alliance (2010) Top threats to cloud computing V1.0. Available https://cloudsecurityalliance.org/research/topthreats
19. Jansen WA (2011) Cloud hooks: security and privacy issues in cloud computing. In: Proceedings of the 44th Hawaii international conference on system sciences, Koloa, Kauai, HI. IEEE Computer Society, Washington, DC, USA, pp 1–1
20. Akl SG, Taylor PD (1983) Cryptographic solution to a problem of access control in a hierarchy. ACM Trans Comput Syst 1(3):239–248
21. Lin SH, Chiu JH, Lee GR (2020) A fast iterative localized RE-authentication protocol for heterogeneous mobile networks. IEEE Trans Comput Electron 56:2267–2276
22. Lin SH, Chiu JH, Shen SS (2020) The performance evaluation of fast authentication schemes in GSMWLAN heterogeneous networks. J Netw 5:956–963
23. Lin SH, Chiu JH, Shen SS (2011) The performance evaluation of fast iterative localized re-authentication for 3G/UMTS-WLAN interworking networks. J Ambient Intell Human Comput 4:209–221
24. Nagy N, Akl SG (2007) Authenticated quantum key distribution without classical communication. Parallel Process Lett 17(3):323–335
25. Bernabe JB (2016) TACIoT: multidimensional trust-aware access control system for the internet of things, soft computing—a fusion of foundations, methodologies and applications, May 2016, https://doi.org/10.1007/s00500-015-1705-6
26. Yanofsky NS, Mannucci MA (2008) Quantum computing for computer scientists. Cambridge University Press. https://doi.org/10.1017/CBO9780511813887

Voting Using Blockchain Technology

Shabnam Sayyad and Farook Sayyad

Abstract A growing list of records that are widely known as barriers, linked using the cryptographic hash of previous blocks, is called blockchain. Blockchain technology allows publicly distributed records that capture data completely in a secure and encrypted manner and ensure that transactions cannot be changed. Bitcoin and other crypto-currencies are the most common forms of blockchain behavior. Nowadays, there is a lot of research into blockchain technology being used to find solutions to real-world problems. After Bitcoin, many types of digital currencies were created. For many years, it was unclear that blockchain could only be used to keep track of transactions in the online currency. There has been little awareness of the use of blockchain. One example is relevant to the benefit of blockchain online voting or voting in the country. In this paper, we propose a voting solution made using Ethereum protocol. It uses smart contract structures to enforce strict rules around election votes. These votes both independently and customally validate and manage all the desirable blockchain structures (such as flexibility). This is achieved by maintaining voter confidentiality and voting integrity. The projected plan shows that it is possible for blockchain technology to be part of applications that wish to provide transparency and security to the public.

Keywords Blockchain · Online voting · Ethereal · Encryption · Blockchain voting features

S. Sayyad (✉)
Department of Computer Engineering, AISSMS College of Engineering, Pune, India
e-mail: ssshaikh@aissmscoe.com

F. Sayyad
Department of Mechanical Engineering, Dr. D Y Patil School of Engineering, Lohegaon, Pune, India
e-mail: fbsayyad@gmail.com

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_26

1 Introduction

For modern technology on the Internet, voting has been considered extremely glamorous due to major security issues. Voting online has started to struggle in today's society. It may seem strange that the acceptance of this method of voting has slowed down a bit, but if you look at the issues involved, the reasons will be clear. Online or remote voting is a very simple process, but it also leads to great potential for the damage of results. A unanimous vote agreement can lead to many changes in the choices made to voters. This means that online voting imposes very strict requirements on the security of all voting categories. We believe that blockchain technology is the link left in the creation of a possible online voting system. It has been a challenge to create an online voting system to satisfy the legal requirements of a democratic country. Blockchain technology has come up with exciting features, which can make the process of virtualization an easy victory. Blockchain offers an unlimited list of applications between online voting. The various features of blockchain technology can make a positive change for blockchain technology offering different features that can make a difference. With the rise of people, the increase of democracy is hard, trying to satisfy everyone. Using blockchain technology, uncertain issues related to people's identities, and visibility across long distances and complex trade-offs, such as election plans can be kept. This problem can be solved by building a very transparent system that people vote on the supportive blockchain technology has reduced our uncertainties about ownership. In this paper, our implementation will be posted on Ethereum's testnet to demonstrate the benefits, vulnerabilities, and efficiency of implementing a blockchain-based system.

2 Definition

Traditional paper ballots can be replaced by voting using blockchain technology. Security issues can be easily solved. Clearly, a public ledger [1] technology that can break the mistrust between central and voting authorities, thus providing a viable and democratic voting solution with the minimum need to know someone using a cell phone.

The blockchain is a "securely-chained singleton machine connected to a state-owned state." "Cryptographically secure" means that digital currency creation is protected by statistical algorithms that are hard to break. "Transactional singleton machine" means that there is only one sub-machine instance that takes care of all the operations in the system. In other words, there is one universal truth that is believed by everyone. By "shared state" means that the land stored on this machine is shared and open to everyone [2].

The blockchain is built on a computer-centric network, which stores performance records in the order of blocks. The first block is identified as a genesis block [3]. Some of the following blocks are linked to the previous block. Hash value of the previous block is stored on every block, thus linking all blocks. Hash is generated based on sales records in blocks. Changing the block records changes the hash and disables the chain. This makes blockchain unchanged.

3 Literature Review

The author uses a system in which a voter is given a mark that allows them to vote in their elections [4]. Privacy is an end-to-end (E2E) Evoting seminar activity that has become a thriving field. Generally, the idea of being a certified E2E refers to these two factors: First, every voter can confirm whether their vote has been filed as desired or listed as impersonators. Second, anyone can verify if all votes are given as written [5]. A proposed number of E2E-certified systems and some of these are used in practice [2]. The voting system, called DRE-i (DRE honestly) [6], is to achieve E2E verification without involving any tallying authorities (TAs) [7]. DRE-I has an environmental guidance plan that requires pre-recorded data to be stored securely and accessible during the voting phase. This introduces the possibility that our competitor will enter a secure storage module and promise the privacy of all votes. To overcome this discrepancy, they provide a DRE-ip voting system (DRE-i with enhanced privacy). DRE-ip achieves E2E authentication without TAs and is at the same time a more valuable privacy guarantee than DRE-i [8]. Two assumptions in which the probability distribution can be (ga, gb, gab): a, b are randomly selected and independently selected from Zq and (ga, gb, gc): a, b, c are randomly selected and independent of Zq are arbitrarily divided computer security $n = \log(q)$. Our proposed system operates over an elliptic curve in ECDSA such as group planning or DSA such as cyclic group duplication when using the Diffie–Hellman (DDH) decision design [5].

Ethereum is an open-source, public, blockchain-based shipping computing system and smart contract.

Operating system (script) blockchain polling can be done in different stages [9]. The most common categories are the following: launch, registration, ballot broadcast, validation, tallying results, and disclosure results.

Different programs use different methods by combining several sections together. The proposed system uses the following categories,

(1) Voter registration is the phase where voting provides personal information. Data is verified by authorized personnel and stored on the blockchain.

(2) Creating a private voter blockchain for voter registration information is a step toward the creation of a new blockchain and Merkle hash tree to store voter information. Merkle hash tree is also known as binary hash tree. The Merkle hask tree [10] is used to store a large collection of data to summarize correctly and to verify the validity of such a large dataset. The Merkle tree has the following features: (i) Ability to verify whether a transaction is entered into

blocks. (ii) Light customers. (ii) Total performance and fraud. (iv) Specified payment guarantee (SPV).

(3) Voter authentication is the step by which physical voter verification is performed based on the record stored on a private blockchain.

(4) The voting and billing phase is the last step where actual voting is done and the results are displayed and shown to the voter.

Blockchain Voting Features

Ready: This item states that users should be able to vote only if eligibility criteria is met. Eligibility criteria may vary from country to country. One of the years is the age of the user casting votes. In India, a person over the age of 18 years is allowed to vote.

Privacy: Privacy is one of the most important aspects of democracy. Voters' privacy should be maintained. No one should know how a particular person voted or how a particular voter voted.

Resistance to corruption: No person should be able to force a vote and must not have the power to distinguish between a voter and the voter who is instructed to vote.

Physical confirmation: Physical verification is required to know if the voter personally voted or a fraudulent person to vote on an identity document. Forgiveness: The voter should be able to alter the vote at any time before the election ends. This is related to the coercion resistance. Even if the coercer forced the voter, a voter should be able to change the vote afterward.

Verifiability: This property states that everyone involved in the voting process should be able to verify the results. This brings transparency in the election. Also, an individual voter should be able to verify whether his/her vote is counted or not.

Immutability: The voter's vote should be immutable. No one should be able to change the vote of any voter without proper concern of the voter. All the records should be immutable.

4 Implementation of Blockchain Voting

There are many platforms to implement blockchain voting, but the most used platform is Ethereum. It is an open-source, public, blockchain-based distributed computing platform and operating system featuring smart contract (scripting) functionality. When using Ethereum, computational expenses are manifested as gas price. An initial gas value is set, which can be used to perform the operations. Blockchain voting can be implemented in various phases. Most common phases are as follows: initialization, registration, ballot casting, verification, tallying results, and revealing results.

We will be concerned about the tokens gained through blockchain blocking in this way. Since our voting is based on blockchain, many features benefit from blockchain itself, such as endpoint validation, transparency, volatility, and so on. This aspect of blockchain solves some of the legal requirements for democratic voting (Fig. 1).

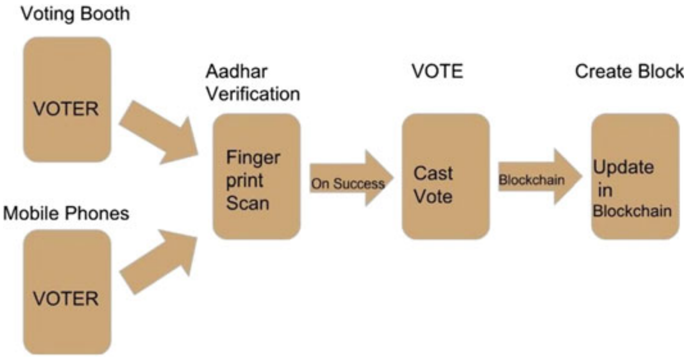


Fig. 1 Architecture diagram of the system

But with this, development comes with certain threats such as the convenience of anonymity has become increasingly difficult. Various research papers have proposed various methods to solve these problems. Those processes are as follows.

Shamir's secret sharing plan is implemented through a voter registration system [11]. The idea is based on the assumption that one would need k points to interpret a polynomial of degree $k - 1$. For example, two points should draw a line segment or three points should draw a curve. So, one can share the secret between person n in such a way that it can only be accessed by k ($k <= n$) people, and we need to hide that secret in the form of a $k - 1$ degree polynomial. This function keeps the privacy of every user even though data is visible to everyone. So, the issue of transparency and privacy is solved at the same time. Homomorphic encryption is another concept that can be used to provide anonymity. Homomorphic encryption is a encryption technology that can be implemented using the ElGamal cryptosystem exonential. The basic idea is to allow the authorities to vote by voting without having to spread that by providing privacy and security to the system. ElGamal's exponential is a program called cryptosystem that is used to encrypt voters' data and provides additional homomorphic system assets to the system.

Another way to use homomorphic encryption is to use Playfair encryption. Homomorphic encryption is nothing but computing in the hidden data that the compilation was originally made from the entered data. Creating a multiplicative form of multiplication in full homomorphic encoding is very strengthening and slow. Therefore, a smaller version of homomorphic encryption is used such as internal encryption. This adds up to two ciphertexts and is a duplicate collected with another clear database.

5 Comparison of Results

One-time signature techniques can also work to prevent anonymity. The unclassified signature is used at the end of the receiver. In this scheme, no observer can check

Table 1 Results of experimentation

	Polys [4]	Bronco vote [5]	Ranked choice voting [7]	Bit congress [8]	Follow my vote
Eligibility	Yes	Yes	Yes	no	Yes
Anonymity	Yes	Yes	Yes	Yes	Yes
Verifiability	Yes	Yes	Yes	Yes	Yes
Integrity	Yes	Yes	Yes	Yes	Yes
Physical verification	–	No	No	–	–
Forgiveness	–	–	Yes	Yes	Yes

whether the send is sent to a specific recipient address or two corresponding addresses. To maintain anonymity in the sender side, ring signature process is used. Here, the sender creates a unique ring structure with a fracture in it using a public key generated by other signatories and completes the ring using a private key (Table 1).

Although blockchain offers many things, there are limitations and drawbacks of blockchain. The first problem is that of a mass attack, which is nothing but if someone has more than 51% of their ability to use the available power, they can change the purchase data. Another issue related to blockchain problem fork [12]. When the system came to a new agreement or a new type, the blockchain network was divided into two nodes, new nodes and old locations, so after a change was made, the old nodes could not adapt to the new nodes, and then, there was a problem.

6 Conclusion

In this paper, he introduced the features of the online voting system and how the various programs met those criteria. Many programs have been able to deal with many aspects successfully. Some of the features that blockchain and remaining are solved by some encryption techniques such as homomorphic encryption. Although most systems have successfully used blockchain features, there are some online voting features that can be configured yet.

References

1. Dricot L, Pereira O (2018) SoK: Uncentralisable Ledgers and their Impact on Voting Systems
2. Kasireddy P How does ethereum work anyway [online]. Available https://medium.com/@preethikasireddy/how-does-ethereum-work-anyway-22d1df506369
3. Lin I-C, Liao T-C (2017) A survey of blockchain security issues and challenges. Int J Netw Secur 19(5):653–659

4. Yang X, Yi X, Nepal S, Kelarev A, Han F (2018) A secure verifiable ranked choice online voting system based on homomorphic encryption. https://doi.org/10.1109/ACCESS.2018.2817518
5. Polys—Online voting system, Whitepaper [online] Available https://polys.me/assets/docs/Polys_whitepaper.pdf
6. Panja S, Roy BK (2018) A secure end-to-end verifiable e-voting system using zero knowledge based blockchain
7. 1.Dagher GG, Marella PB, Milojkovic M, Mohler J (2018) BroncoVote: secure voting system using Ethereum's blockchain. In: ICISSP 2018—4th international conference on information systems security and privacy
8. Wanga B, Suna J, Hea Y, Panga D, Lua N Large-scale election based on blockchain. In: International conference on identification, information and knowledge in the internet of things, Available online at www.sciencedirect.com
9. Hardwick FS, Gioulis A, Akram RN, Markantonakis K (2018) E-voting with blockchain: an e-voting protocol with decentralisation and voter privacy
10. Andrew Blockchain fundamentals #1: What is a Merkle Tree? [online]. Available https://medium.com/byzantine-studio/blockchain-fundamentals-what-is-a-merkle-tree-d44c529391d7
11. Bartolucci S, Bernat P, Joseph D (2018) SHARVOT: secret SHARe-based VOTing on the blockchain. In: ACM/IEEE 1st international workshop on emerging trends in software engineering for blockchain
12. Ray S Blockchain forks [online]. Available https://hackernoon.com/blockchain-forks-b0dca84db0b0

Deepfake Detection Approaches Using Deep Learning: A Systematic Review

Anushree Deshmukh and Sunil B. Wankhade

Abstract Deepfake algorithms can make forged pictures and videos that people cannot differentiate them from true ones. The suggestion of technology that locate and prove the truth of virtual visual media is as a result essential. Deepfakes generates realistic forged images or videos of targeted persons by swapping their faces another person saying or doing things that are not really done by them and public start trusting in such forged videos, as it is not identifiable with the normal human eye. This paper offers a survey of tools and algorithms used to make deepfakes and, additional significantly, methods to locate deepfakes. We present huge discussions on challenges, studies, advances and strategies associated to deepfake. By reviewing the history of deepfakes and cutting-edge deepfake detection strategies, this gives a comprehensive assessment of deepfake techniques and helps the development of latest and more robust strategies to deal with an increasing number of tough deepfakes.

Keywords Deepfakes · Deep learning · Autoencoders · Generative adversarial networks (GAN)

1 Introduction

Deepfake (stemming from "deep learning" and "fake") is a technology which creates fake images or videos of targeted humans by swapping their faces another character saying or doing things that are not absolutely done by them and humans start believing in such fake, as it is not always recognizable with the everyday human eye. Deep learning models such as autoencoders and generative adversarial networks have been applied widely in the computer graphics vision to solve various issues [1, 2]. These

A. Deshmukh (✉)
Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, India
e-mail: anushree.deshmukh@mctrgit.ac.in

S. B. Wankhade
Department of Information Technology, Rajiv Gandhi Institute of Technology, Mumbai, India
e-mail: sunil.wankhade@mctrgit.ac.in

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*,
Lecture Notes in Networks and Systems 146,
https://doi.org/10.1007/978-981-15-7421-4_27

models have been used by deepfake algorithms to examine facial languages and actions of a person and synthesize facial images of another person making equivalent expressions and movements [3]. Deepfake algorithms usually require a large dataset to train models to make realistic images and videos. As public figures such as legislator or celebrities may have a large number of videos and images available online, they are probable targets of deepfake. Deepfakes were used to superimpose faces of legislator or celebrities to bodies in porn images and videos.

The first deepfake video found in 2017 where face of a celebrity was superimposed to that of a porn actor. It is threat to world security when deepfake procedures can be employed to make videos of world leaders with forged speeches for false purposes. Deepfakes therefore can be abused to cause political or religious misunderstanding between countries, to fool public and affect results in election campaigns, or create confusion in financial markets by creating fake news. It can be even used to create forged satellite broadcasting images of the Earth to hold items that do not really exist to create chaos in military**.**

There is also advantages of deepfakes such as creating voices of those who have lost theirs voice or updating episodes of Web series without reshooting them. However, the number of venomous uses of deepfakes largely dominates that of the positive ones. The growth of advanced deep learning networks and the accessibility of big amount of data have completed the forged images and videos almost unique to people and even to algorithms. The method of creating those forged images and videos is also much easy today as it wants as little as a self-photo or a small video of a targeted individual. Less effort is required to produce an impressively substantial tempered footage. Recent advances can even create a deepfake with just a motionless image. Deepfakes therefore can be a threat affecting not only community figures but also ordinary people.

Finding the fact in digital domain therefore has become gradually critical. It is even more interesting when dealing with deepfakes as they are majorly used to help venomous drives and almost anyone can create deepfakes these days used with present deepfake tools.

Therefore, here have been abundant methods proposed to detect deepfakes. Most of them are also based on deep learning, and thus,a battle between malicious and optimistic uses of deep learning methods has been rising.

In Sect. 2, we present the ideologies of deepfake algorithms and how deep learning has been used to enable such riotous technologies. Section 3 reviews different methods for detecting deepfakes as well as their advantages and disadvantages. We discuss challenges, investigation movements and directions on deepfake recognition as well as interactive program and forensics issues along with research gap in Sect. 4. Section 5 concludes the paper.

2 Deepfake Creation

Deepfakes have become popular due to the quality of tampered videos and the easy-to-use skill of their applications to a wide range of users with various computer skills from expert to beginner. These applications are typically developed based on deep learning practices. Deep learning is well known for its ability of instead of complicated and high dimensional data. The first attempt of deepfake creation was Fake App, developed by a Reddit user using autoencoder--decoder blending structure. In that method, the autoencoder extracts hidden features of face images and the decoder is used to reconstruct the face images. To swap faces between source images and target images, there is a need of two encoder decoder sets where each pair is used to train on an image set, and the encoder's parameters are joint sandwiched between two network sets.

This approach enables the common encoder to find and learn the comparison between two sets of face images, which are comparatively unchallenging because faces normally have similar structures such as eyes, nose, mouth places. Figure 1 demonstrations a deepfake making procedure where the feature set of face A is linked with the decoder B to rebuild face B from the original face A. Table 1 shows some tools for deepfake creation along with its algorithm used and key features.

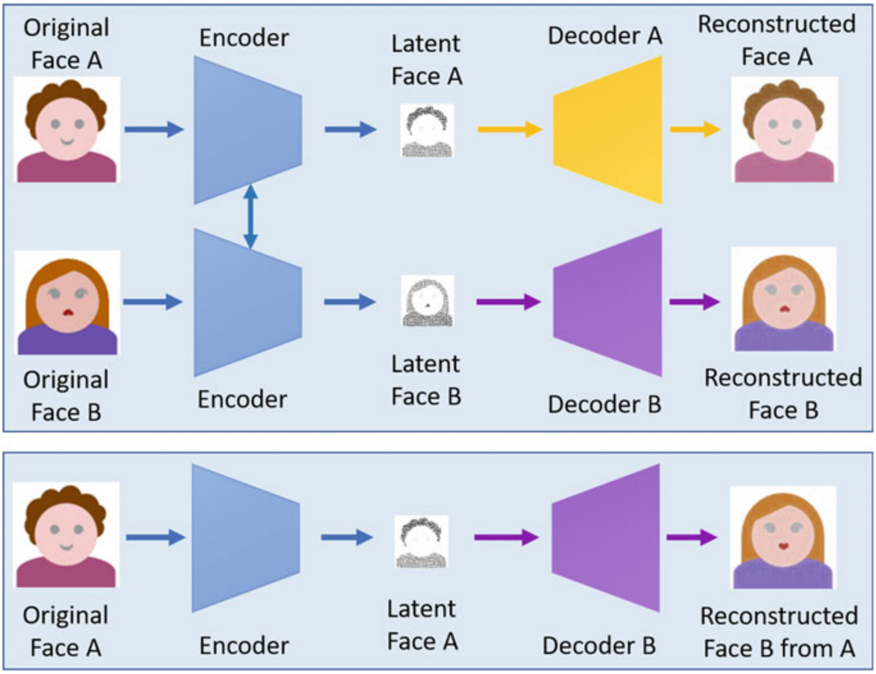


Fig. 1 A deepfake creation model

Table 1 Summary of notable deepfake tools

Tools	Algorithm used	Key features,
DFaker	GAN with DSSIM	Reconstruction of face is done by DSSIM loss function Keras library-based implementation
Faceswap-GAN	GAN v2.2 with MTCNN	The auto-encoder architecture is provided with adversarial loss and perceptual loss (VGGface) are added to
Faceswap	GAN with MTCNN and DFL-H128	Parameters of the encoder are shared and pairs of encoder--decoder
DeepFaceLab	GAN-er with MTS3FD	Expand from the Faceswap model with new models Uses multiple face
DeepFake-tf	GAN with DSSIM	Like DFaker but tensor-flow is used for implementation

This segment gives a review of deepfake detection methods where we categorize them into two major classes: fake image detection approaches and fake video detection approaches as shown in Fig. 2. The latter is distinguished further: visual artifacts within video frame-based methods and temporal features across frames-based ones. Whereas most of the approaches based on temporal features use deep learning recurrent classification models, the methods use visual artifacts within video frame can be executed by either deep or shallow classifiers.

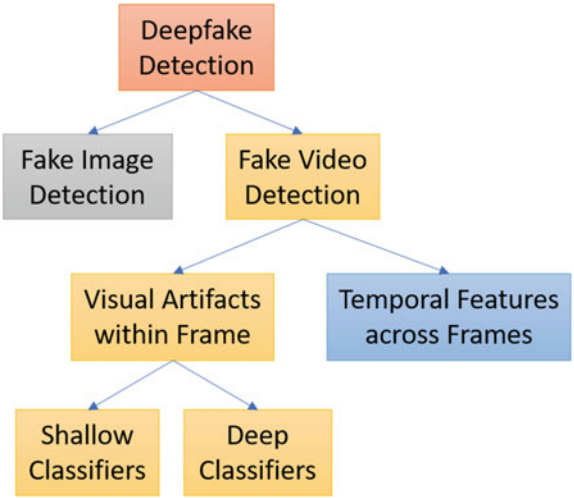


Fig. 2 Classification of deepfake detection

3 Deepfake Detection

Deepfake finding is normally deemed a binary arrangement problem where classifiers are used between reliable videos and interfered ones. This kind of method requires a large database of real and fake videos to train classification models. The number of fake videos is progressively available, but it is still limited in terms of setting a level for validating many discovery methods.

To report this matter, Korshunov and Marcel [4] produced a distinguished deepfake data set containing of 598 videos based on the GAN with the open source code Face swap-GAN. Videos from the openly available VidTIMIT database were used to produce low and high quality deepfake videos, which can efficiently emit the facial features, mouth actions, and eye blinking. These are then used to test various deepfakes. Test outcomes show that the general face recognition systems supported VGG and Facenet [4] are incapable to detect deepfakes successfully. Additional approaches such as lip-syncing and image quality metrics with SVM makes very high error rate when allied to detect deepfake videos from this newly shaped dataset. This increases fears about the serious need of upcoming development of more strong systems that can detect deepfakes from the original.

3.1 Fake Image Detection

Face swap has a few convincing applications in video compositing, transformation in portraits, and specially in individuality safety as it can swap faces in photos by ones from a group of typical images. However, it is one of the approaches that fake attackers enter authentication structures to gain access. The usage of deep learning such as CNN, GAN, SVM, random forest and multi-layer perceptron has exchanged face images more challenging for forensics as it can reserve pose, facial appearance and light of the photos. Among deep learning produced images, those produced by GAN are possibly most tough to notice as they are genuine and good quality based on GAN's ability to learn supply of the input data and gives novel outcomes with same input distribution.

Maximum works on recognition of GAN produced images do not consider the capability of the recognition models although the development of GAN is continuing, and several new extensions of GAN are often introduced. Xuan et al. [5] used an image preprocessing phase. This rises the pixel level statistical comparison between actual image and false image and needs the forensic classifier to study additional essential and meaningful features, which has better simplification competence than preceding image forensics methods or image stag analysis networks.

Additionally, Agarwal and Varshney [6] company the GAN-based deepfake recognition as a hypothesis challenging issue where a statistical outline was presented by means of the information theoretical study of verification. The least distance among deliveries of genuine images and images produced by a specific GAN is defined, explicitly the vision fault. The logical outcomes demonstrate that this distance rises when the GAN is fewer correct, and in this situation, it is easier to distinguish deepfakes. In case of high-resolution image, a very accurate GAN is required to produce fake images that remain tough to detect.

3.2 Fake Video Detection

Maximum image-based deepfake recognition approaches cannot be used for videos because of the robust deprivation of the frame data after audiovisual compression. Also, videos have chronological features that are diverse amongst sets of frames and thus challenging for approaches intended to sense individual fake images. This subcategory emphases on deepfake video recognition approaches and classifies them into two clusters: approaches that uses chronological features and those that explore visual artifacts inside frames.

3.2.1 Temporal Features Across Video Frames

Based on the opinion that temporal coherence is not imposed efficiently in the procedure of deepfakes, Sabir et al. [7] use of spatio-temporal features of audiovisual streams to sense deepfakes. Video manipulates to carry out on a frame-by-frame basis so that low level objects formed by face manipulations are supposed to further clear themselves as temporal artifacts with irregularities among frames.

A recurrent convolutional model (RCN) was projected based on the combination of the convolutional network Dense Net and the gated recurrent unit cells to exploit temporal inconsistencies across frames see Fig. 3. The proposed technique is verified on the Face Forensics++data set, which contains 1,000 videos, and displays promising outcomes.

Guera and Delp [8] detailed that deepfake videos comprise intra-frame conflicts and temporal conflicts among frames. Then they proposed the temporal-aware pipeline technique that uses CNN and long short term memory (LSTM) to detect deepfake videos. CNN is employed to extract frame-level features, which are then fed into the LSTM to create a temporal sequence descriptor. A fully connected network is

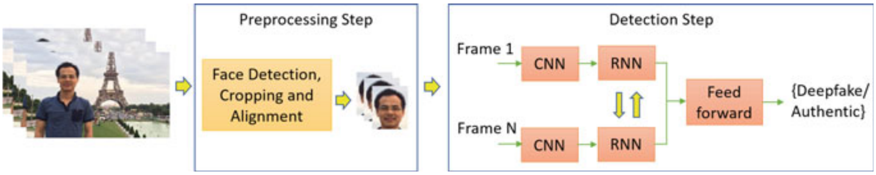


Fig. 3 A two-step process for face manipulation detection

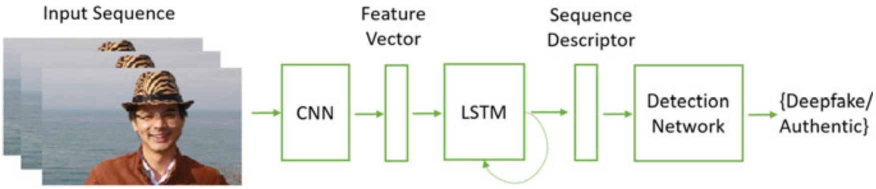


Fig. 4 A deepfake detection using CNN and LSTM

used afterward for classifying doctored videos from real ones based on the sequence descriptor (Fig. 4).

Similarly, the use of a physical signal, eye blinking, to notice deepfakes was planned in [9] based on the opinion that a individual in deepfakes has a lot fewer frequent blinking than that in original video. A fit adult human would usually blink somewhere between 2 and 10 s, and every open and close the eye would take 0.1–0.4 s. Deepfake procedures, however, frequently use face images existing online for training, which usually demonstrate persons with open eyes. Therefore, without blinking images of persons, deepfake algorithms do not have the ability to produce fake faces that can blink typically. To distinguish actual and false videos, Li et al. [9] initially decompose the videos into frames where face areas and then eye region are mined based on six eye signs. After some stages of pre-processing such as aligning faces, extraction and scaling the bounding boxes of eye sign points to generate new arrangements of frames, these cropped eye area arrangements are distributed into long-term recurrent convolutional networks (LRCN) for dynamic state prediction. The LRCN contains of a feature extractor built on CNN, a sequence learning built on long short-term memory (LSTM), and a state prediction built on a fully connected layer to forecast probability of eye open and close state. The eye blinking illustrates robust temporal dependencies and therefore the application of LSTM helps to capture these time-based patterns efficiently. The blinking rate is calculated built on the predicted outcomes where a blink is defined as a highest above the threshold of 0.5 with interval less than 7 frames. This technique is assessed on a dataset collected from the internet containing of 49 videos and their matching fake videos produced by the deepfake algorithms.

3.2.2 Visual Artifacts Within Video Frame

As observed in the preceding subsection, the approaches using temporal patterns across video frames are typically based on deep recurrent network models to sense deepfake videos. This subsection inspects the additional method that usually decomposes videos into frames and explores visual artifacts within single frames to get distinguished features. These features are then dispersed into either a deep or shallow classifier to discriminate between false and true videos. We thus assemble approaches in this subsection based on the categories of classifiers, i.e., either deep shallow.

Deep Classifiers

Deepfake videos are usually formed with few resolutions, which need a face warping method (i.e., rotation and shear) to matches the original ones. Because of the resolution contradiction between the warped face part and the nearby context, this procedure leaves objects that can be noticed by CNN methods such as VGG16, ResNet50, ResNet101 and ResNet152. A deep learning technique to spot deepfakes based on the objects observed throughout the face warping step of the deepfake generation algorithms was proposed in [10]. The proposed structure is assessed on two deepfake datasets, i.e., the UADFV and Deepfake TIMIT. The UADFV dataset [11] contains 49 actual videos and 49 forged videos with 32,752 frames in total. The Deepfake TIMIT dataset contains bad quality videos of size 64 * 64 and another good quality videos of 128 × 128 with 10,537 unique images and 34,023 forged images mined from 320 videos for each set. Performance of this technique is associated with other predominant approaches such as the face tampering recognition technique two-stream NN, head pose [11], and two deepfake recognition MesoNet methods, i.e., Meso-4 and MesoInception-4 [12]. Advantage of the proposed technique is that it need not to produce deepfake videos as forged examples before training the detection methods. As an alternative, the false examples are produced dynamically by mining the face area of the unique image and aligning it into several scales before applying Gaussian blur to a scaled image of chance choice and warping back to the unique image. This decreases a huge time and computational resources associated to other approaches, which need deepfakes are produced in advance. Recently, Nguyen et al. [13] proposed the usage of capsule networks for distinguishing doctored images and videos. The deepfake face swap dataset created by Afchar et al. [12].

4 Research Issues

Deepfakes are gradually damaging to confidentiality, society safety and democracy. Diverse approaches for detecting deepfakes are proposed by many researchers as soon as this menace was familiarized. Initial efforts were based on handcrafted features gained from objects and irregularities of the fake video synthesis procedure. Current methods, additionally, uses deep learning to automatically mine prominent and discriminative features to spot deepfakes. Using recognition approaches to spot deepfakes is critical but considering the actual intent of persons reproducing deepfakes is even more important. This needs the decision of users built on social context in which deepfake is revealed. This is serious as deepfakes are getting more photorealistic and it is extremely projected that recognition software will be lagging behind deepfake creation techniques.

4.1 Limitations of Existing System

In the technique where the recognition is done by eye blinking technique [9], the dataset used was very small thus system accuracy will be reduced. The experimental outcomes show great performance of the planned technique in detecting false videos, which can be further enhanced by seeing dynamic pattern of blinking. Detection is done by frame by frame basis so there might be chances of inconsistency [7]. RNN gives better result than CNN [8, 10, 12].

4.2 Effects of Deepfake

Misinformation: Folks are more probable to have a response to misinformation in the form of forged image, audio, and video content, which allows the doctored media to spread more rapidly than purely written fake data. Additional, images and video have been recommended to generate a Mandela effect, the formation of memories that never happened.

Exhaustion of serious thinking: It will take more effort for people to determine whether data is true, specially when it does not originate from reliable actors. Uncertainty around content reliability might also put off an distinct from sharing correct content, dropping the distribution of true information.

The deceiver's dividend: The presence of entirely artificial content offers an avenue for actors to deflect charges of impropriety based on footages and video, by appealing the source material has been faked.

These results are troubling and will be most persistent in the future, as deepfake excellence increases and social alertness lags.

5 Conclusion

Deepfakes have started to tear down trust of people in media contents as seeing them is no longer corresponding with believing in them. This is critical nowadays as the technologies for creating deepfakes are progressively approachable and social media platforms can spread those fake contents rapidly. This paper has revised the state-of-the-art methods and a summary of typical approaches are provided. It is noticeable that a fight between those who use progressive machine learning to create deepfakes with those who make effort to detect deepfakes is rising. Deepfakes' good quality has been increasing and the performance of discovery methods needs to be enhanced consequently. Detection methods are still in their early stage and various methods have been suggested and evaluated but using fragmented data sets. An approach to progress performance of detection methods is to create a rising updated standard data set of deepfakes to authorize the ongoing development of discovery methods. This will simplify the training process of discovery models, particularly those based on deep learning, which involves a large training set. The current detection methods mostly focus on disadvantages of the deepfake generation pipelines, i.e., finding weakness of the participants to attack them. This kind of information and knowledge is not always available in adversarial surroundings where attackers usually attempt not to reveal such deepfake creation technologies. This is a real challenge for detection method development and a future research needs to focus on introducing more robust, scalable and generalizable methods.

References

1. Tewari A, Zollhoefer M, Bernard F, Garrido P, Kim H, Perez P, Theobalt C (2018) High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2018.2876842
2. Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, Metaxas DN (2019) StackGAN++: Realistic image synthesis with stacked generative adversarial networks. IEEE Trans Pattern Anal Mach Intell 41(8):1947–1962
3. Lyu S (29 Aug 2018). Detecting deepfake videos in the blink of an eye. Retrieved from https://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072
4. Korshunov P, Marcel S (2019) Vulnerability assessment and detection of deepfake videos. In: The 12th IAPR international conference on biometrics (ICB), pp 1–6.
5. Xuan X, Peng B, Dong J, Wang W (2019) On the generalization of GAN image forensics. https://arXiv.com/1902.11153.
6. Agarwal S, Varshney LR (2019). Limits of deepfake detection: A robust estimation viewpoint. https://arXiv.com/1905.03493.
7. Sabir E, Cheng J, Jaiswal A, AbdAlmageed W, Masi I, Natarajan P (2019). Recurrent convolutional strategies for face manipulation detection in videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 80–87
8. Guera D, Delp EJ (2018, Nov) Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE international conference on advanced video and signal based surveillance, IEEE, AVSS, pp 1–6
9. Li Y, Chang MC, Lyu S (2018, Dec) In ictu oculi: exposing AI created fake videos by detecting eye blinking. In: 2018 IEEE international workshop on information forensics and security (WIFS), IEEE, pp 1–7
10. Li Y, Lyu S (2019). Exposing deepfake videos by detecting face warping artifacts. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 46–52
11. Yang X, Li Y, Lyu S (2019, May). Exposing deep fakes using inconsistent head poses. In: 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 8261–8265
12. Afchar D, Nozick V, Yamagishi J, Echizen I (2018, Dec) MesoNet: a compact facial video forgery detection network. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, pp 1–7
13. Nguyen HH, Yamagishi J, Echizen I (2019, May). Capsule-forensics: using capsule networks to detect forged images and videos. In: 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 2307–2311