Lecture Notes in Computational Vision and Biomechanics 25

Jude Hemanth
Valentina Emilia Balas *Editors*

Biologically Rationalized Computing Techniques For Image Processing Applications

Springer

Lecture Notes in Computational Vision and Biomechanics

Volume 25

The research related to the analysis of living structures (Biomechanics) has been a source of recent research in several distinct areas of science, for example, Mathematics, Mechanical Engineering, Physics, Informatics, Medicine and Sport. However, for its successful achievement, numerous research topics should be considered, such as image processing and analysis, geometric and numerical modelling, biomechanics, experimental analysis, mechanobiology and enhanced visualization, and their application to real cases must be developed and more investigation is needed. Additionally, enhanced hardware solutions and less invasive devices are demanded.

On the other hand, Image Analysis (Computational Vision) is used for the extraction of high level information from static images or dynamic image sequences. Examples of applications involving image analysis can be the study of motion of structures from image sequences, shape reconstruction from images, and medical diagnosis. As a multidisciplinary area, Computational Vision considers techniques and methods from other disciplines, such as Artificial Intelligence, Signal Processing, Mathematics, Physics and Informatics. Despite the many research projects in this area, more robust and efficient methods of Computational Imaging are still demanded in many application domains in Medicine, and their validation in real scenarios is matter of urgency.

These two important and predominant branches of Science are increasingly considered to be strongly connected and related. Hence, the main goal of the LNCV&B book series consists of the provision of a comprehensive forum for discussion on the current state-of-the-art in these fields by emphasizing their connection. The book series covers (but is not limited to):

- Applications of Computational Vision and Biomechanics
- Biometrics and Biomedical Pattern Analysis
- Cellular Imaging and Cellular Mechanics
- Clinical Biomechanics
- Computational Bioimaging and Visualization
- Computational Biology in Biomedical Imaging
- Development of Biomechanical Devices
- Device and Technique Development for Biomedical Imaging
- Digital Geometry Algorithms for Computational Vision and Visualization
- Experimental Biomechanics
- Gait & Posture Mechanics
- Multiscale Analysis in Biomechanics
- Neuromuscular Biomechanics
- Numerical Methods for Living Tissues
- Numerical Simulation
- Software Development on Computational Vision and Biomechanics
- Grid and High Performance Computing for Computational Vision and Biomechanics
- Image-based Geometric Modeling and Mesh Generation
- Image Processing and Analysis
- Image Processing and Visualization in Biofluids
- Image Understanding
- Material Models
- Mechanobiology
- Medical Image Analysis
- Molecular Mechanics
- Multi-Modal Image Systems
- Multiscale Biosensors in Biomedical Imaging
- Multiscale Devices and Biomems for Biomedical Imaging
- Musculoskeletal Biomechanics
- Sport Biomechanics
- Virtual Reality in Biomechanics
- Vision Systems

More information about this series at http://www.springer.com/series/8910

Jude Hemanth · Valentina Emilia Balas
Editors

Biologically Rationalized Computing Techniques For Image Processing Applications

Springer

Editors
Jude Hemanth
Karunya University
Coimbatore
India

Valentina Emilia Balas
Faculty of Engineering
Aurel Vlaicu University of Arad
Arad
Romania

ISSN 2212-9391 ISSN 2212-9413 (electronic)
Lecture Notes in Computational Vision and Biomechanics
ISBN 978-3-319-61315-4 ISBN 978-3-319-61316-1 (eBook)
DOI 10.1007/978-3-319-61316-1

Library of Congress Control Number: 2017945695

Printed on acid-free paper

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

Contents

Artificial Bee Colony Algorithm for Classification of Semi-urban LU/LC Features Using High-Resolution Satellite Data

J. Jayanth, T. Ashok Kumar, Shivaprakash Koliwad and V.S. Shalini

Abstract Attempts to classify high-resolution satellite data with conventional classifier show limited success since the traditional-per-pixel classifiers examine only the spectral variance ignoring the spatial distribution of the pixels corresponding to the land use/land cover classes. The work is carried out in two stages on panchromatic sharpened IRS P-6 LISS-IV (2.5 m) multispectral (MS) imagery of the year 2014 of Mangalore coastal zone along the west coast of Karnataka state of India. In the first stage, in order to overcome the limitations experienced in the parametric and nonparametric classifications, the swarm intelligence optimisation technique based on Artificial Bee Colony (ABC) algorithm has been studied for twelve land cover classes that are mapped. In the second stage, to bring out a greater separability between the spectrally overlapping classes, a texture-based image classification approach has been introduced and a methodology is developed to determine the optimal window size, interpixel distance and the best combinations of texture bands in multispectral data. The five texture measures, viz. entropy (ENT), angular second moment (ASM), contrast (CON), MEAN and homogeneity (Hmg) derived from the grey-level co-occurrence matrix (GLCM), are investigated

J. Jayanth (✉)
Department of Electronics and Communication Engineering,
GSSS Institute of Engineering & Technology for Women,
Mysore, Karnataka 570 016, India
e-mail: jayanthnov8@gmail.com

T. Ashok Kumar
PES Institute of Technology and Management, Shivamogga,
Karnataka 577 204, India
e-mail: ashokkumar0868@gmail.com

S. Koliwad
Department of Electronics & Communication Engineering,
Malnad College of Engineering, Hassan, Karnataka 573 202, India
e-mail: spksagar2006@gmail.com

V.S. Shalini
Department of Electronics and Communication Engineering,
ATME College of Engineering, Mysore, Karnataka 570 028, India
e-mail: shalinivs29@gmail.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_1

in the study. The major observations and contributions of this work are as follows: in the first stage, the image classifier employing the ABC algorithm exhibits higher classification accuracy when compared with maximum likelihood classifier. In the second stage, the results show that combining textural features and spectral bands in classification approach has proven very useful in delineating the spectrally overlapping classes, particularly at higher class hierarchy level.

Keywords Artificial Bee Colony · MLC · GLCM · Texture

1 Introduction

In recent years, remote sensing data classification became attractive due to its technical, economic and environmental benefits. Basically, this process is difficult because regions of a landscape are frequently not properly segregated which may lead to overlap in the specified region of interest. Precisely, for each pixel, grey-level values are allocated depending on the mediocre reflectance of different land cover classes on the given study area. Supervised, unsupervised and semi-supervised are the three popular learning techniques for land cover classification [1–3]. Classification of Remotely Sensed (RS) data has been carried out by several authors through computational artificial intelligence techniques such as fuzzy logic [4], neural network [5], support vector machine [6], K-means [7]. Metaheuristics have been also widely used for remote sensing data classification, for example particle swarm optimisation [8], ant colony optimisation [9], bee colony [10], artificial immune system [11] and genetic algorithm [12].

RS data classification for specified boundaries of classes with overlapping regions has been a challenging task for the research communities [13, 14]. This provides a reason to develop a classification model, which takes concern of this issue. Literature survey also indicates that texture information [5, 6, 8, 9, 15], surface temperature [6], digital elevation model (DEM) [12] have been studied as ancillary data in the RS data classification which is captured by different sensors. Among the several types of ancillary data, the grey-level co-occurrence matrix (GLCM)-based texture statistics derived from the image data are reported to be performing satisfactorily in RS classification [11]. Ultimate goal in all classification techniques is to best utilise the spectral, spatial and temporal resolutions and polarisation signature of the data and other inherent characteristics associated with it, and devise classification techniques which show improvements in accuracy, stability and speed. The motivation behind this work is to provide a faultless classifier, which can furnish accurate information about various land use/land cover (LU/LC) which is a prerequisite and indispensable issue in the efficient monitoring and management of natural resources and quick delivery of the end products. But it seems that the present traditional approach of classification fails to make the best use of the rich data obtained through the modern sensor and cripples in taking advantage of the capabilities of today's high-speed computational machines. In the above context, the current research work is primarily

aimed at developing an efficient and reliable classification strategy by integrating advanced image processing techniques through swarm intelligence to facilitate improving classification accuracy of high spatial resolution of RS data over semi-urban LU/LC features. In this regard, ABC algorithm is used to classify data and image texture analysis is used for characterising the spatial variations for extracting the information from the data.

2 Maximum Likelihood Classifiers

The maximum likelihood classifier (MLC) is one of the most commonly used algorithms for supervised classifications in RS data analysis [5–8]. Classification error is been reduced in MLC algorithm, assuming that the spectral data in the content has been distributed normally. Due to this impression, analyst chooses the most homogenous training sites by picking up many smaller training sites rather than few larger training sites. Due to this assumption, distribution of classes is been categorised depending on the pattern which is entirely outlined by the *MEAN vector* and the *covariance matrix* [6]. Hence, MLC needs a training data set that contains a relationship between the multivariate object properties and known classes. Jayanth J et al. [8] mentioned that MLC algorithm provides less accurate results for various land cover classes. MLC algorithm provides an accurate result when compared with parallelepiped classifier, but computational time is more [9]. However, one of the major drawbacks in MLC is the complexity involved due to multimodal distribution in the classes which may lead to poor classified results [7]. Hence, MLC method itself cannot solve the problem of spectral confusion in RS classification.

3 Artificial Bee Colony

In 2005, Dervis Karaboga defined Artificial Bee Colony (ABC) algorithm, which has been inspired by the intelligent behaviour of honey bees. Bees choose their food source in search space by flying in different aspects through waggle dance or randomly without using experience. When the bees are updated with their new nectar value, it is compared with old nectar content which is stored in the memory and updates with the highest nectar value. To manage the balance exploration and exploitation process, local search and global search methods are combined to validate and test the instance.

- Pixels (DN values) in the image are represented by bee.
- LC classes, such as pool, grass dry, are the food sources.

The four essential elements in the proposed algorithm are: initialising a set of features, fitness function, local search strategy and prediction strategy.

Classification procedure of ABC algorithm is shown in Fig. 1.

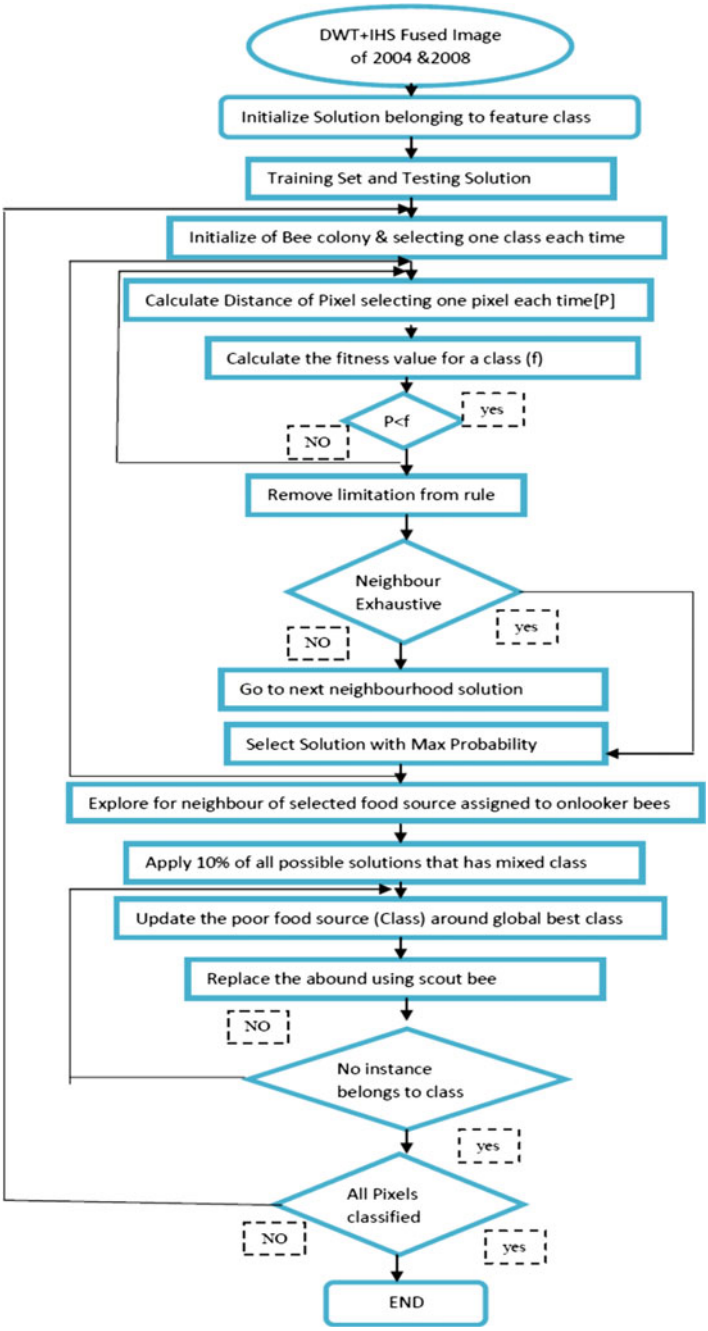


Fig. 1 Flow chart for Artificial Bee Colony algorithm classifier

3.1 Advantages of ABC

The principal advantage of ABC classifier over the traditional classifiers is as follows: bees being very optimal well-defined workers, they distribute the workload among themselves and their dancing behaviour helps in optimal design.

- ABC obeys proximity principle which responds to quality factor of the class.
- ABC can handle high-dimensional data and represents the acquired knowledge from waggle dance which is easy to intuit and assimilate.
- Simple and able to handle noisy data by creating a lower bound and upper bound for all the attributes.

3.2 Factors Affecting the Performance of the Artificial Bee Colony

Only the most commonly affecting issues are dealt with the following:

- Since bees do not depend on prior assumption about the data, it interrupts the performance of training data at the end of the list while assigning task to the agent, and may affect the performance of the classifier.
- ABC uses sequential covering algorithm, and hence, a small variation in data can make classification unstable and cause overfitting inside the data.
- Since each task has been selected by agents, the updating of agents from new task and its resource allocated to its agents can consume classification time.

4 Extraction of Textural Features

Texture represents the pixel relationship between spatial values and their grey-level properties which carries useful information for feature discrimination purpose in images. There is no single, unambiguous, widely accepted definition of texture in the literature. Analysis of texture in an image is referred to a mathematical model which shows differences in spatial characteristics of the image as a means of information extraction as shown in Table 1.

Since the first principal component (PC-1) contains the information which is common to all bands and accounts for maximum variance by removing the redundancy of information [16], the PC-1 image was extracted from the Linear Image Sensing Sensor (LISS)-IV data of 2.5 m spatial resolution to obtain GLCM and derive textural features thereof. All the textural features were extracted in

Table 1 Definitions of the texture measures based on GLCM

$\text{Entropy} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} p(i,j)\log_2[p(i,j)]$	$\text{Mean}: \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} ip(i,j)$
$\text{Energy (ASM)} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} [p(i,j)]^2$	$\text{Homogenity(IDM)}: \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{[P(i,j)]^2}{1+(i-j)^2}$
$\text{Contrast (Inertia)}: \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (i-j)^2 p(i,j)$	

ENT Entropy; *ASM* Angular second moment; *CON* Contrast; MEAN; *Hmg* Homogeneity

Geomatica V 10.0 at grey level (quantization levels) 16. We observed that the higher the grey level, the darker the image is and no marked difference is seen [7]. In the first phase of the work, five texture features (ENT, ASM, CON, MEAN and Hmg) [8] were computed from the GLCM at window sizes (W): 5 × 5, 7 × 7, 11 × 11, 17 × 17, 25 × 25, 35 × 35 and 51 × 51 at $\theta = 0°$ (assuming texture to be isotropic) keeping interpixel distance (D) unity. The above five features were also extracted at D: 3 and 7 keeping W: 25 × 25 for ENT, ASM, CON and Hmg, and 35 × 35 for MEAN.

Despite the fact that textural features exhibit non-Gaussian distribution of data, and Jeffries–Matusita distance performs better as a feature selection criterion for multivariate normal (Gaussian) classes [12], we employed the transformed divergence (TD) and Jeffries–Matusita (JM) distance measure as feature selection criterion for determining the best combination of multispectral bands and textural features. The JM distance varies between 0 and 1414, where 0 signifies no separability and the upper bound 1414 indicates high separability. Both the TD and JM distance measures are based on the measure of the statistical separation between pairs of class signatures [2], and the same has been adopted by [1–4].

5 Materials

5.1 Data Products Used

Indian Remote Sensor (IRS)-P6 LISS-IV satellite data captured on 16 April 2004 and 26 December 2007 consisting of three multispectral (MS, 5.8 m spatial resolution) bands recorded at green (0.52–0.59 μm), red (0.62–0.68 μm) and infrared (0.77–0.86 μm) wavelengths, and a panchromatic imagery of CARTOSAT-1 captured on 7 January 2008 (2.5 m spatial resolution); Census Data from Mangalore Urban Development Authority have been used in this study. All the satellite data are geo-referenced and projected with reference to global positioning system (GPS) readings.

Table 2 LU/LC classification hierarchy levels and details of the training and validation sites

Level-I	Level-II	Training sites		Validation sites
		MLC	Strata	
Built-up	RCC and asbestos sheet-roofing (RCC_sheet)	1032	11	239
	Clay tiled roofing (tiled_roof)	441	11	42
Roads	National highway (NH)	355	4	68
	Interior tar roads (int_rd)	490	5	147
Wastelands	Open ground/playgrounds (open_gnd)	1080	4	34
	Beach sand (sand)	270	2	21
Water	Pool water (pool)	93	1	18
	Sea water (sea)	884	2	19
Veg	Veg_mix[a] (veg_mix)	1848	6	178
	Acacia tree clusters (acacia)	1661	5	43
Grass dry	Grass dry (dry_grass)	424	3	66
	Total:	**8578**	**54**	**875**

[a]Thick leaf/fruit yielding trees such as mango, jack fruit, coconut, banyan

5.2 Study Area

The area under investigation is the Mangalore coastal region on the coastal belt of Karnataka state, India, between 12° 51′ 32″–12° 57′ 44″ N latitudes and 74° 51′ 30″–74° 48′01″ E longitudes. The image dimension of the study area is 1664 × 2065 pixels in MS data.

6 Results and Discussion

The objective of the behavioural study is to investigate the response of the Artificial Bee Colony (ABC) algorithm for classification on hierarchy level I and II over a semi-urban LU/LC area (coastal region of mangalore district) is shown in Table 2. The secondary objective of this analysis is to investigate the effectiveness of integrating GLCM-derived texture measures, viz. ENT, ASM, CON, MEAN and Hmg in classifying the panchromatic sharpened data of 2.5 m resolution at class hierarchy level-I and level-II, and to determine the optimal window size (W) and interpixel distance (D) for the texture measures under investigation.

6.1 Performance of ABC at Class Hierarchy Level-I and Level-II

The study was extended to compare the performance of both the classifiers at class hierarchy level-I and level-II. Hence, the training data sets, correspondingly the

Table 3 Classwise comparison of the accuracy of MLC and ABC at class hierarchy level-I

Class name	Producer's accuracy (%)		User's accuracy (%)	
	Classifier type			
	MLC	ABC	MLC	ABC
Grass_dry	72.18	73.68	93.20	91.59
Wasteland	91.11	88.89	96.47	95.24
Road	73.28	88.55	57.14	65.91
Built-up	70.05	82.35	77.51	77.39
Veg	94.58	81.28	87.67	92.70
Water	100	100	100	100
OCA (%)	81.42	83.44		

Table 4 Classwise comparison of the accuracy of MLC and ABC at hierarchy level-II

Class name	Producer's accuracy (%)		User's accuracy (%)	
	Classifier type			
	MLC	ABC	MLC	ABC
Grass_dry	75.94	69.92	90.99	92.08
Open_gnd	83.02	86.79	89.80	93.88
Rd_interior	83.70	83.70	60.63	62.10
Roof_RCC	85.00	83.75	79.07	83.75
Roof_Sheet	81.08	89.19	76.92	70.21
Roof_tiled	82.86	71.43	50.00	60.98
Sand	89.19	89.19	100	97.06
Veg_acacia	52.00	74.40	89.04	79.49
Veg-Mix	76.92	80.77	84.51	87.50
Water_Pool	100	100	93.75	100
Water_Sea	100	100	100	100
OCA (%)	77.88	80.03		

validation sets, were merged into 6 classes from 11. The classification results are tabulated in Tables 3 and 4.

At CHL-I, ABC shows an improvement of 3% OCA when compared with MLC algorithm (ABC 83.44% and MLC 81.42%). At class hierarchy level-I (CHL-I), ABC shows an improvement of 15.2% in producer's accuracy (PA) (88.55%) and 8.77% in user accuracy (UA) (65.91%) over MLC (PA 73.28% and UA 57.14%) for urban class *road*. Further, for class *built-up*, the ABC shows an improvement of 12.3% in PA over MLC. For the class wasteland, ABC shows 2.2 and 1.2% lesser PA and UA when compared with MLC. The class *veg* (vegetation) has experienced a trade-off in its UA and PA between both the classifiers. Hence, in both the class hierarchy levels, the ABC is able to maintain approximately 2% higher OCA (2.02 and 2.15% CHL-I and level-II, respectively) over MLC. It would also a

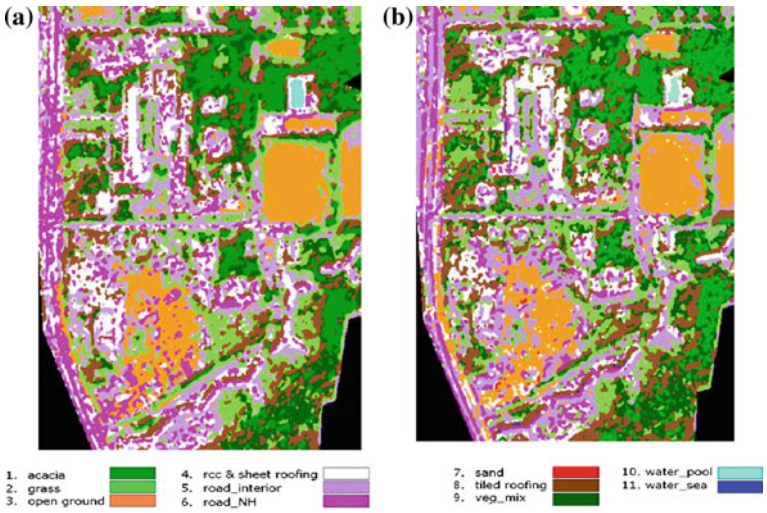


Fig. 2 Land cover classification in the area of Mangalore using **a** MLC and **b** ABC

noteworthy observation that as the class hierarchy level is increased from 6 classes to 11 classes, the OCAs of both the classifiers are found to be decreasing by an amount of 3–3.5% (Fig. 2).

Based on the classification results, ABC algorithm has been selected as the best candidate for classification and evaluation of textural features at class hierarchy level-I and level-II.

6.2 *Texture: Selection of Optimal Window Size and Interpixel Distance*

The aim of this work was to determine the optimal window size (W) and interpixel distance (D) for all the five textural measures. The extracted textural bands with five features (ENT, ASM, CON, MEAN and Hmg) at 9 different window sizes keeping $\theta = 0°$ were combined (stacked) separately with the three bands of the multispectral (MS) data. The supervised classification based on ABC algorithm, trained with 8578 locations, was carried out on the above data sets and one MS data set at CHL-II. Thereafter, a 3×3 majority filter was passed on all the classified images as a post-classification smoothing requirement. Finally, the accuracy assessment was carried out on the classified images using a validation data set comprising of 875 test points (details are in Table 2). The overall classification accuracy (OCA) and the kappa coefficients were computed and tabulated in Tables 5 and 6.

Table 5 OCA for five textural measures at seven different window sizes (W)

Window size (W):	5 × 5	7 × 7	11 × 11	17 × 17	25 × 25	35 × 35	51 × 51
Entropy (ENT)	64.11	65.83	67.89	75.09	78.17	77.6	76.34
Energy (ASM)	64.23	63.43	65.71	70.17	74.63	72.34	69.26
Contrast (CON)	66.29	67.43	67.66	71.43	76	76	73.71
MEAN	62.63	67.2	68.69	67.77	68.34	69.37	68.34
Homogeneity (Hmg)	64.46	65.14	67.87	70.51	72.69	71.09	72.69

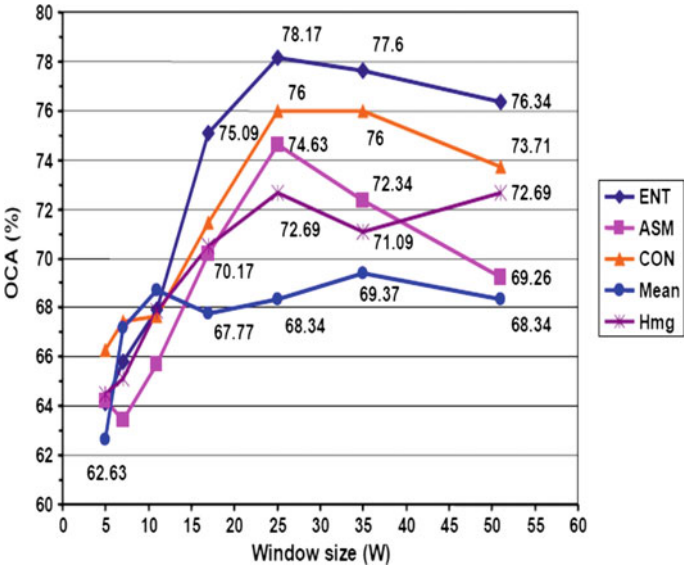


Fig. 3 Plot of OCA for five textural measures at seven different window sizes (W)

From Table 5; Fig. 3, the optimal window size for each of the five textural measures was determined from the band combination which provided the highest OCA, and the same is assumed to be the effect of the contribution of integrating the respective textural feature on MS band. Thereafter, the five textural features were extracted for D: 3 and 7 keeping W: 25 × 25 for features: ENT, ASM, CON and Hmg, and 35 × 35 for MEAN, and combined with MS bands creating an additional set of 10 data sets.

Further, Table 6; Fig. 4 show the classification accuracy obtained for the combinations of ENT (W: 25), ASM (W: 25), CON (W: 25), MEAN (W: 35) and Hmg (W: 25) at optimal window sizes (indicated within the brackets) at interpixel distance of 1, 3 and 7, keeping $\theta = 0°$. The difference in the magnitude of OCA is 1.71, 1.3, 0.46, 0.46 and 1.8% for ENT, ASM, CON, MEAN and Hmg, respectively, between D: 1 and 3. Only the texture measure Hmg shows some significant increase in OCA by an amount of 1.8% from D: 1 to 3. Also, Fig. 4 visualises that

Table 6 OCA for five textural measures at three different interpixel distances (D) at optimal window size (W)

Interpixel distance (D):	1	3	7
Entropy (ENT) (W: 25)	78.17	76.46	73.37
Energy (ASM) (W: 25)	74.63	73.26	70.97
Contrast (CON) (W: 25)	76	76.46	72.8
MEAN (W: 35)	69.37	69.83	69.03
Homogeneity (Hmg) (W: 25)	72.69	74.51	66.63

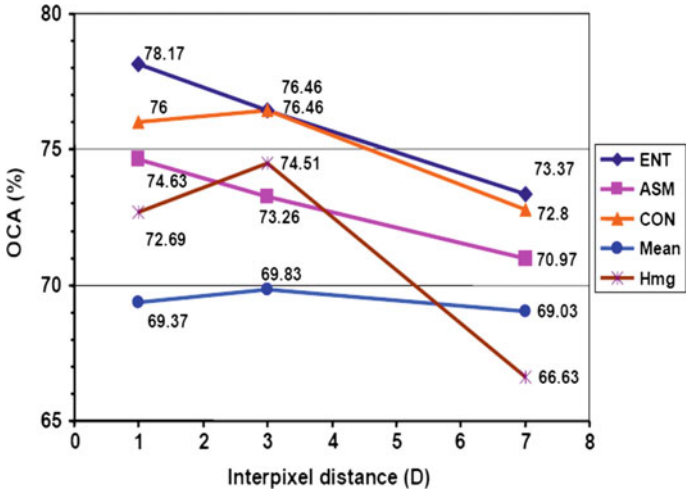


Fig. 4 Plot of OCA for five textural measures at three different interpixel distances (D) at optimal window size (W)

two of the features, viz. ENT, ASM, are moving downwards and the remaining three features, viz. CON, MEAN and Hmg, are moving upwards on OCA for change in D from 1 to 3. But among them, MEAN is not showing any significant difference in its OCA for variable D. Above D: 3, all the features exhibit a downfall in their performance. Therefore, an interpixel distance in the range of 1–3 exhibits satisfactory performance for all the texture features under investigation.

6.3 Effectiveness of Texture Feature Combinations

The objective of this experiment is to understand how a single and combinations of more than one GLCM texture measures perform with MS bands in MS data classification. The classification was carried out by employing MLC. The results of the classification using multiple texture features are provided in Table 7. A single

Table 7 Overall classification accuracy (OCA) and Kappa coefficient for data sets having combination of multispectral and texture bands at CHL-II

Sl. No.	Band ID No.	Band combinations	OCA (%)	Kappa statistics
1	1	MS	62.63	0.5721
2	2	1	78.17	0.7406
3	3	2	74.63	0.7007
7		1 + 2[a]		
8	7	1 + 3	76.57	0.7217
13		2 + 5[a]		
17		1 + 2 + 3[a]	75.09	
20	15	1 + 3 + 4	73.83	0.6867
21	16	1 + 3 + 5	76	0.7144
22	17	1 + 4 + 5	72.57	0.677
23	18	2 + 3 + 4	73.26	0.6802

MS Multispectral band, *1* Entropy, *2* ASM, *3* Contrast, *4* MEAN, *5* Homogeneity, [a]Band combinations excluded from further analysis as the combinations resulted in total misclassification of one or more classes, *ID No* Identification number provided only to the band combinations which resulted in valid classified images

texture at a time is able to provide a complete all-class classified image and hence treated as a valid classification. Still, the combinations of two textures together have shown a satisfactory performance by maintaining accuracy level approximately higher than 75% (e.g. ENT+CON, MEAN+Hmg, ASM+CON, CON+MEAN, ENT +Hmg and ENT+MEAN band combinations in Table 7). It is interesting to note that bands MEAN and Hmg alone have produced an OCA of 69.37 and 74.51%, respectively; however, when they are combined together, the OCA is increased to just 75.31%. Contrary to this observation, bands ENT and CON have produced an OCA of 78.17 and 76.46%, respectively, when used separately, but when combined together (MS+ENT+CON), the accuracy still stands at 76.57% leaving no hope of any reasonable improvement in accuracy for combination of two textures. Instead, the combination has degraded the single texture accuracy of the ENT from 78.17 to 76.57% with no appreciable increase in accuracy from 76.46 to 76.57% in respect of CON. Likewise, a simultaneous combination of three, four and five textures in MS has merged a large number of classes together, leaving as small as only two classes in the classified images as illustrated in Fig. 5. Thus, greater number of texture bands in MS has made the classification unreliable and produced invalid classified images. Therefore, it is difficult to generalise which band combination of two or more than two texture features together is most efficient in classification. It is also apparent from the above that particularly in the present classifier MLC, an increase in the number of texture features need not necessarily contribute to the improvement in accuracy; instead, such combinations degrade the classifier's performance.

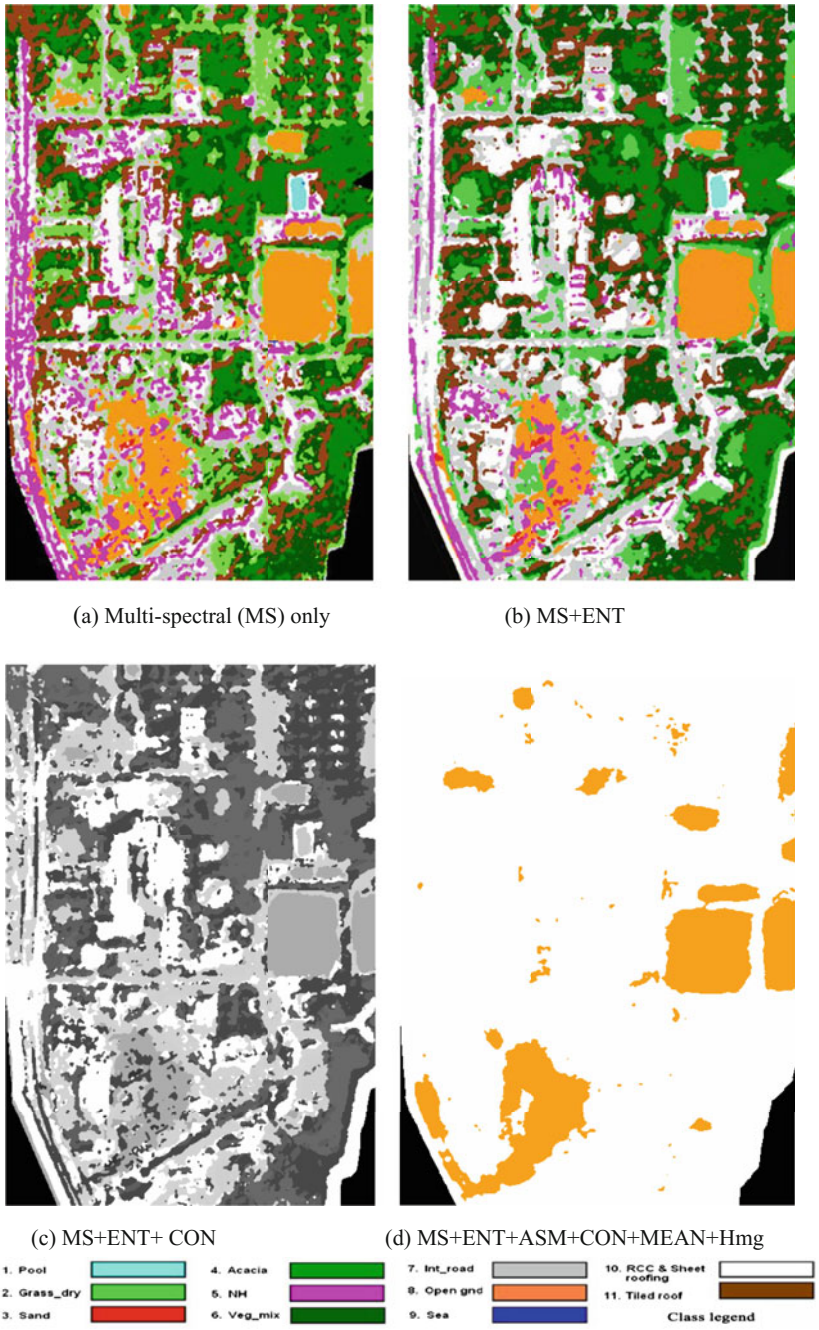


Fig. 5 Subset of classified images at CHL-II

Nevertheless, the results indicate that incorporation of texture measures into classification improves the overall classification accuracy. The discussion would be more complete if a quantitative study is made on the confusion matrices to understand the effect of the texture features on individual LU/LC classes. Hence, only two confusion matrices corresponding to MS band and MS+ENT bands are presented here in Tables 8 and 9, respectively, as the latter provides the highest accuracy in the whole study carried under CHL-II. The confusion matrices corresponding to Fig. 5c and d are not able to classify due to misclassification by the ABC classifier. So in the following discussion, the MS band and MS+ENT band combinations are referred to as case-1 and case-2, respectively.

The PA and UA for classes *pool water* and *sea water* are 100% in both the cases (Tables 8 and 9) since they are spectrally distinct and spatially homogeneous. The next two spatially homogeneous classes are *sand* and *open_gnd*. The PA for *sand* remains the same (71.43%) in both the cases, but the UA attained 100% in case-2 from the 88.24% of case-1. It is obvious because classes such as *pool water*, *sea water* and *sand* exhibit the highest TD of 2000 which indicates total separability when paired with other classes. However, even though the class *open_gnd* is spatially homogeneous, it is not spectrally distinct. The TD between the *open_gnd* and *rd_interior* is not satisfactory (1897.97). Hence, 10 validation points of class *rd_interior*, 4 each of *sand* and *RCC_sheet,* have been misclassified as *open_gnd* and reduced the UA of *open_gnd* to 64.71% in case-1, whereas these misclassifications are markedly lessened in case-2 and a UA of 88.57% is achieved for *open_gnd*. This also results in an improvement in Kappa value from 0.63 to 0.881 in comparison with case-1.

The class *roof_tiled* represents individual houses of approximately 15 m × 15 m in area located amidst of thick vegetation (*veg_mix*) along the sides of *interior road*. The major source of misclassification of *roof_tiled*, as seen in case-1, is from *veg mix* and *RCC_sheet*. In case-2, texture could make the UA and Kappa see a small improvement of 3.69% and 0.039, respectively, over MS band, but no difference in PA (92.86%). Hence, texture did not make a notable improvement on class *roof_tiled* which is distributed as relatively small individual entities over the study area. For class *RCC_sheet* (*RCC* and *asbestos sheet-roofing*), the most of the misclassification has come from the *interior road*, and a major portion of the validation sites has also been misclassified to *rd-NH, rd_interior road* and *roof_tiled* in MS data. But inclusion of feature ENT has significantly reduced the misclassification of *RCC_sheet* to *rd_NH*, *grass_dry* and *roof_tiled*, resulting in an improvement in PA from 53.14 to 76.99% and Kappa value from 0.64 to 0.724 when compared with MS band. This in turn has greatly benefited *rd_NH*, and hence, class *NH* has experienced an improvement in its UA from 48.33 to 82.35% and Kappa from 0.43 to 0.808 in MS+ENT band. The above analysis draws attention to the fact that texture has no effect on spatially homogeneous and spectrally distinct classes such as water bodies and sand. However, texture has shown an appreciable performance on spatially homogeneous but spectrally overlapping classes such as *open ground*, *RCC*, *NH* and *interior roads*.

Table 8 Confusion matrix and conditional kappa values of the classification results obtained for MS band alone for 11 classes

classes	1	2	3	4	5	6	7	8	9	10	11	Row total	UA%
1	**18**											18	100
2		**55**					12			7		74	74.32
3			**15**							2		17	88.24
4				**38**		101						139	27.34
5					**58**		9			53		120	48.33
6		5		4		**62**				1	2	74	83.78
7		4			3		**84**	1		32	1	125	67.20
8			4				10	**33**		4		51	64.71
9									**19**			19	100
10		2	2		7	1	32			**127**		171	74.27
11				1		14				13	**39**	67	58.21
Column total	18	66	21	43	68	178	147	34	19	239	42	**875**	
PA%	100	83.33	71.43	88.37	85.29	34.83	57.14	97.06	100	53.14	92.86		**OCA**
													62.63
kappa	1.0	0.72	0.87	0.23	0.43	0.79	0.60	0.63	1.0	0.64	0.56		

Class legend: *1* pool; *2* grass_dry; *3* sand; *4* acacia; *5* NH; *6* veg_mix; *7* inl_rd; *8* open_gnd; *9* sea; *10* rcc_sheet; *11* tiled_roof

Table 9 Confusion matrix and conditional kappa values of the classification results obtained for MS+ENT (W: 25 × 25, D = 1) bands for 11 classes

classes	1	2	3	4	5	6	7	8	9	10	11	Row total	UA%
1	**18**											18	100
2		**56**				2	3	1		2		64	87.50
3			**15**									15	100
4				**26**		28						54	48.15
5		1			**56**		1			10		68	82.35
6		3		17		**132**				2	2	156	84.62
7		5			6		**108**	1		32	1	153	70.59
8			4				10	**31**				35	88.57
9									**19**			19	100
10		1	2		6	1	35	1		**184**		230	80.00
11				1		15				9	**39**	63	61.90
Column total	18	66	21	43	68	178	147	34	19	239	42	**875**	
PA%	100	84.85	71.43	60.47	82.35	74.16	73.47	91.18	100	76.99	92.86		**OCA**
													78.17
kappa	1.0	0.864	1.0	0.454	0.808	0.806	0.646	0.881	1.0	0.724	0.599		

Class legend: *1* pool; *2* grass_dry; *3* sand; *4* acacia; *5* NH; *6* veg_mix; *7* int_rd; *8* open_gnd; *9* sea; *10* rcc_sheet; *11* tiled_roof

Another major spectrally overlapping classes are the mixed vegetation (*veg_mix*) and acacia forest (*veg_acacia*), which exhibit the lowest TD measure of 434.44. The confusion matrix corresponding to MS shows that out of the given 178 validation sites to *veg mix*, 101 have been misclassified as *veg_acacia*, 14 as *roof_tiled* and 1 site to *RCC_sheet*. However, there is no serious misclassification from other classes. Hence, even though it has produced the lowest PA of 34.83%, it could maintain a reasonably high UA of 83.78%. On the contrary, the same MS band has produced a relatively higher PA of 88.37% for *acacia*, but has made a fall in UA to a very low value of 27.34%, since 101 sites of *veg_mix* are misclassified as *veg_acacia*. A trade-off is seen between the PA and UA of these two classes. But the addition of texture band on MS could raise the PA of *veg_mix* from 34.83% to 74.16% with no appreciable improvement in the UA and Kappa. Contrary to this, texture has made a greater contribution to the UA (27.34–48.15%) of class *acacia* at the cost of a fall in PA from 88.37 to 60.47%, but without losing in Kappa. The Kappa, in turn, is improved from 0.23 to 0.454, approximately double that of the MS band. Finally, the class *grass_dry* is also benefited by texture as it could record an increase in Kappa from 0.72 to 0.864 in comparison with case-1. Hence, it is evident from above that addition of texture in MS band has the potential to improve the accuracy even for those classes which are spectrally overlapping and spatially non-homogeneous. For all the 11 classes, texture has maintained higher Kappa value over the MS band.

Finally, it is quite evident from the two classified images shown in Fig. 5b, c that the data set having texture feature entropy (ENT) integrated with MS is fairly free from the *salt-and-pepper* noise, which is very common and annoying in classified images (Fig. 5a). This improvement is also seen in other classified images too which employ the texture and MS bands together. Besides, since linear structures such as roads and buildings are seen more distinct, the shape information is found to be greatly pronounced in a texture-embedded classified image.

6.4 Investigation of Texture at Class Hierarchy Level-I and Level-II

The purpose of this phase of investigation is to examine the effect of the texture features at classification hierarchy level-I, in comparison with level-II. Results of the work have been tabulated in Table 10. The corresponding plot is also available in Fig. 6. Here, 6 data sets having texture measures with optimal W and D in combination with MS bands, but one feature at a time, were investigated. At class hierarchy level-I, the OCA obtained for all data sets is remarkably higher than level-II. It is also obvious from Fig. 6 that the MS band has gained an improvement of 18.17% in OCA (from 62.63 to 80.80%) when CHL is decreased from level-II to level-I without any texture integrated in it. The incorporation of texture at CHL-I

Table 10 OCA (%) and Kappa for various combinations of spectral and texture measures one at a time at CHL-I and CHL-II (6 and 11 classes, respectively)

Band combinations:		MS	ENT	ASM	CON	MEAN	Hmg
Class hierarchy level-I	OCA %	80.80	8.23	81.37	87.54	84.69	85.49
	Kappa	0.749	0.791	0.753	0.835	0.798	0.808
Class hierarchy level-II	OCA %	62.63	78.17	74.63	76.46	69.83	74.51
	Kappa	0.572	0.740	0.700	0.720	0.650	0.699

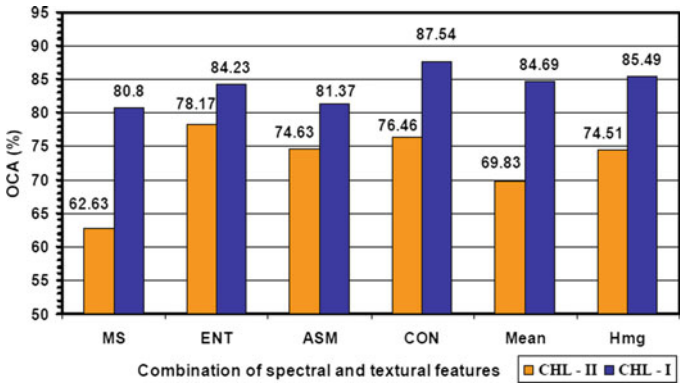


Fig. 6 Plot of OCA versus various combinations of spectral and texture measures one at a time at class hierarchy level-I and level-II (6 and 11 classes, respectively)

has improved the OCA in the range of 0.57% (with ASM: 81.37%) to 6.74% (with CON: 87.54%) in comparison with MS band of CHL-I (80.80%). Whereas, even though, the OCA (62.63%) is small for MS band at CHL-II, the difference in improvement seen upon adding texture ranges from 6.74% (with MEAN: 69.83%) to as high as 15.54% (with ENT: 78.17%). Hence, texture is found to be very effective in improving the classification accuracy at classification hierarchy level-II than level-I.

In summary, it is seen that texture feature ENT (W: 25 × 25, D: 1) not only produced the highest accuracy of 78.17% in class hierarchy level-II, but also maintained higher accuracy over its counterparts at all window sizes greater than W: 11 × 11 and at all interpixel distances. In contrast to the above, it is apparent from the plot in Fig. 6 that in CHL-I, the highest accuracy of 87.54% is obtained for the texture feature CON at W: 25 × 25, D: 1. The ENT and CON have also occupied the first and the second positions, respectively, in performance at CHL-II when investigated one texture at a time with MS. But the combination MS+ENT +CON, which is expected to perform better than ENT, stood second in accuracy (OCA: 76.57%) among all the band combinations with MS. Meanwhile, in CHL-II, the lowest OCA (69.37%) is recorded for texture band MEAN at W: 35 × 35, D: 1, but shows a steady performance for a wide range of W and D. Hence, it is difficult to ascertain which combination of the texture features of more than one band is

important in improving the overall classification accuracy. In the course of the analysis, one of the principal findings is that the urban or semi-urban LU/LC classes such as *RCC* and *sheet-roofing*, *roads* and *open grounds* have been substantially benefited by integrating texture features in the classification of multispectral bands.

7 Conclusion

The study demonstrated the effect of integration of spectral and textural features in classification of semi-urban Lu/LC features on LISS-IV data of 2.5 m spatial resolution at CHL-I and CHL-II. Textural measures, viz. entropy, energy, contrast, MEAN and homogeneity, were investigated with multispectral bands and various texture feature combinations. The following conclusions were drawn from the present study:

1. This work has presented a new method for classifying RS data using Artificial Bee Colony algorithm. As a result, ABC algorithm has provided high-quality result on the study area. OCA for ABC method shows an improvement of 3% in comparison with MLC.
2. Combining textural and spectral features in high-resolution satellite data has proven to be effective in improving the classification accuracy of spectrally overlapping but spatially homogenous urban land use/land cover features such as RCC roofing, asbestos sheet-roofing, interior road, NH and open grounds (playground). Texture combinations have not made marked difference over spectrally distinct and spatially homogeneous classes such as sand and water bodies in comparison with multispectral bands. Integration of texture greatly reduced the 'salt-and-pepper' noise in the classified images.
3. Literature survey indicates that there is no unique textural window size describing a land cover class as it depends on the spatial resolution of the data, the land cover spread which varies from one geographical area to another, and human cultural habitation; hence, it is solely problem dependent [16–18].

Acknowledgements The authors graciously thank Dr. Dwarakish, Associate Professor, NITK, Surathkal; DR. K.S. Shreedhara, UBDTCE, Davangere; and the staff at KSRSAC, Mysore, for their assistance and help rendered during field work.

References

1. Hala E, Mohamed NH (2012) Mapping potential landfill sites for North Sinai cities using spatial multicriteria evaluation. Egypt J Remote Sens Space Sci 15(2):143–150
2. Du P, Tan K, Xing X (2012) A novel binary tree support vector machine for hyperspectral remote sensing image classification. Optics Commun 285:3054–3060. doi:10.1016/j.optcom.2012.02.092

3. Zheng J, Cui Z, Liu A, Jia Y (2008) A K-means remote sensing image classification method based on AdaBoost. In: fourth international conference on natural computation ICNC '08, vol 4, pp 27–32. doi:10.1109/ICNC.2008.903
4. Yang H, Du Q, Chen G (2012) Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. IEEE J Sel Top Appl Earth Observations Remote Sen 5(2):544–554. doi:10.1109/JSTARS.2012.2185822
5. Liu X, Li X, Liu L, He J, Ai B (2008) An innovative method to classify remote-sensing images using ant colony optimization. In: IEEE transactions on geoscience and remote Sensing 46(12). doi:10.1109/TGRS.2008.2001754
6. Jayanth J, Koliwad S, Ashok Kumar T (2015) Classification of remote sensed data using Artificial Bee Colony algorithm. Egypt J Remote Sens Space Sci 18(1):119–126. doi:10.1016/j.ejrs.2015.03.001
7. Jayanth J, Ashok Kumar T, Koliwad S, Krishnashastry S (2015) Artificial bee colony algorithm for classification of remote sensed data. In: international conference on industrial instrumentation and control (ICIC)
8. Jayanth J, Koliwad S, Ashok Kumar T (2015) Identification of land cover changes in the coastal area of Dakshina Kannada district, south India, during the year 2004–2008. Egypt J Remote Sens Space Sci 19(1):119–126. doi:10.1016/j.ejrs.2015.03.001
9. Jayanth J, Ashok Kumar T, Koliwad S (2012) Comparative analysis of Image fusion techniques for remote sensing. In: international conference on advanced machine learning technologies and applications (AMLTA 2012) Cairo, Egypt, December 8–10. Proceedings of communications in computer and information Science. In: Hassanien AE, Salem AM, Ramadan R, Kim T (eds) (2012) vol 322 Springer, Berlin, pp 111–117
10. Kumar U, Dasgupta A, Mukhopadhyay C (2012) Sequential maximum a posterior (SMAP) algorithm for classification of urban area using multi-resolution spatial data with derived geographical layers. In: Proceedings of the India conference on geo-spatial technologies and applications, department of computer science and engineering, Indian Institute of Technology Bombay (IITB), 12–13 April 2012
11. Chang Che-Wei, Ho Chien-Chang, Chen Jyh-Horng (2012) ADHD classification by a texture analysis of anatomical brain MRI data. Frontier Sys Neuro Sci 12:123–129
12. Choodarathnakara AL, Ashok Kumar T, Koliwad S, Patil CG (2012) Soft classification techniques for RS data. IJCSET 2(11):1468–1471
13. Srikrishna Shastri C, Ashok Kumar T, Koliwad S (2016) Advances in classification techniques for semi urban land features using high resolution satellite data. Int J Adv Remote Sen GIS 5(3):1639–1648
14. Ashok Kumar T, Koliwad S, Dwarakish GS (2008) Effectiveness of decision tree algorithm in land use/land cover classification of NITK, Surathkal campus using high resolution satellite data. Int J Earth Sci Engg 18(1):56–65
15. Ashok Kumar T, Koliwad S, Dwarakish GS (2008) Classification of high-resolution satellite imagery: a non-parametric approach. National Institute of Technology Karnataka Research Bulletin, vol 17 no 1, pp 21–33
16. Ashok Kumar T, Koliwad S, Dwarakish GS, Patil CG (2008) Non-parametric classifier: an integrated approach with texture for land-cover classification of pan-sharpened LISS-IV imagery. In: Souvenir cum abstracts, Indian national cartographic association (INCA)-XXVIII international congress on collaborative mapping & space technologies Ahmedabad, Gujarath, India, November 4–6
17. Tuceryan M, Anil Jain K (1998) Texture Analysis. The handbook of pattern recognition and computer vision. World Scientific Publishing Co, pp 207–248
18. Ashok kumar T (2013) Advanced Image processing Techniques for Land feature Classification. Lambert Academic Publishing, pp 25–86

Saliency-Based Image Compression Using Walsh–Hadamard Transform (WHT)

A. Diana Andrushia and R. Thangarjan

Abstract Owing to the development of multimedia technology, it is mandatory to perform image compression, while transferring an image from one end to another. The proposed method directly highlights the salient region in WHT domain, which results in the saliency map with lesser computation. The WHT-based saliency map is directly used to guide the image compression. Initially, the important and less important regions are identified using WHT-based visual saliency model. It significantly reduces the entropy and also reserves perceptual fidelity. The main aim of the proposed method is to produce the high-quality compressed images with lesser computational effort and thereby achieving high compression ratio. Due to the simplicity and high speed of WHT, the proposed visual saliency-based image compression method is producing reliable results, in terms of peak signal-to-noise ratio (PSNR), compression ratio, and structural similarity (SSIM), compared to the state-of-the-art methods.

Keywords Saliency detection · Image compression · Walsh–Hadamard transform · PSNR

1 Introduction

Salient region detection drastically attracts the attention toward many of the computer vision and pattern recognition tasks such as image compression, object recognition, content-based image retrieval, image collection browsing, image editing, visual tracking, and human–robot interaction. It aims to detect the salient

A. Diana Andrushia (✉)
Department of ECE, Karunya Universtiy, Coimbatore, Tamil Nadu, India
e-mail: andrushia@gmail.com

R. Thangarjan
Department of CSE, Kongu Engineering College, Erode, Tamil Nadu, India
e-mail: thangs_68@yahoo.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_2

region of an image under biological plausibility. In this paper, Walsh–Hadamard transform (WHT)-based visual saliency detection is used for the image compression application.

Demand on data compression is increasing rapidly as the modern technologies are growing high. High storage capacity is required for uncompressed images. Many images are having the common characteristics as their neighboring pixels are highly correlated with redundant information [1]. Image compression-based techniques are aiming to reduce the redundant information by eliminating the spectral and spatial redundancies. It results in the reduction of consumption of expensive resources in the form of transmission bandwidth and hard disk space. Image compression techniques are generally classified into two types. One is spatial coding and another one is transform coding. In the transform coding type, discrete cosine transform (DCT), discrete Fourier transform (DFT), Walsh–Hadamard transform (WHT), etc., are used to perform natural image compression. Each transform is having its own advantages and disadvantages in the compression domain. Transform type coding is used in the proposed method.

WHT-based visual saliency detection and WHT-based image compression based on saliency map are the two main phases of this proposed method. The proposed WHT-based visual saliency detection is transform domain approach. So the frequency domain approaches are only considered for the performance comparison. As the WHT-based compression also transform domain approach so frequency domain methods only considered for fair performance comparisons in the compression phase.

The proposed work is experimented through MIT dataset. It is one of the benchmark datasets for visual saliency detection which consists of the indoor, outdoor images. The performance metric of receiver operating characteristics (ROC), area under the curve (AUC), precision, recall, and F-measure is obtained to analyze the proposed visual saliency detection. Peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and compression ratio are obtained for saliency-based image compression. The performance metrics are yielding significant results while comparing with state-of-the-art methods.

WHT is used in the saliency detection as well as image compression. WHT is chosen because the number of computations in this transform is significantly less compared to the other transforms, and WHT is the key transform to provide energy compactness. Finding the salient information in image/audio/video will reduce the number of computations and lesser hardware in the compression techniques. The proposed visual saliency-based compression method achieves 90% compression ratio. It is due to the lesser computations of WHT transform. This is the most added advantage of the proposed method compared to the state-of-the-art methods.

The reminder part of this chapter constructed as follows: Sect. 2 clearly explains about the backgrounds of visual saliency detection, saliency-based image compression. Section 3 elaborates the proposed methodology. Section 4 explains the experimental results with the performance metrics. Finally, the conclusion of the chapter is given in the last section.

2 Backgrounds

2.1 Saliency Detection

The world is full of visual data. Humans selectively perceive the visual information that is getting in through their eyes. Visual attention is the process of selecting particular information from the plenty of raw data perceived. For example, while sincerely watching a cricket match in the play ground, a sudden change in the action of the umpire with red shirt picks the attention of the spectators in the gallery despite the colossal load of visual inputs such as actions of the batsmen, bowlers, and fielders is there. The eyes of the spectators gaze the umpire momentarily before shifting to other events in the visual scene. The phenomenon of drawing the focus of attention to certain regions in a given scene or image is called visual attention [2]. In the jargon of computer vision, these regions are known as *salient* regions.

Detection of salient regions finds application in a wide spectrum of processes such as automatic object detection, image retrieval, remote sensing, automatic target detection, image and video segmentation, robotics, scene understanding, computer–human interaction, driver assistance, action recognition, background subtraction, image and video compression, video summarization, medical imaging, and automatic cropping. The cognitive process that directs human to select highly relevant data from a scene/image is named as visual attention.

Recently, a number of computational models have been developed to highlight salient regions. As far as computational models are concerned, there are two types of models in the literature, namely bottom-up and top-down approaches. Bottom-up approach works from the low-level visual features and moves up to construct a saliency map. The top-down approach is goal-driven, and it uses prior knowledge to achieve the goal such as target detection, scene classification, and object recognition [3]. The top-down approach starts from a particular visual processing task.

The computational models of visual attention either bottom–up or top–down can be constructed in spatial domain or in frequency domain, in order to highlight the salient regions as saliency map. Spatial domain methods require more computation time to obtain the features compared to the frequency domain methods [4]. The computational complexities of these models are very high, and these models are not performed with multi-scale spatial and frequency analysis. Many models which come under frequency domain approaches have used only local features to identify the salient regions.

The very first method of saliency detection is developed by Itti et al. [5]. Local contrast information is used to develop the method. The local features are only used by Ma et al. [6], Harel et al. [7], and Goferman et al. [8] to obtain the visual saliency detection. The global features are considered in [9–12] to construct the visual saliency model. However, the accurate identification of salient regions should also involve the global features. In recent years, many researchers have shown more interest to build computational visual models in the transform domain. Mainly

Fourier transform (FT) and wavelet transform (WT) have also been extensively used to highlight the salient regions in the transform domain. Every approach has its own pros and cons.

FT gives promising results for applications involving only stationary signals. The amplitude and phase spectrum of Fourier transform is used in [13]. Guo et al. [12] used the phase spectrum of quaternion Fourier transform (PQFT) to highlight the saliency, and it is also applied for efficient video compression. Hou et al. [14] used Fourier transform and log spectrum to construct the spectral residual approach for the saliency detection.

WT has the capacity to provide multi-scale spatial and frequency analysis because it codes the signal at different bands and bandwidths. WT can represent singularities in a much better manner than FT can. And moreover, WT can be applied for nonstationary signals also [15]. WT is used in [16] to find the salient object. WT-based orientation feature maps are obtained in different scale. The order map is also found by using Fourier analysis. The local, global information are used in the WT-based salient point detection [17]. WT-based salient detection is used in multi-scale image retrieval problem.

Wavelets are very good to represent point singularities, but when it comes to directional features they fall short [18]. The main reason is that wavelets are generated by isotropic elements. The quantity of decomposition level should be very large when approximating a curve using WT. The disadvantages of WT have been overcome by using multi-directional and multi-scale transforms.

The higher directional wavelet transforms of ridgelet, curvelet, shearlet are also used for the visual saliency detection. The directional features are captured effectively, and the potential salient regions are identified. Bao et al. [19] proposed visual saliency detection based on shearlet transform, in which the local and global contrasts are used to obtain local and global saliency map. Initially, the potential salient regions are identified in order to update the feature maps in shearlet domain.

Even though the transform based saliency detection methods are producing reliable results, suffered highly from computation complexity. Larger computations are required for these methods.

In general, the transform domain visual saliency detection methods are using the following steps.

- Transform the input image into transform domain,
- Obtain the feature maps for various features,
- Combine the various feature maps,
- Use the top-down features if required for particular application,
- Apply inverse transforms to get the saliency map.

The saliency map is the topographical map which shows the visual saliency in the visual scene. So the visual saliency detection methods are showing the outputs as visual saliency map. Saliency-based image segmentation, image compression, and image retrieval are the popular areas of research.

The Walsh–Hadamard transform (WHT) has lesser computations and extremely fast transform. It is computed only by addition and subtraction. Lesser hardware is required for the practical implementation [20]. The highly correlated pixels are captured by the WHT in the visual space. Hence, in this chapter WHT is used to detect the salient regions in transform domain.

2.2 Visual Saliency-Based Image Compression

Usually, the important regions of an image may be small and highly degraded at low bitrates. The standards of compression such as JPEG/JPEG-2000, MPEG 4 are not handling the salient regions well. Guo et al. [21] and Hadi et al. [22] investigated saliency-based compression techniques. To adhere the saliency values, the transmitted coefficients are modified. These methods cannot handle salient regions well and also suffer from complicated computations. Barua et al. [23] developed wavelet-based image compression technique for images and videos. The algorithm is designed to obtain the saliency values in wavelet domain and then corresponding image/video coefficients are transmitted. It preserves the important region of an image/video.

Nabil Ouerhani et al. [24] proposed adaptive color image compression based on biologically inspired visual attention. The initial stage perceptual salient regions of interest are identified automatically. The adaptive coding scheme allocates higher number of bits for the salient regions. The results are compatible with the JPEG standards.

Li et al. [25] performed video compression based on computational models of visual attention. The salient regions are encoded with higher quality compared with non-salient regions. The salient regions are awarded with higher priority rather than others. But it may generate visible artifacts in the non-salient part where the quality of image is poorer. The artifacts also sometimes draw the end-user's attention. In several cases, the high level of artifacts becomes salient and captures the viewer's attention. But the notable artifacts are not to be salient.

Hadizadeh [26] dissertation reveals the visual saliency methods for video compression and transmission. Saliency-based video coding is investigated. The main concept is that high salient regions are having higher ability to percept than lesser salient regions. The quality of image/video is handled toward the user most attended regions. This method effectively performs video coding expect in two major cases. If any region is richly salient, then its saliency will be increased after the compression, provided the quality of the image/video remains high. The reason is that the users are noticing the high salient regions in the scene. If the region is lesser salient, then its saliency will be decreasing after the compression task, because the lower saliency regions are ended with lesser quality.

Ho-Phuoc et al. [27] proposed the visual saliency-based data compression for image sensors. The adaptive image compression is presented in each block. First the saliency value is obtained, and then the Haar wavelet transform is applied for the compression. This framework gives lesser memory and compact operators. The data stored in the image sensors are very much reduced, and image quality is not altered.

Zundy et al. [28] proposed the content-based image compression using visual saliency methods. Initially, the saliency map is obtained from the video, automatically or by user input. The salient regions are performed with nonlinear image scaling. Salient image regions are given higher pixel count, and non-salient regions are given lesser pixel count. Existing compression techniques are utilized to compress the nonlinearly down-scaled images, and in the receiver end it is up-scaled. This method supports for anti-aliasing effect which reduces the aliasing in highly scaled regions.

In order to reduce redundant information in the dynamic scenes, the visual saliency in videos is proposed by Tu et al. [29]. Based on the video visual saliency map, the redundant information is removed. In this paper, video visual saliency is the catalyst of video compression technique. DCT is utilized to perform the video compression, and this technique is adopted for MPEG-1, MPEG-4, and H.265/HEVC standards.

Yu et al. [30] used visual saliency to guide image compression. At individual scales, the saliency is measured through Laplacian pyramid. The proposed compression algorithm decreases the entropy of the image with respect to the saliency map in each scale. Dhavale et al. [31] proposed visual computational model-based image compression. It successfully locates the regions of interest of the human and thereby applies for image compression.

Duan et al. [32] proposed the image compression technique based on saliency detection and independent component analysis. The input image is transformed first using ICA. The transformed coefficients are numbered with set zero coefficient percentage. The sparse nature of independent component analysis is used in this method. It is compared with DCT-based compression method.

Many of the state-of-the-art methods are failed to show the energy compactness. In the field of image compression, energy compactness is the key point and these methods also suffer from higher computational complexity. In order to revoke the key points, the WHT is used in the proposed image compression method. The redundant pixels are captured by the WHT with lesser computations.

2.3 Walsh–Hadamard Transform (WHT)

Discrete Fourier transform (DFT), discrete cosine transform (DCT), and Walsh–Hadamard transform (WHT) are widely used in the image processing applications. These linear image transforms are chosen in the image processing application because of their flexibility, energy compaction, and robustness. These transforms effectively extract the edges and also provide energy compaction in the

state-of-the-art methods. Among all these transforms, WHT is very gorgeous one because of its simplicity and its computational efficiency. The major properties of WHT are same as that of other image transforms. The basis vector components of WHT are orthogonal, and it is having binary values (±1) only.

WHT is orthogonal, non-sinusoidal transform which is used in image filtering, speech processing, and medical signal analysis. To be more specific the lunar images/signals are well processed, coded, and filtered by WHT. It is known well because of its simplicity and fast computation. WHT is the substitute of Fourier transform. It is computationally simpler because it requires no multiplication or division operations. Every computation is performed by simple addition and subtraction operation. WHT is one of the very fast transforms which can be implemented in $O(N\log_2 N)$ additions and subtractions. So the hardware implementation of WHT-based applications is also so simpler [20].

So it is beneficial in terms of energy consumption and lesser computation. WHT is real, orthogonal, and symmetric $H = H^* = H^{\mathrm{T}} = H^{-1}$. Walsh–Hadamard transform is represented in terms of Walsh–Hadamard transform matrix (WHTM).

It consists of set of N rows denoted by H_{j} for $j = 0, 1, 2 \ldots N - 1$. The properties of WHT matrix are:

i. H_j takes values as +1 and −1,
ii. The size of WHT matrix is usually the power of 2,
iii. H_j [0] = 1 for all j.

The size of the WHTM is generally the power of two. The second-order Hadamard matrix is given by, $H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

The WHTM with the order of 4 is $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$.

Each row in the matrix is called as basis vector of WHTM [33]. Generally, the basis vectors are orthonormal and orthogonal. Orthonormal means dot product of each basis vector themself is one. Orthogonal means dot product of any two basis vectors is zero.

The computation of WHT involves very simple step only, when the image is projected into basis images, each pixel is multiplied by ±1, whereas the FFT needed complex multiplication. So WHT is more efficient than FFT in terms of computation complexity.

Consider the image C of size $N \times N$ with the pixels of $c(x, y)$, the 2D WHT is defined as

$$H(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} c(x, y) g(x, y, u, v) \tag{1}$$

where $H(u,v)$ is the WHT-transformed image, $g(x,y,u,v)$ is the WHT kernel function. The beautiful property of WHT which is related to the energy conservation is given below

$$\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}|c(x,y)|^2=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}|H(u,v)|^2 \tag{2}$$

The energy conservation property exists between the spatial and Walsh–Hadamard domains. The average brightness of the image is calculated in terms of zero sequence term of WHT matrix

$$H(0,0)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}c(x,y) \tag{3}$$

In this proposed method, Walsh–Hadamard transform plays vital role. WHT is used to detect the salient regions in the transform domain. Visual saliency-based natural image compression is proposed in this chapter. The redundant pixels are identified with the help of saliency map and WHT. The saliency map is also used to identify the correlation among the pixels effectively.

3 Proposed Method

The framework of the proposed work is given in Fig. 1. The proposed method involves two phases. The first phase is the saliency map computation based on WHT and entropy. The second phase consists of WHT-based image compression in which saliency map is acting as catalyst.

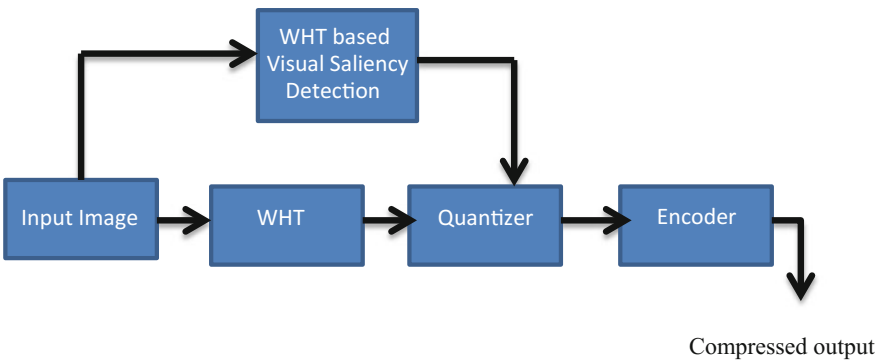


Fig. 1 Framework of proposed saliency-based image compression

Visual saliency detection is used to reduce the correlation between the pixels. Therefore, the redundant information in the pixels is identified. WHT-based quantization table is utilized. The quantization block is used to reduce the number bits needed to incorporate the transformed coefficients. In order to get the compressed image, entropy coding is used in the encoder block.

3.1 WHT-Based Saliency Map Computation

The human visual system is richly structured, and high redundant information is available in the sensory input. The predominant redundant information of human arisen from the input and the human visual system has the ability to reduce the redundant information of visual data [34, 35]. In this proposed method, WHT is used to seize the highly correlated features in the visual field and overwhelm them. So the salient visual features will be highlighted. Figure 2 is describing about the WHT-based saliency computation.

In general, natural images can be classified into salient and non-salient regions. The uncommon regions of an image are identified as salient regions. The uncommon regions of an image are generally caused by the preattentive features such as color, orientation, texture, and shape. In order to identify the salient regions, the suitable color channels must be used. Luo et al. [36] revealed that the salient regions are identified easily in specific color channels. Under different conditions, the images may have different contrasts in the diverse color channels. Usually YCbCr

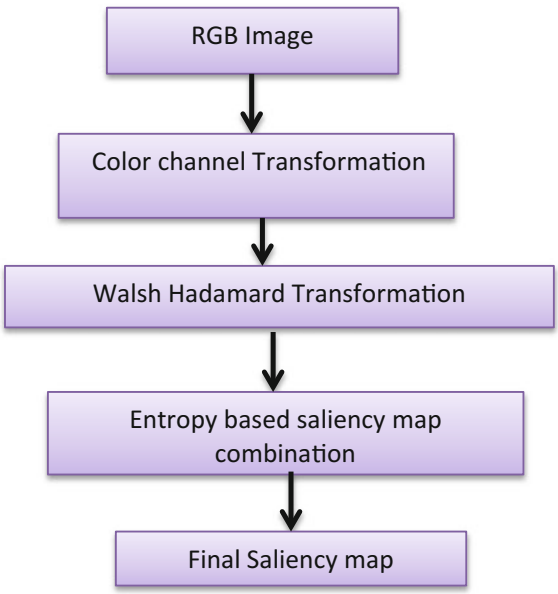


Fig. 2 Framework of WHT-based saliency map computation

color channels, Lab color channels are used in the saliency detection literature. In the human visual system, the color space of images is separated into various decorrelated color channels. RGB is one of the highly correlated color spaces, whereas lab color space is a well decorrelated color space with uniformity [3].

Recently, RGBYI color channels are used in the saliency detection [37]. In this proposed method also, RGBYI color channel is employed because this color channel mimics the visual cortex of human system by involving color channel pairs such as yellow/blue, blue/yellow, red/green, and green/red. It also includes color difference channels and four broadly tuned color channels. Let r, g, and b be the red, green, and blue color channels of an input image. The grayscale image is obtained by Eq. (4). Four broadly color channels and their pairs are obtained by Eqs. (5), (6), and (7)

$$\text{Intensity} = \frac{r+g+b}{3} \tag{4}$$

$$\text{RE} = r - \frac{g+b}{2}, \quad \text{GR} = g - \frac{r+b}{2}, \quad \text{BL} = b - \frac{r+g}{2} \tag{5}$$

$$\text{YELLOW} = \frac{r+g}{2} - \frac{|r-g|}{2} - b \tag{6}$$

$$\text{REGR} = R - G \,\text{and}\, \text{BLYE} = B - Y \tag{7}$$

The input image is represented through r, g, b, RE, GR, BL, REGR, BLYE, intensity color channels. The components of WHT reflect the global features in the visual field. So the highly correlated features are seized by WHT components. The 2D WHT is performed by each row-wise and column-wise. So the transformation of WHT is very much simpler compared to the fast algorithms, and so, the computational complexity of WHT is extremely lower than other transforms. Let consider I^{C} be the color image. Each channels of I^{C} are applied with Walsh–Hadamard transform, and the transformed image is represented as

$$X = \text{sign}\left(\text{WHT}\, I^{\text{C}}\right) \tag{8}$$

Sign (.) represents signum function. WHT represents 2D Walsh–Hadamard transform. It holds the sign information of WHT components and omitting the other information such as amplitude, phase. Binary information (1 and −1) is only considered. So WHT is very compact by assigning single bit for each component. The WHT components are normalized by setting 1 for all positive coefficients and −1 for negative components. Sign (.) function usually boosts the image features as binary decision process. It reduces the highly correlated features in the visual field by normalizing WHT coefficients, thereby yielding visual saliency in the transform

domain. The inverse Walsh–Hadamard transform is applied to get the visual saliency map in the transform domain.

$$Y = \text{abs}(\text{IWHT}(X)) \tag{9}$$

All the nine color channels are performed with these steps. The final saliency map is obtained by combining the saliency maps at each color channel. The concept of entropy is vested to combine the saliency maps. The salient region and non-salient regions are differentiated from each other form in the saliency map at certain level. The salient regions may produce higher values, whereas non-salient regions produce lesser values. The weight of the saliency map is calculated by taking reciprocal of the entropy.

$$\text{SMAP} = \sum_{i=1}^{9} \frac{Y_i}{M\{H(Y_i)\}}. \tag{10}$$

'M' is for the normalization operator [0 1], where H (.) is the entropy. It is defined as

$$H(y) = -\sum_{i=1}^{n} p_i \log p_i. \tag{11}$$

p_i is the probability of different values of y. It varies from 0 to 255. 'n' is the possible values of 'y'. If the saliency map is having higher entropy, then the weight will be a lower one. The weight of the saliency map and its entropy are inversely proportional to each other. To smooth the saliency map, the Gaussian low-pass filter is applied before the updation of entropy.

The major steps in the WHT-based saliency detection are given below:

i. In the initial stage, input image is converted into RGBYI color channels with suitable scale,
ii. Apply Walsh–Hadamard transform (WHT) for each color channel and obtain the saliency maps for each channel,
iii. The final saliency map is obtained by combining the saliency maps using entropy concept.

3.2 Saliency-Based Image Compression

The multimedia data of image, audio, and video require extensive capacity to store. If these data are compressed, then quite a considerable amount of space is only

needed for transmission or storage. In the lower capacity, space is not sufficient to store large amount of data. So it is mandatory to perform compression while transferring the large amount of data from one end to other end. The reduction of redundant information in the multimedia data is the main goal of the compression techniques.

The proposed method highlights how the visual saliency-based methods are used to reduce the redundant information in the input and thereby performs compression. The framework of the proposed method is shown in Fig. 1. In this proposed method, there are three major blocks. Initially, the input image is divided into subblocks in order to apply Walsh–Hadamard transform. It is followed by quantizer unit. The quantitation table depends on the output from visual saliency detection and the coefficients of WHT. The modified quantitation table is used in the encoder unit. Entropy-based encoding technique is vested in the proposed method.

3.2.1 Preprocessing Stage

In order to design the quantitation table for quantizer, it is necessary to perform preprocessing steps in the initial stage. Human visual system is more sensitive to luminance than chrominance. RGB color space is highly correlated. The inter-component correlation among Red, Green, and Blue color channels is high. In order to obtain good image compression performance, it is necessary to reduce the correlation among the color channels [38].

YCbCr color space is one of the decorrelated color spaces. So RGB color space is not well suited for image compression. In this proposed method, input image is converted from RGB color space to YCbCr color space. Each color channel of YCbCr is subsampled into 8×8 nonoverlapping blocks.

3.2.2 Transform Domain

HT is non-sinusoidal orthogonal transform. It has wide application in the field of image processing. It is so popular among the research community because of its decorrelating capability. It also requires simple hardware implementation. The kernel function of WHT based on square and rectangular pulses with the highest magnitudes of ± 1. For two-dimensional images, the WHT and IWHT kernels are identical. It is due to the symmetric matrix in the kernels. Orthogonal rows and columns are in the symmetric matrix. So the forward and inverse WHT kernels are identical. The 2×2 smallest orthogonal Hadamard matrix is given by,

$$H_2 = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}$$

The WHT coefficients of an image I^C with the size of $N \times N$ are computed by

$$H(x,y) = \frac{1}{N}\sum_{r=0}^{N-1}\sum_{c=0}^{N-1} I^C(r,c)(-1)\sum_{i=0}^{n-1}[b_i(r)p_i(x) + b_i(c)p_i(y)] \tag{12}$$

Each 8×8 nonoverlapping block is applied with WHT, where $H(x,y)$ is the result of Hadamard transform. $I^C(r,c)$ is the image pixel value in spatial domain at (r,c) position. At frequency domain, the index is represented as (x,y).

N is the dimension of an image $N = 2^n$. The binary representation of r at ith bit is denoted as $b_i(r)$. $p_i(x)$ in terms of $b_i(x)$ is given below:

$$p_0(x) = b_{n-1}(x) \tag{13}$$

$$p_1(x) = b_{n-1}(x) + b_{n-2}(x) \tag{14}$$

$$p_2(x) = b_{n-2}(x) + b_{n-3}(x) \tag{15}$$

$$\text{simillarly } p_{n-1}(x) = b_1(x) + b_0(x) \tag{16}$$

The transform coefficients are used in the construction of quantization matrix.

3.2.3 Quantitation

Quantitation step is used to reduce the number of bits needed to store the transformed coefficients. WHT coefficients are representing the frequency components in the range of low, high, and middle. The saliency mask is generated to highlight the high-frequency components. It obtains the coefficient depending on the saliency map.

$$\text{SMAP}'(x,y) = \begin{cases} 0, & \text{if } x \le 0.08 \text{ and } y \le 0.08 \\ 1, & \text{otherwise} \end{cases} \tag{17}$$

Visual saliency-based modified Hadamard matrix is obtained by multiplying the WHT coefficients with the human visual saliency map.

$$H'(x,y) = H(x,y) * \text{SMAP}'(x,y) \tag{18}$$

In this proposed method, quantitation is performed for each block with respect to the saliency maps which are obtained in Sect. 3.1. Now, the quantitation matrix contributes the perceptual quality of human attention. The quantization matrix can be obtained by

$$Q(x,y) = \frac{q}{H'(x,y)} \tag{19}$$

where $Q(x, y)$ is the quantization matrix; q is the step size of the uniform quantizer; $H'(x, y)$ is visual saliency-based modified Hadamard matrix. The quantization matrix is rounded off and is represented as.

$$Q'(x, y) = \text{round}[Q(x, y)] \tag{20}$$

3.2.4 Encoding

According to human visual perception, the removal of high-frequency components is not at all achieving considerable impact on the input. The reduction of high-frequency contents yields the output of better compression [39]. Mainly, the human fixations felt into salient regions of an image. Huffman coding and arithmetic coding are needed with additional steps to encode the transform coefficients. Due to these extra steps, there are certain losses in the information. However, entropy coding is a lossless technique. In this proposed method, entropy-based encoding is used.

In this proposed method, saliency map-based quantitation matrix is obtained which reduces the correlation between the two adjacent pixels of an input image. If the correlation between pixels is reduced, then automatically the performance will be increased.

i. Convert RGB color space to YCbCr color space,
ii. Divide the input image into 8 × 8 blocks,
iii. Each block is applied with Hadamard transform,
iv. Obtain saliency map-based quantitation table,
v. Perform the quantitation and rounding.

The redundant pixels are captured by the WHT with the assistance of second-order correlation statistics.

4 Experimental Results

According to the proposed method which is mentioned in Sect. 3, there are two major parts. First part is Walsh–Hadamard-based saliency map generation. Second part is saliency map-based image compression. Following subsections clearly intimate the results of WHT-based saliency detection and saliency-based image compression.

4.1 Results of WHT-Based Saliency Detection

In this section, the objective and subjective results of WHT-based saliency detection are discussed. The proposed method is compared with the seven state-of-the-art saliency detection approaches. However, the transform domain approaches are only taken for considerations. To validate the results of the proposed method, 1003 images are used from MIT dataset [40]. It consists of indoor, outdoor images, and portraits. All these images are captured from 15 viewers under free-viewing task.

The selected extant methods are IT in Itti et al. [5], VS in Li et al. [13], SR in Hou et al. [14], SE in Murray et al. [41], WT in Imamoglu et al. [3], ST in Lei Bao et al. [19], and FT in Achanta et al. [9]. It is the very first computational model in the field of visual saliency detection. The other six models proposed are in various transform domains. Fourier transform-based methods [9, 13, 14], wavelet transform-based methods [3, 41], and shearlet transform-based methods [19] are considered for the fair comparison. In order to enumerate the consistency of the results, two major evaluation metrics are obtained.

Receiver operating characteristics (ROC) and area under curve (AUC) are the major metrics. It is found in many of the major saliency detection methods for objective evaluation. Receiver operating characteristics (ROC) are plotted between true-positive rate (TPR) and false-positive rate (FPR). In order to do a fair comparison, all saliency maps are obtained by the state-of-the-art models, and the proposed method is normalized in the same range [0, 255] of original images. The better visual image saliency detection method is tried to have huge AUC values.

Figure 3 shows the subjective comparison of saliency maps of the proposed method and seven existing methods on MIT benchmark dataset. The proposed WHT method can highlight the salient region even though the image background is cluttered (row 5 in Fig. 3). The images with complex background are also effectively handled by the proposed method. The performances of the proposed method are evaluated, with the parameters of precision P, recall R, and F-measure F_α which are considered as first test criteria. These parameters are defined as:

$$P = \frac{\sum_x \sum_y (g(x,y) \times s(x,y))}{\sum_x \sum_y s(x,y)} \tag{21}$$

$$R = \frac{\sum_x \sum_y (g(x,y) \times s(x,y))}{\sum_x \sum_y g(x,y)} \tag{22}$$

$$F_\alpha = \frac{(1+\alpha) \times P \times R}{\alpha \times P + R} \tag{23}$$

where $g(x,y)$ is the ground truth, $s(x,y)$ is the saliency map of the proposed method. α is taken as 0.3 in this work [3]. The performance metrics are calculated for proposed and state-of-the-art methods. The accurate assignment of salient pixels is highlighted by precision P. Recall R relates the correct detection of salient

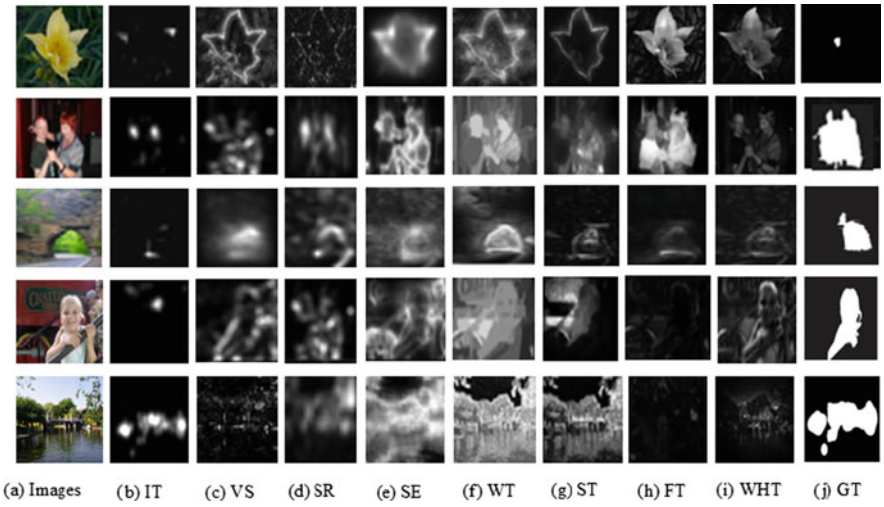


Fig. 3 Examples of saliency maps over MIT-1003 dataset **a** input images (image 1–image 5) **i** proposed WHT method **j** ground truths and **b–h** saliency maps of other existing methods

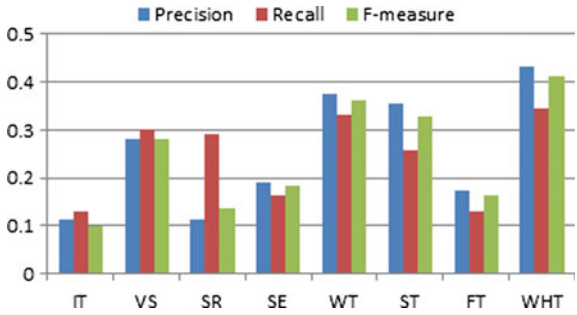


Fig. 4 Precision, recall, *F*-measure of the proposed WHT method and state-of-the-art methods on MIT-1003 dataset

regions in accordance with ground truth. *F*-measure F_α is the harmonic mean of *P* and *R*. Binary images of the saliency map are obtained by considering the mean value of saliency map and Otsu automatic threshold algorithm [42].

Each saliency map is the combination of salient regions and non-salient regions. Imamoglu et al. [3] explained true-positive rate (TPR) is the amount of salient pixels in the ground truth $g(x, y)$ intersecting with the amount of salient pixels in the saliency map $s(x, y)$. False-positive rate (FPR) is the amount of non-salient pixels from ground truth $g(x, y)$ intersecting with the salient pixels in the saliency map $s(x, y)$. ROC curve is the plot which is drawn between TPR and FPR. The relationship between the false positives and false negatives is linked in each point of ROC curve. The ROC curves and PR charts are given in Figs. 4 and 5, respectively. From Figs. 4 and 5, it is understood that proposed method shows higher performance with respect to the state-of-the-art methods.

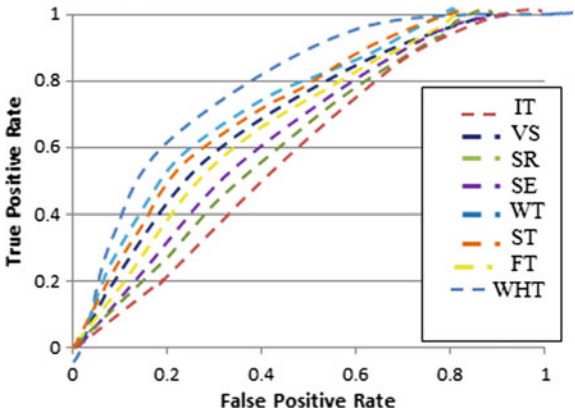


Fig. 5 ROC of the proposed method and state-of-the-art method on MIT-1003 dataset

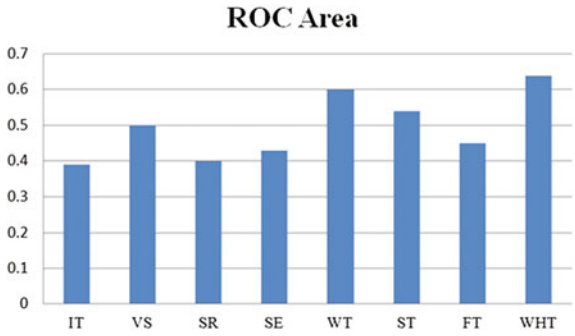


Fig. 6 AUC of the proposed WHT method and other existing methods

Figure 6 shows the ROC area of MIT benchmark datasets wherein the proposed method outperforms the state-of-the-art methods.

4.2 Results of Visual Saliency-Based Image Compression

The evaluation of visual saliency-based compression is performed for MIT benchmark dataset. Section 4.1 clearly shows the performance analysis of saliency detection of MIT dataset. To evaluate the consistency of the proposed method, the major performance metrics of peak signal-to-noise ratio (PSNR), compression ratio (CR), structural similarity index (SSIM), and mean square error (MSE) are obtained. The efficiency of the proposed method is assessed with certain state-of-the-art methods. These methods are performed to find image compression based on saliency analysis and DCT, DWT transforms. The methods are abbreviated as SM_DCT [32] and SM_DWT [23]. The proposed method is also compared

with JPEG [43] and DWT [44] methods. The proposed method is named as SM_WHT.

The input images (image 1–image 5) of Fig. 3 are used for the evaluation. The experiment is performed on a machine with Intel i3 2.4 GHz CPU and 4 GB RAM. The proposed method and all the comparative methods are implemented in Matlab R2012b. The average running time taken to process 400 × 300 image is 0.31 s. The image quality of the proposed method and state-of-the-art methods are analyzed by the performance metric of PSNR and SSIM. The compression ratio is achieved through the performance of metric of Bpp.

4.3 Performance Analysis

All the compression methods are having the main objective to obtain best quality of images with lesser utilization of bit. PSNR is one of the important parameters to evaluate the quality of an image. If the PSNR value is high, then the image quality is good. If it is low, then the quality of an image is low. It is also a type of objective measurement which is based on the mean square error (MSE) and it is given by,

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{24}$$

The mean squared error of an image is defined as

$$\mathrm{MSE} = \frac{1}{M \times N} \times \left[\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(A_{ij} - B_{ij}\right)^2\right] \tag{25}$$

where M × N denotes the size of the input image. A_{ij} is the input original image, and B_{ij} is the compressed image. The images of MIT dataset are resized as 200 × 200. The proposed method is experimented on different images of MIT datasets. The image quality of the proposed method is obtained by the PSNR which is shown in Fig. 7.

SSIM is another performance metric which is also used to evaluate the image quality. It is used to highlight the structural similarity of input and output images based on the human visual system characteristics. It quantifies the structural similarity other than error visibility of two images. SSIM gives better interpretation of image quality compared to the PSNR metric. It is defined as

$$\mathrm{SSIM}(u, v) = \frac{(2m_u m_v + d_1)(2\sigma_{uv} + d_2)}{\left(m_u^2 + m_v^2 + d_1\right)\left(\sigma_u^2 + \sigma_v^2 + d_2\right)} \tag{26}$$

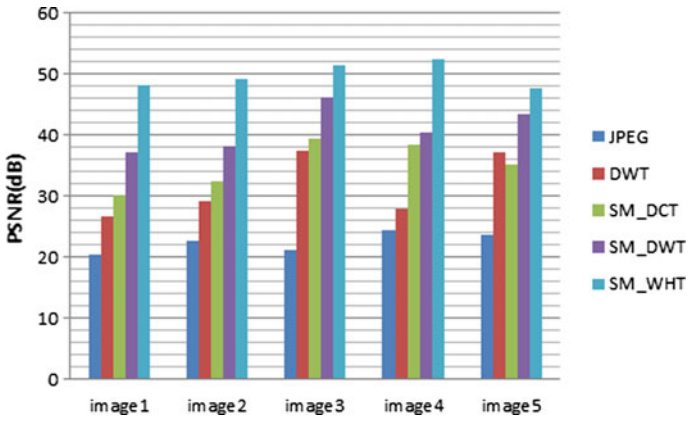


Fig. 7 PSNR (dB) of proposed and existing compression methods

where u and v are the block size of the images. $m_u m_v$ are the mean intensity values of u, v. σ_u^2, σ_v^2 are the standard deviation of u, v. d_1, d_2 are the constants.

From Fig. 8, it is understood that the proposed saliency-based WHT compression method produces better results than existing state-of-the-art methods. It is reaching nearly one. The proposed method uses the saliency-based quantitation table for image compression.

The compression ratio is the important performance metric in the scenario of image compression. If the compression ratio is higher, then a lot of memory is needed to store the data. If it is less, then lesser memory is only needed to store the output image. The compression ratio is defined as the total number of bits in the input image to the compressed image. It is used to estimate how much amount of compression has taken place. In the image, some regions are most important and others are less important. The most important regions are called salient regions. The compression ratio of salient regions is lesser compared to other regions. Whereas other regions are highly compressed, the compression ratios obtained from proposed method and state-of-the-art methods are updated in Table 1.

$$\text{Compression Ratio (CR)} = \frac{\text{No of bits in the compressed image}}{\text{No of bits in the original image}} \tag{27}$$

Table 1 clearly shows the compression ratios of the proposed method and state-of-the-art methods. The average compression ratio of the proposed method is 90%. The proposed SM_WHT image compression method extremely performs well compared to the saliency-based DWT and DCT methods. It is mainly due to the lesser number of computations in WHT.

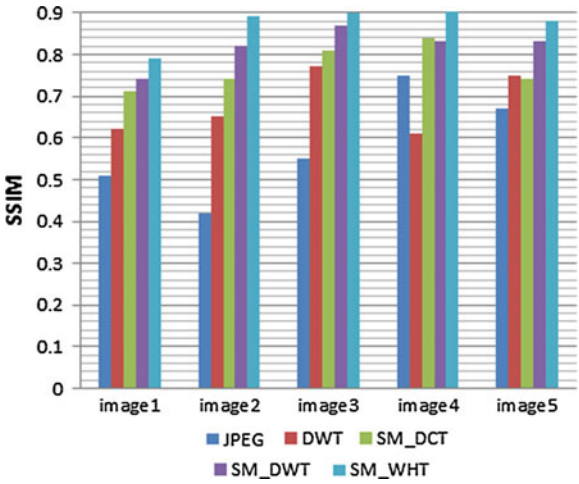


Fig. 8 SSIM of proposed and existing compression methods

Table 1 Compression ratio of the proposed method and state-of–the-art methods

	Compression methods				
Images	JPEG	DWT	SM_DCT	SM_DWT	SM_WHT
Image 1	0.75	0.77	0.82	0.83	0.88
Image 2	0.79	0.84	0.89	0.88	0.91
Image 3	0.84	0.86	0.88	0.90	0.92
Image 4	0.77	0.88	0.86	0.92	0.90
Image 5	0.80	0.84	0.90	0.82	0.89

5 Conclusion

Visual saliency-based image compression is exhausted in this chapter. Initially, the most important regions of an image are highlighted by the saliency map. Then the saliency-based quantitation table is used in the compression method. The performance metric of PSNR, SSIM, and compression ratio is obtained in order to estimate the performance of the proposed method. The proposed visual saliency-based compression method yields the average compression ratio of 90% compared to the state-of-the-art compression methods. The redundant pixels are captured by the WHT with the assistance of second-order correlation. Lesser computations in the WHT transform are the most added advantage in the proposed compression technique. The proposed method can be extended by considering the motion images as the inputs. The hardware implementation of the proposed method can be performed as the future work.

References

1. Guo C, Zhang L (2010) A novel multi resolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19, 185–198
2. Arya R, Singh N, Agrawal RK (2015) A novel hybrid approach for salient object detection using local and global saliency in frequency domain. Multimed Tools Appl. doi:10.1007/s11042-015-2750-y
3. Imamoglu N, Lin WS, Fang YM (2013) A saliency detection model using low-level features based on wavelet transform. IEEE Trans Multimedia 15:96–105
4. Lin RJ, Lin WS (2014) Computational visual saliency model based on statistics and machine learning. J Vision 14(9), 1–18
5. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259
6. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: ACM International conference on multimedia, pp 374–381, Berkeley (2003)
7. Harel J, Koch C, Perona P (2006) Graph-based visual saliency. Neural Inf Process Syst 545–552
8. Goferman S, Zelnik Manor L, Tal A Context-aware saliency detection. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2376–2383
9. Achanta R, Hemami S, Estrada F, Susstrunk (2009) Frequency-tuned salient region detection. IEEE conference on computer vision and pattern recognition, pp 1597–1604
10. Cheng MM, Zhang GX, Mitra NJ, Huang X, Hu SM (2015) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
11. Perazzi F, Krahenbuhl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection. In IEEE conference on computer vision and pattern recognition (CVPR), pp 733–740
12. Guo C, Ma Q, Zhang L (2008) Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 1–8
13. Li J, Levine MD, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. IEEE Trans Pattern Anal Mach Intell 35:996–1010
14. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition (CVPR), pp. 1–8
15. Merry RJE (2005) Wavelet theory and application-a literature study. Eindhoven University of Technology, The Netherlands
16. Li ZQ, Fang T, Huo H (2010) A saliency model based on wavelet transform and visual attention. Sci China Inf Sci 53(4):738–751
17. Tian Q, Sebe N, Lew MS, Loupias E, Huang TS (2001) Image retrieval using wavelet-based salient points. Electron Imag 10(4):835–849
18. Candes E, Donoho D (2004) New tight frames of curvelets and optimal representations of objects with piecewise singularities. Commun Pure Appl Math 57:219–266
19. Bao L, Lu J, Li Y, Shi Y (2014) A saliency detection model using shearlet transform. Multimed Tools Appl. doi:10.1007/s11042-014-2043-x(2014)
20. Yu Y, Yang J (2016) Visual saliency using binary spectrum of Walsh–Hadamard transform and its applications to ship detection in multispectral imagery. Neural Process Lett. doi:10.1007/s11063-016-9507-0
21. Guo C, Zhang L (2010) A novel multi resolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19(1):185–198
22. Hadizadeh H, Bajic I (2014) Saliency-aware video compression. IEEE Trans Image Process 23(1):19–33
23. Barua S, Mitra K, Veeraraghavan A (2015) Saliency guided Wavelet compression for low-bitrate Image and Video coding. **I**n: IEEE global conference on signal and information processing (2015)

24. Ouerhani N, Bracamonte J, Hugli H, Ansorge M, Pellandini F (2001) Adaptive color image compression based on visual attention. International conference on image analysis and processing, pp 26–28
25. Li Z, Qin S, Itti L (2011) Visual attention guided bit allocation in video compression. Image Vision Comp 29(1):1–14
26. Hadizadeh H (2013) visual saliency in video compression and transmission. Thesis dissertation (2013)
27. Ho-Phuoc T, Dupret A, Alacoque L (2012) Saliency-based data compression for image sensors. Sensors, IEEE
28. Zundy F, Pritch Y, Sorkine-Hornung A, Mangold S, Gross T (2013) Content-aware compression using saliency-driven image retargeting. In: 20th IEEE international conference on image processing (ICIP) (2013)
29. Tu Q, Mena A, Jiang Z, Ye F, Xu J (2015) Video saliency detection incorporating temporal information in compressed domain. Signal Process: Image Commun 38, 32–44
30. Yu SX, Lisin DA (2009) Image compression based on visual saliency at individual scales. In International symposium on visual computing, USA
31. Dhavale N, Itti L (2003) Saliency-based multi-foveated MPEG compression. In: Proceedings of signal processing and its applications pp 229–232 (2003)
32. Duan L, Ke C (2012) A natural image compression approach based on independent component analysis and visual saliency detection. Adv Sci Lett 5, 1–4
33. Lakshmi Priya GG, Domnic S (2014) Walsh-Hadamard transform kernel-based feature vector for shot boundary detection. IEEE Trans Image Process 23(12):5187–5197
34. Petrov Y, Li Z (2003) Local correlations, information redundancy, and sufficient pixel depth in natural images. J Opt Soc Am A 20(1):56–66
35. Li Z, Atick JJ (1994) Toward a theory of the striate cortex. Neural Comput 6(1):127–146
36. Luo W, Li H, Liu G, Ngan KN (2012) Global salient information maximization for saliency detection. Sig Process: Image Commun 27(3):238–248
37. Ma X, Xie X, Lam K-M, Zhong Y (2015) Efficient saliency analysis based on wavelet transform and entropy theory. J Vis Commun Image R 30:201–207
38. Acharya T, Tsai PS (2005) JPEG2000 standard for image compression: concepts, algorithms and VLSI architecture. Wiley 60(2005)
39. Gupta R, Khanna MT, Chaudhur S (2013) Visual saliency guided video compression algorithm. Signal Process: Image Commun 28:1006–1022
40. Judd T, Ehinger K, Durand F, Torralba A (2009) Learning to predict where humans look. In: IEEE conference on computer vision and pattern recognition (CVPR)
41. Murray N, Vanrell M, Otazu X, Parraga CA (2011) Saliency estimation using a non-parametric low-level vision model. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 433–440
42. Gonzalez RC, Woods RE, Eddins SL (2004) Digital signal procesing using Matlab. Prentice Hall, Englewood Cliffs, NJ
43. Wallace GK (1991) The JPEG still picture compression standard. Commun ACM 34:31–44
44. Chowdhury MMH, Khatun A (2012) Image compression using discrete wavelet transform. Int J Comput Sci 9(4)

Object Trajectory Prediction with Scarce Environment Information

Jin Sung Park, Daniela López De Luise and Jude Hemanth

Abstract This paper presents a prototype called HOLOTECH that implements a model prediction using a limited description of the environment to support blind people. The goal is to perform fast detection and identification of obstacles to provide information about collision riskiness and location. The prediction is not probabilistic but statistic, in order to improve the inferences results. The model works fine using low-precision images drifted from real-time camera of a regular Android cell phone supported with ultrasonic sensors. The main focus of this work is how to pre-process images on the fly in order to be able to train and to tune the plastic learning module, improving the object's trajectory prediction.

Keywords Computational intelligence · Blind people · Expert systems · Neural networks · Image processing · Statistical analysis

1 Introduction

Prediction of trajectories of moving objects in a real-time environment is always a challenging task. The success rate depends on the methodologies and the techniques used for prediction. Literature survey reveals many of such methodologies and their associated applications. Hash algorithm-based trajectory detection is proposed in [1]. The main objective of this work is to improve the processing efficiency. A probabilistic model of the trajectory prediction is given in [2]. Maritime search

J.S. Park (✉) · D.L. De Luise · J. Hemanth
Computational Intelligence and Information Systems Lab, Pringles, Argentina
e-mail: zeroalpha2000@gmail.com

D.L. De Luise
e-mail: daniela_ldl@ieee.org

J. Hemanth
e-mail: judehemanth@karunya.edu

J. Hemanth
Department of ECE, Karunya University, Coimbatore, Tamil Nadu, India

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_3

and rescue is the application exposed in the paper but can be extended to other applications also. Traditional machine learning algorithms have been also used for object trajectory prediction. One such work is reported in [3]. A trajectory prediction model based on shape and motion features is available in [4]. This methodology can be used for any application. The similarity between the trajectories of moving objects for real-time applications is evaluated and reported in [5]. This methodology is useful for a variety of applications in reducing the computational complexity of the process. The concept of collision avoidance in moving object trajectories is discussed in [6]. Different algorithms are discussed in this work with F-measure as the performance measure. Another model of trajectory prediction based on floating car trajectory data for urban environment is discussed in [7]. Fuzzy methodologies are used in this work for formulating the algorithm.

Time prediction method for trajectory prediction is used in [8]. This method is found to be effective for different foreground and background situations. The application of trajectory prediction in social robotics is discussed in [9]. Both probabilistic and symbolic models are used in this work for prediction. One more work in the area of trajectory prediction is reported in [10]. Stereo vision system-based trajectory prediction for moving objects is proposed in [11]. The industrial application is focused in this paper, which can be extended to any real-time applications. Another probabilistic model-based trajectory prediction is reported in [12]. This model is mainly framed for moving vehicles with in-built intelligent systems. A novel integrated trajectory prediction and control framework are developed in [13]. This model is tested on ground vehicles but can be extended to any applications. A trajectory prediction model for 3-D data is proposed in [14]. The concept of segmentation is also involved within this model. Another similar work is reported in [15].

In this work, a prototype for assisting blind people is being developed with less environment information. Statistics of the experiments conducted on the real-time video sequences are discussed in this work.

2 The HOLOTECH Model and Prototype

Mobile devices are useful in many indoor/outdoor environments. Besides, they constitute a simple and accessible technology, which nowadays can be expanded with Android applications. The prototype uses also a Nano Arduino GRAVITECH device [16], two ultrasonic sensors HC-SR04, and a pedometer. Figure 1 shows the global configuration of the hardware required to implement the model.

The goal is to collect information with redundancy, in order to improve inference precision about which objects are at the environment. The prototype is working on Android SDK 22, with MATLAB for Cascade Object detection connected to an expert system though a simple feature vector. The Arduino programs are encoded in C++ [17].

Fig. 1 HOLOTECH global disposition

The device is located on the chest of the user, to complement the position and activity of the eyes while walking.

Video sampling from the built-in camera feeds the Android module located in the cell phone, the same with the data form the Arduino's software. The core of the system is a set of neural networks (NNs) connected to an expert system (ES). The NNs are trained filters using the features derived by Cascade Object Detection (COD) for specific obstacle recognition. The workflow has many steps:

- *First step* input video. For testing purposes, the camera records at a resolution of 320 × 240 pixels. Ultrasonic input is generated by two sensors located at a 45° inclination compared to the floor, with 90° between each other.
- *Second step* slice into images every 30 s.
- *Third step* after slicing, the images are processed to determine which objects are present in the sequence.
- *Fourth step* object segmentation; convert into a gray scale, segment, chop, and extract the border [18, 19].
- *Fifth step* evaluate location and distance of obstacles: process with haar-like object-detection feature. Evaluate a short subsequence of images with a fixed sliding window in time; it is possible to obtain the acceleration, orientation, and displacements in the scene. The distance of the obstacles is raw data from the ultrasonic sensors.
- *Sixth step* encode position/distance and activate the proper sound in the alarm system (see [20]): The system implements a specific language varying frequency, tone, and vibration. To do so, an ES [21] analyzes the information extracted and determines the urgency of the communication. It implements a context-free grammar that is flexible enough to cover more new situations.
- *Seventh step* sound production restated by the model's grammar. It communicates any risk situation (as evaluated by the ES). Rules and patterns [22] are defined in a unique way, in order to avoid confusion or ambiguity.

3 Train, Testing, and Results

In order to validate the model and the procedure to statistically analyze the data and its precision, the training of the NN was done using an image bank of the subject [23].

For a change and to confirm that the NN can be used to recognize different kind of subjects, we have shifted to recognize cars. There were some limitations to the recognition of the subject, nevertheless they did not affect the statistical analysis to calculate and extract useful information about the environment. Because the information extraction of the distance of the subjects from a image produce extra processing, the use of the Arduino's ultrasonic was used to deal with it in a fast and precise information.

3.1 Image Acquisition and Feature Extraction

Real-time videos were obtained using a video camera (2X) to validate the NN where the duration for each video is 10 s. The cars are identified as objects which are the region of interest. Initially, each video is sliced into images which results in 200–300 frames for each video dataset. The dimension of each image is 720 $\times$ 480 pixels, and the average size is 30 KB. The format of raw images is jpg, which is converted to avi format for further processing [24]. Among these huge number of frames, 10 frames are chosen for further processing. The selection of these frames is done randomly, and they are different for each of the input video. It may be noted that frame selection is less significant, since the objective of this work is only to detect the objects. For validation purposes, some other techniques like feature-detection methods were used [25].

Furthermore, feature extraction is done in each of these frames using the NN. The features are nothing but the recognition of the subject in a region of interest in each frame. On each frame, depending on the context and situation, the NN can find more multiple features. However, frames are similar. The same feature can be found in each sequence with a displacement in their position. That displacement can be defined as the motion vector of the object of interest (cars).

3.2 Classifier Training and Precision Evaluation

The classifier used for training is haar-cascade classifier; this classifier was selected for its fast response and high success rate to recognize the object. Because the classifier does not support multiple classifications of different objects at the same time, we used many classifiers to classify all possible objects and afterward build an

expert system to integrate all specific classifiers. For this paper only, the description of the training done to recognize a car is made. Most of the procedure for the training is similar among the objects. On each case, a collection of positive images where the object is present and a collection of negative images, where the object is not present were used for the training. To start and to test the efficiency of the classifier, an image bank was used for training, and the data recorded in real time was used to test it afterward.

The image bank UIUC Image Database [23] has a collection of images of cars in two perspectives and was used for the positive sample of the training and some random set of images for the negative set. The classifier configuration was not changed and was made to use the default values where the initial value is 14, splits 1, minimum hit rate 0.995, max false alarm 0.5, weight trimming 0.95, max tree splits 0, and minimum positive sample per cluster 500. In the training, each stage, the information displayed is:

N = current feature for this cascade (seq-> total),
%%SMP = percentage of samples used, if trimmings enabled (v_wt)
F = '+' if is Flipped, if symmetry is specified (v_flipped), '−' otherwise
ST.THR = stage threshold,
HR = Hit Rate based on Stage threshold (v_hitrate/numpos),
FA = False alarm based on Stage threshold (v_falsealarm/numneg)
EXP.ERR = Strong classification error of adaboost algorithm, based on threshold = 0 (v_experr). This is a description of the information on the classifier.

An example result of training's stage one is that more the features are used by the HR, the worse the results are.

Positive sample: 500 502 0.996016
Negative sample: 500 0.45045
Background processing time: 0.00
Pre-calculation time: 0.00.

(Table 1).

Stage training time: 152.00.
Number of used features: 7.
Parent node: 0.
Chosen number of splits: 0.
Total number of splits: 0.
Result of stage 1 training.

As can be seen in the example, with a result that is a short version of the log, the HR and FA are pretty balanced, giving an excellent EXP.ERR rate. However, in each stage of the cascade, the result will vary. Once the training is done, an XML containing the classification logic is generated. Using this XML and some testing images, the real-time images are collected. The next step is to test the performance of the model for real-time cases.

Table 1 Training statistical information for Cascade Detector

N	%SMP (%)	F	ST.THR	HR	FA	EXP. ERR
1	100	−	−0.703392	1.000000	1.000000	0.168000
2	100	+	−1.186522	1.000000	1.000000	0.168000
3	100	−	−1.753016	1.000000	1.000000	0.132000
4	84	+	−1.960341	1.000000	1.000000	0.130000
5	86	−	−1.567593	0.998000	0.734000	0.121000
6	74	+	−1.470981	0.996000	0.516000	0.108000
7	69	−	−1.175175	0.996000	0.382000	0.075000

Using 10 sets of images to get an overall test, the classifier with highest hit rates was used to calculate the kappa statistics value (k) of the model. This test is done using the tool provided from opencv. The formula for such statistics is given in [26]:

$$K = P(a) - P(\mathrm{e})/(1 - P(\mathrm{e}))$$

where $P(a)$ is the probability of an event called "a", and $P(\mathrm{e})$ is the statistical probability of its occurrence. The performance evaluation is as follows:

As it can be seen from the Table 2, hits are balanced with missed cases but in just one case there is a false positive that indicates the approach to avoid most of the false alarms. Once the model is ready, using it with the real-time image, we found situations where the application of statistical analysis was possible. For the data construction for each successive frame, the model was applied, and its features extracted and logged and recorded like in the Fig. 2. Each feature has a representation in the frame in form of a blue block where each one has a position x, y, h, w.

Table 2 Performance testing for the detector

File name	Hits	Missed	False
tests/img1.jpg	0	1	0
tests/img2.jpg	1	0	0
tests/img3.jpg	1	0	0
tests/img4.jpg	1	0	0
tests/img5.jpg	0	1	1
tests/img6.jpg	1	0	0
tests/img7.jpg	0	1	0
Total	5	4	1

Fig. 2 Figure of the feature detected

4 Data Flexibility

In order to properly evaluate the performance of this proposal, a testing set consisting of 118 pictures, extracted from a video dataset, was used for validation purposes. As mentioned, training set is part of a collection of videos collected with a cell phone and the bank of images, with the goal to expose main parts of cars, trees, and faces. Therefore, images derived by the slicing process have an excellent performance.

But it is not the case with the testing set used for this section: In order to assess the robustness of the proposed model, the dataset has many videos with lower quality. Frequently, the objects of interest are far, fuzzy, or distorted. As an example see Fig. 3, a training video of 10 s derived into two images. The squares show the detected object.

The prototype successfully detected a car, a face, and a tree. There is still an extra blue box detecting part of a gate that is a false-positive detection. It is important to note that every video considered for training has a 10-s duration, and two images were systematically extracted from each one. Images were paired in order to keep together those that belong to the same original video. A similar process was performed with the testing data. The images in Fig. 4 are part of the dataset used for the present statistical analysis. They were extracted from video-ID 15.

Fig. 3 Two images extracted from a training video. Objects are clear and performance is very good

Fig. 4 Two images extracted from video 15. Lightning changes and object's position are highly variable

The characteristics of this set are as follows:

- scenes can vary with much velocity (see in the example that even the orientation of the car has changed).
- as a consequence of the previous tip, lightning and number of objects to be detected also change at same velocity. Note that in the case of a pedestrian, it is hard to find these conditions, since the displacement is quite slow as training set. Nevertheless, this type of video was not discarded since the goal in this section is to evaluate the tolerance of the implemented proposal.
- The size of the object also may change. Figure 5 shows an example.
- As is well known, resolution of the original video highly affects the results. So every image was scaled to resolutions of 72 pixels/in.

The remaining of this section explains how all of the previous hints affect the performance. To do that, all the paired images of the testing dataset were processed, resulting in vectors with the following data:

Pair ID	Image A	A. x1	A. y1	A. w1	A. h1	…	A. Count	Image B	B. x1	…	B. Count

Pair ID: the unique identification of the case and the video
Image A: name of the first image extracted from Pair ID
A.x1 … A.x4: x coordinate of the box enclosing object 1 … 4 in image A
A.y1 … A.y4: y coordinate of the box enclosing object 1 … 4 in image A
A.w1 … A.w4: width of the box 1 … 4 in image A
A.h1 … A.h4: height of the box 1 … 4 in image A
The rest of the variables are defined similarly to them (Image B, B.x1 … B.x4, B.y1 … B.y4, B.w1 … B.w4 and B.h1 … B.h4)

The resulting dataset, consisting of 59 records with 37 attributes, was processed using WEKA 3.6.13 (c) and Inforstat 2014 (c), both of them are free platforms from

Fig. 5 Two images extracted from video 22. Size and orientation are also variable

Waikato University and Universidad Nacional de Córdoba, respectively, under Linux UBUNTU 16.04 distribution.

The first thing to test is how the number of detected objects is related to the image characteristics. Using selector of attribute of WEKA, with the fsSubsetEval algorithm (set to perform local search), it is possible to know which variables determine A count (the number of objects in image A) and B count. This algorithm evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them (b1). It assesses the predictive ability of each attribute individually and the degree of redundancy among them, preferring sets of attributes that are highly correlated with the class but have low inter-correlation.

The algorithm takes a search approach that may vary. For the present analysis, the Best First algorithm was selected. The variables that best predict the number of detected objects in A are as follows:

A.x1
A.y1
A.w1
A.y2
A.w2
A.h2
A.w3
A.h3
A.w4
A.h4

It is remarkable that variables like A.x1 and A.y1 are good predictors, since they are just coordinates. This is probably due to the low number of cases that introduce a bias in the test database, enhanced by the searching approach.

To confirm that, other search algorithms were used: Genetic search, Greedy Stepwise (backward), and Linear Forward Selection. Table 3 shows the resulting statistics for other algorithms:

Table 3 Comparison of variable selectors

BF	GS	GT	LFS
A.x1	A.x1	A.x1	A.x1
A.y1	A.y1	A.y1	A.y1
A.w1	A.w1	A.w1	A.w1
A.h1		A.h1	
A.y2	A.y2	A.y2	A.y2
A.w2	A.w2	A.w2	A.w2
A.h2	A.h2	A.h2	A.h2
A.w3	A.w3	A.w3	A.w3
A.h3	A.h3	A.h3	A.h3
A.w4	A.w4	A.w4	A.w4
A.h4	A.h4	A.h4	A.h4
B.x3		B.x3	B.x3
	A.x2		
	B.x4		

In the table, the headings are, respectively, as follows:

BF: Best First
GS: Genetic Search
GT: Greedy Stepwise
LFS: Linear Forward Selection

As can be seen, all of them have similar information: Size of the box is the main predictor of every object (w1, w2, w3, w4 and h1, h2, h3, h4). Data seem to have certain bias in the first object, probably due to the fact that every video starts with a centered object. In this way, all of the approaches take x1 and y1 as a predictor of an object. Similar results can be obtained for B count (the number of objects in the second image). Results are shown in Table 4:

In this way, it was confirmed that the main prediction ability depends on width and height (consider that all of the images have the same resolution).

Then, the next step is to determine the minimum value of them that leads to an object's inference.

To do that, the data will be filtered: Only records with null value in A count or B count are considered. The linear regression was performed with Infostat, considering A count as dependent and w1 … w4, h1 … h4 as regressors. The statistics are shown below:

$$R^2 = 0.97$$

With: coeff. (parameter estimated), probable value error of the estimation (EE), lower and upper interval thresholds (LI and LS respectively), T estimator, p statistics, and Cp Mallows. As p-value is less than 0.05 for w2, h2, h3, w3, the regression is linear only for those cases. The rest of the cases have a nonlinear regressive behavior (Table 5).

Table 4 Comparison of variable selectors

BF	GS	GT	LFS
B.x1		B.x1	B.x1
B.y1		B.y1	B.y1
B.w1	B.w1	B.w1	B.w1
B.h1	B.h1	B.h1	B.h1
B.X2	B.x2		
B.w2	B.w2	B.w2	B.w2
B.h2	B.h2	B.h2	B.h2
B.w3	B.w3	B.w3	B.w3
B.h3	B.h3	B.h3	B.h3
B.w4	B.w4	B.w4	B.w4
B.h4	B.h4	B.h4	B.h4
A.x4	A.x4	A.x4	A.x4
	A.x2		
	B.y3		
		B.x2	

Table 5 Statistical validation

Coef	Est.	E.E.	LI (95%)	LS (95%)	T	*p*-value	Cp Mallows
const	0.11	0.06	−0.01	0.23	1.80	0.0775	
A.h4	0.43	0.31	−0.20	1.06	1.36	0.1807	9.83
A.w1	−3.5 E-03	0.08	−0.16	0.15	−0.04	0.9645	8.02
A.h1	0.02	0.15	−0.29	0.33	0.14	0.8910	8.04
A.w2	−0.43	0.07	−0.56	−0.29	−6.27	<0.0001	46.52
A.h2	0.86	0.14	0.59	1.13	6.36	<0.0001	47.71
A.h3	1.32	0.24	0.84	1.80	5.55	<0.0001	38.24
A.w3	−0.66	0.12	−0.90	−0.42	−5.49	<0.0001	37.54
A.w4	−0.20	0.16	−0.52	0.11	−1.30	0.1995	9.68

Nevertheless, the data are coordinates where the subjects are present in the environment in front of the user. So the next step is using the statistical data to feed the main expert system where it will sync with the constant raw data of the distance from the sensor, sound language to give a prediction of danger and in need alert the user.

5 Conclusions and Future Work

This paper reviewed one of many approaches and presented a proposal that can process lightweight images and derive data vectors. The analysis shows that it is possible to infer the recognition in real-time and description of some of the main

behavior's features. As the model underneath is pretty global, it can be applied to different problems in several fields.

For future work, the system needs to apply the sound language, to extended object catalog recognition (probably houses, fences, etc.). Also it is pending to train the classifier with more data and change the bank image with an unbiased dataset. It is also pending to consider other versions of Android platform to make it possible to apply to more devices, and to be more flexible. Besides, it can evaluate the possibility of customizing certain details (for instance, type of sounds and specific set of obstacles of interest). All of these changes and improvements should also be complemented with an extensive statistical analysis and usability report.

References

1. Zhou Y et al (2017) An efficient data processing framework for mining the massive trajectory of moving objects. Comput Environ Urban Syst 61:129–140
2. Zhang J et al (2017) Probabilistic modelling of the drifting trajectory of an object under the effect of wind and current for maritime search and rescue. Ocean Eng 129:253–264
3. Valsamis A et al. (2016) Employing traditional machine learning algorithms for big data streams analysis: The case of object trajectory prediction. J Sys Softw http://dx.doi.org/10.1016/j.jss.2016.06.016 (In press)
4. Sheng-Zheng W, Hao-Bing N, Chao-Jian S (2014) A drifting trajectory prediction model based on object shape and stochastic motion features. J Hydrodyn 26:951–959
5. Magdy N et al (2016) A generic trajectory similarity operator in moving object databases. Egyptian Informatics J. doi:10.1016/j.eij.2016.07.001 In press
6. Lettich F (2016) Detecting avoidance behaviors between moving object trajectories. Data Knowl Eng 102:22–41
7. Kong X et al (2016) Urban traffic congestion estimation and prediction based on floating car trajectory data. Future Gener Comput Sys 61:97–107
8. Gan W, Lee M, Wu C, Kuo CJ (2016) Object tracking with temporal prediction and spatial lrefinement (TPSR). Sig Process Image Commun 47:303–312
9. Bourdonnaye FL, Setchi R, Zanni-Merk C (2016) Gaze trajectory prediction in the context of social robotics. IFAC-PapersOnLine 49:126–131
10. Abraham S, Lal PS (2012) Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations. Transp Res Part C 23:109–123
11. Mironov K, Vladimirova I, Pongratz M (2015) Processing and Forecasting the Trajectory of a Thrown Object Measured by the Stereo Vision System 48:28–35
12. Fu et al X (2014) Probabilistic trajectory prediction in intelligent driving. In: Proceedings of the 19th world congress of the international federation of automatic control, pp 2664-2672
13. Li X et al (2017) Development of a new integrated local trajectory planning and tracking control framework for autonomous ground vehicles. Mech Syst Signal Process 87:118–137
14. Xu F, Lam K, Dai Q (2011) Video-object segmentation and 3D-trajectory estimation for monocular video sequences. Image Vis Comput 29:190–205
15. Ciccio CD et al (2016) Detecting flight trajectory anomalies and predicting diversions in freight transportation. Decis Support Syst 88:1–17
16. ArduinoNano (2015) http://arduino.cc/en/Main/ArduinoBoardNano
17. Igoe T (2012) Making Things Talk, 2nd edn
18. Krasula L, Klima M, Rogard E, Jeanblanc E (2012) MATLAB-based applications for image processing and image quality assessment part II. Experimental results

19. Bala A (2012) An improved watershed image segmentation technique using MATLAB. Int J Sci Eng Res 3(6)
20. Park JS, López De Luise D, Pérez J (2015) HOLOTECH prototype. Sound language for environment's understanding. Int J Learn Technol Inderscience Publishers, Geneva
21. http://www.mathworks.com/help/vision/ug/train-a-cascade-object-detector.html. 2015
22. López De Luise D (2012) Ingeniería en Inteligencia Computacional (Computational Intelligence Engineering). In: Rovarini P (eds) (UTN-FRT), 104 pp
23. https://cogcomp.cs.illinois.edu/Data/Car/
24. Chen Q, Kotani K, Lee FF, Ohmi T A fast search algorithm for large video database using hog based features
25. Abada L, Aouat S Facial shape-from-shading using features detection method. doi:10.1504/IJAIP.2016.074774, pp 3–19
26. Carletta Jean (1996) Assessing agreement on classification tasks: the kappa statistic. Comput Linguist 22(2):249–254
27. Arditi A, YingLi T (2013) User Interface preference in the design of a camera-base navigation and wayfinding Aid. J Vis Impairement Blindness
28. HS MAHDI Image understanding using object identification and spatial relationship
29. Bairagi B, Dey B, Sarkar B, Sanyal S Selection of robotic systems in fuzzy multi criteria decision-making environment doi:10.1504/IJCSYSE.2015.067798, pp 32–42

A Twofold Subspace Learning-Based Feature Fusion Strategy for Classification of EMG and EMG Spectrogram Images

Anil Hazarika and Manbendra Bhuyan

Abstract We addressed an algorithm intuitively modeling multi-view information for pattern recognition application, specifically for electromyography (EMG) classification. The objective of the framework is to extract the low-dimensional embeddings (LDEs) inherent in multiple views that comprehensively represent the class information. We have shown that the algorithm is capable of providing robust solution to multitask learning relying on multi-view information. On two sets of EMG data, the learned LDEs comprehensively represent the multi-view information they were trained to represent, with consistency in performance across multiple sets of partitioned data sets. The significant aftermaths of the adopted learning strategy affirm the practical usability of the algorithm in healthcare applications for making correct diagnosis. Further, implementation of the algorithm for spectrogram image-based recognition is also of note.

Keywords Feature fusion · Canonical correlation analysis (CCA) · Electromyography (EMG) · Spectrogram image

1 Introduction

Multi-view information embedding (MVIE) plays a significant role in learning models (LMs) [1–8]. In LMs, MVI is extracted from different feature spaces, which have same or different statistical nature into a concatenation space that comprehensively represents the object for classification. Importantly, difficulty in LMs is the *curse of dimensionality* due to MVIE that enhances the learning parameters, and subsequently, it degrades the overall performance [9, 10]. This can be avoided by

A. Hazarika (✉) · M. Bhuyan
Department of Electronics and Communication Engineering,
Tezpur University, Assam, India
e-mail: anilh@tezu.ernet.in

M. Bhuyan
e-mail: manab@tezu.ernet.in

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_4

choosing an appropriate dimension of feature space that retains most of the energy contents through a given strategy. Much redundant and irrelevant information contained in the input feature space is necessary to remove prior to learning. Further, for inherent variations with complex data distribution pattern of input feature space, information embedded from single view feature space may not suitable for possible characterization of the object in supervised models. For example, an electromyography (EMG) signal is non-stationary, and further, the nature and exact data distribution patterns are not preknown, which makes difficult to accurately label the view or signal that can characterize the class-specific information. As a consequence, in supervised models, multi-view learning has a paramount impact for effective learning by reducing the bias in feature registration. Consequently, fusion technique, specifically feature fusion (FF), has been more and more demanding for multitask learning problems, such as remote sensing, pattern recognition, computer vision, robotics, and bioinformatics [11–15], and has been shown as a promising paradigm in classification.

The main objective FF is to make use of relevant integral sets of feature vectors by adopting feature projection technique (FP) such as principal component analysis (PCA) [9, 16, 17], linear discriminant analysis (LDA) [18, 19], uncorrelated LDA (ULDA) [20], and self-organizing feature map (SOFM) [9]. PCA well approximates the distribution of original space without considering the class information. Nonetheless, in MLs, wide separation margin among multi-group features is an important requirement that has been taken care of by LDA, ULDA, and SOFM. On the other hand, FP such as canonical correlation analysis (CCA) [21] and its extension multi-set canonical correlation analysis (MICCA) extracts low-dimensional feature vectors (FVs) from the projection surface by optimizing learning criteria. The integrity of this learning is that it finds unique feature representation (FR) based on the degree of proximity between two or multiple views, which significantly reduces the noise levels in the feature space too. It is worth noting that the learned patterns from projection surface may not have physical interpretation, but have much robustness in real-life applications, especially when the best feature sets are unknown [22].

Recently, CCA projections have been successfully applied in many pattern recognition applications, such as image [13, 22, 23], handwritten character recognition [2, 24, 25], and other applications [11, 26–28]. In general, CCA learning does not incorporate uncorrelated, i.e., orthogonal components between the views/variables. Recently, extended versions of CCA have also been introduced by adopting generalized learning to overreach the limitations [29]. Nonetheless, for non-stationary pattern analysis, inclusion of both components to the feature space through given strategy promises to play significant role in improving the generalized ability of models [14], specifically for EMG classification. In view of that, the study addresses a supervised subspace multi-view learning-based fusion strategy (MvFS) that can encode MVIE from twofold view subspaces to classify one-dimensional EMG signals and EMG spectrogram images that include templates from neuromuscular disorders (NDs), i.e., amyotrophic lateral sclerosis (ALS), myopathy, and control subject. Further, the algorithm is implemented for

spectrogram images generated from templates. Two major challenges are to avoid feature biasing and utilization of learned low-dimensional patterns to effectively explore the class information.

The etiology of NDs varies with muscle fibers, nerves, and tissues, from where diseases start off. As a result, it is difficult to track the origin of the disease depending on the symptoms raise off. Consequently, precise diagnosis of subjects suffering from NDs due to mild sickness has a paramount impact on the supervision of an effective treatment [20, 30]. ALS is a fatal and relentlessly progressive neurodegenerative disease which affects the motor neuron of the motor cortex, brain stem, and spinal cord and eventually leads to muscle weakness, loss of muscle mass, and inability to control movement [31]. Myopathy is due to damage of motor fibers that leads to muscle weakness [32, 33]. The cause of disease changes the morphology and physiology of motor units (MUs) of muscle, and thereby, significant changes in motor unit action potentials (MUAPs) and motor units (MUs) are observed through EMG signals [32]. Particularly, EMG recordings are acquired under maximum voluntary contraction (MVC). In MVC, a large number of MUs start to fire and eventually form a complex EMG pattern, known as interference pattern by superimposing of individual MUAP [34]. As a consequence, it is difficult to identify individual MUAP. Further, in mild case, abnormalities are often subtle in its respective patterns which intricate the conventional clinical practice. Therefore, researches have been focused on the development of automated computer-aided diagnosis system, i.e., quantitative EMG method (qEMG), by incorporating multiple template features. Aside from aforementioned advantages of qEMG, diagnosis aided by its findings obviates the need of *muscle biopsies* or *ultrasound*, and *magnetic resonance imaging* [20].

Over the decade, a number of qEMG methods by utilizing different combinations of feature vectors elicited from templates (i.e., MUAP) to define discriminant vectors have been developed [35–41]. With the development of decomposition techniques, MUAP-based methods (MUAPbs) that utilized the morphological features (i.e., duration, amplitude, and area) were also developed [42–44], some of which used their statistical measures [45]. An increasing trend in the usage of multi-resolution wavelet (WT) in EMG analysis is also of note [34, 46–53]. For instant, Subasi et al. [54, 55] developed discrete WT (DWT)-based techniques. In [32], a multi-view representation (i.e., time and DWT) extracting from MUAPs was adopted. Both time and DWT features are further truncated based on prior knowledge of MUAPs associated with studied classes and by using statistical measures, respectively. Recently, Doulah et al. [56] used DWT-based statistical features elicited from dominated MUAP (dMUAP). Nonetheless, the dMUAP is evaluated with guidance information of morphological patterns belonging to the studied classes, which requires skilled and experienced person (Figs. 1 and 2).

Despite being success of MUAP and DWT methods, they also pose certain limitations. First, use of wavelet functions often warrants the manual intervention as cautioned [20], and further, it classifies the classes considering WT coefficients from each level, which may not be feasible for all MUAPs. Further, the choice of coefficients is impractical. Next, MUAPbs treat all available MUAPs equally,

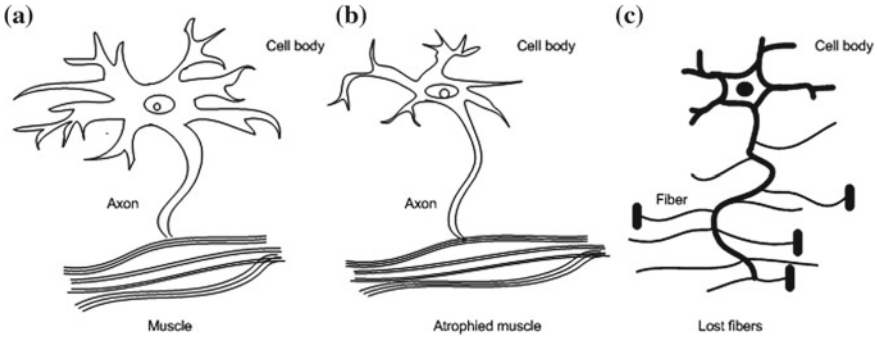


Fig. 1 Physiological changes associated with two classes of NDs in reference to normal pattern, i.e., **a** normal, **b** ALS, **c** myopathy

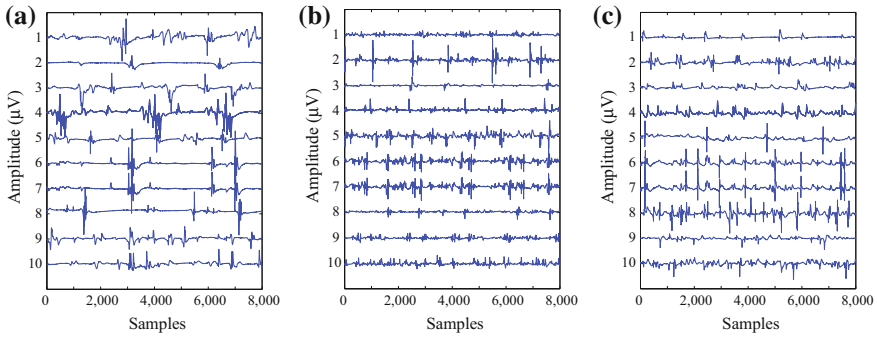


Fig. 2 Multi-view with multiple templates associated with three typical EMG patterns, **a** ALS (*left*), **b** myopathy (*middle*), and **c** normal (*right*) subject

while disregarding the non-stationary MUAP that could lead to false findings. In general, existing strategies usually focus on utilizing different combinations of feature vectors extracted from the specific diagnosed template. This approach simplifies the problems and subsequent learning models; nonetheless, it may cause feature biasing since data distribution structure within the template is not known due to wide variations of templates which could result from (a) nature of disease; (b) subclass of disease; (c) nature of force to recruit the muscle; and (d) inaccurate clinical setting. In view of that, multi-view with twofold learning can bring more prior information about the disease associated with the subject to the learning process than the single view approach. The adopted learning technique, i.e., MvFS with encoded information the subspaces, followed by vector transformation and fusion for classification shown in Fig. 3.

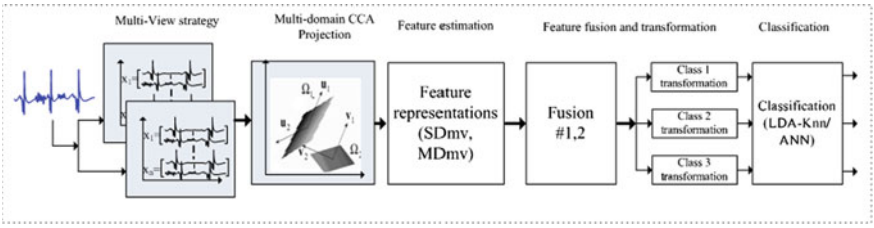


Fig. 3 Proposed multi-view feature fusion strategy (MvFS)

2 Method: Multi-view Template-Based Analysis

Figure 3 shows the work flow diagram of learning model, which includes (a) multi-view strategy, (b) multi-domain projection (twofold), (c) MVI patterns (i.e., SDmv and MDmv), (d) fusion and transformation, and (e) LDA-KNN and ANN classification. In what follows, we will elaborate each step in detail including database description.

2.1 Database Description

To learn and validate the algorithm, two sets of data, (a) publically available database of 250 recordings (50 ALS, 50 myo, and 150 nor) [57] and (b) data set that collected from authors institute, Guwahati Neurological Research Centre (GNRC) database that includes of 60 recordings, 20 from each class are considered. The first data set consists of ten normal subjects—six males and four females (age-group: 21–37 years, mean-27.7 ± 4.5), eight cases with ALS—four males and four females (age-group: 35–67 years, mean-52.8 ± 11.8), and seven myopathy subjects—five males and two females (age-group: 19–63 years, mean-36.3 ± 14.6). Second set data include four ALS subjects—three females and one male (age-group: 38–52 years, mean-43.5 ±7), four cases of myopathy (age-group: 42–59 years, mean-47.5 ±7.8), and healthy control—two females and two males (age-group: 26–34 years, mean-29.3 ± 3.4). None of the healthy subjects had signs or history of NDs.

2.2 CCA Learning

Normally, the frequency of EMG signal falls in the range of 0–1 kHz, and its dominated energy concentrated in 20–500 Hz [9]. So, to begin with, templates are filtered by designing a Kaiser window-based low-pass filter with the order $n = 20$, cutoff frequency f_c = 0.5 kHz, and sampling frequency f_s = 1 kHz [58], followed

by re-sampling of templates to even number for ease of implementation. It is pertinent to mention herein that for differences in population (age, disease profiles, etc.) or measurement setting, the templates in two different populations under same study group might be significantly different [1]. So, to desist the feature biasing, the subject groups are partitioned into subgroups and subsequently templates registered into multi-template vector $X_1(p) = [x_1(p), \ldots, x_q(p)]^{\mathrm{T}}$, with $p = 1, \ldots, 2n$, which is further uniformly decomposed into multi-view vectors $X_i(n_i)$ as follows:

$$X_i(n_i) \in \mathbb{R}^{q \times p/L} = [x_i(k,1), \ldots, x_i(k, p/L)] \tag{1}$$

Here, L is the number of sequences, and each row $x(k.,)$ represents the consecutive signal sequences. It is assumed that sequences are symmetrical in pattern. Here, symmetry in the context of MUAP pattern is investigated via EMGLAB [59], which is an efficient tool to decompose as well as to count the dominated MUAPs in a particular signal sequence.

CCA finds intrinsic representation by integrating two views based on their correlations. It searches two sets of vector, one for each view, such that they are maximally correlated in the mapping space. The low-dimensional feature vectors (FVs) are inherent in view to carry unique characteristic [21, 26]. Further, the correlation between two vectors in embedded space indicates the degree of proximity. In view of classification, we propose subspace learning technique, i.e., mCCA that learns subspace resorting to multi-view vectors.

Taking two views, X_1 and X_2 from Eq. (1), mean of each row from the view matrices is removed to make centered data matrices. CCA finds two projection matrices A and B, one for each view in subspaces Ω_1 and Ω_2 by optimizing the learning criteria as follows:

$$\max_{A_i \neq 0, B_i \neq 0} \frac{A^{\mathrm{T}} \Sigma_{12} B}{\sqrt{(A^{\mathrm{T}} \Sigma_{11} A)(B^{\mathrm{T}} \Sigma_{22} B)}} \tag{2}$$

In formulating our problem, we also assume that $A^{\mathrm{T}} \Sigma_{11} A = B^{\mathrm{T}} \Sigma_{22} B = 1$. Hereby, optimization criteria in Eq. (2) is reduced as follows:

$$\max_{A_i \neq 0, B_i \neq 0} A_i^{\mathrm{T}} \Sigma_{12} B_i \tag{3}$$

subjected to $A_i^{\mathrm{T}} A_i = B_i^{\mathrm{T}} B_i = 1$ and $A_i^{\mathrm{T}} A_j = B_i^{\mathrm{T}} B_j = 0, i \neq j$. Further, $\Sigma_{12} (= X_1^{\mathrm{T}} X_2)$ and Σ_{21} are two cross-covariance matrices of X_1 and X_2, respectively. Formulations in Eqs. (2) and (3) attempt to find the eigenvalues corresponding to the top eigenvalues of following generalized equations:

$$X_1 X_2^{\mathrm{T}} \left(X_2 X_2^{\mathrm{T}}\right)^{-1} X_2 X_1^{\mathrm{T}} A = \alpha^2 X_1 X_1^{\mathrm{T}} A \tag{4}$$

$$X_2 X_1 \left(X_1 X_1^{\mathrm{T}}\right)^{-1} X_1 X_2 B = \alpha^2 B X_2 X_2^{\mathrm{T}} \tag{5}$$

Further, both equations are interrelated, and solution one yields the other. However, the introduction of regularization restricts the limitation of overfitting and singularity. To solve eigenvalue problems, singular value decomposition (SVD) is an efficient technique, which involves $d \times d$-dimensional matrix. Assume that X_1 and X_2 form unitary orthogonal bases for two linear subspaces. Let the SVD of $X_1^{\mathrm{T}} X_2 \in R^{d \times d}$ be

$$X_1^{\mathrm{T}} X_2 = U \Lambda V^{\mathrm{T}} = [U_1, U_2]\mathrm{diag}(\Lambda_{\mathrm{d}}, 0)[V_1, V_2]^{\mathrm{T}} = U_1 \Lambda_{\mathrm{d}} V_1 \tag{6}$$

where U and V are two left and right singular orthogonal matrices of Σ_{12} and Σ_{21}, i.e., $UU^{\mathrm{T}} = VV^{\mathrm{T}} = I_{\mathrm{d}}$. Further, mathematically, it infers that U_2 falls in the null space of X_1 (i.e., $X_1^{\mathrm{T}} U_2 = 0$), which indicates that it is uncorrelated component [29]. Additionally, by reducing the dimensionality based on the degree of correlation between two views, learning parameters can be further reduced, which in turn reduces the computational cost. Singular values represent canonical correlations, and the associated vectors are given by

$$A = X_1 U_1 = [A_1, \ldots, A_{\mathrm{d}}], \quad B = X_2 V_1 = [B_1, \ldots, B_{\mathrm{d}}] \tag{7}$$

Diagonal elements of Λ_{d} that are in descending order measure the strength of FVs. The noteworthy factor is that the correlation between same index pairs is nonzero, due to orthogonal nature of FVs, e.g., correlation, $r = 0$, for pair A_1, B_2. Intuitively, first few pairs show significant proximate behavior (Fig. 4), i.e., they well capture the idiosyncratic information from pair variable. Further, each FV of one set akin to corresponding FV in other set notwithstanding the data variation. Thus, high-dimensional views of each object are well confined to the subspace that preserves most of the energy contents by making use of subspace learning. Further, in the proposed multi-view learning, i.e., mCCA, overfitting of model and singularity of both scatter matrices are averted by dint of regularization [29]. Nonetheless, it does not mutate the projections as Eq. (2) is independent of scaling to A and B. This learning technique is used in aforementioned strategy for subsequent multitask learning.

2.3 Variability Measurement (VM)

Two inherent properties of non-stationary signals are variations and complex data distribution which make difficult the feature selection process. Further, it causes feature biasing in the learning model. As a consequence, VM followed by feature selection is the one of the key strategies to achieve the optimal level of model performance. Statistical technique [60–62] is one of the key techniques to

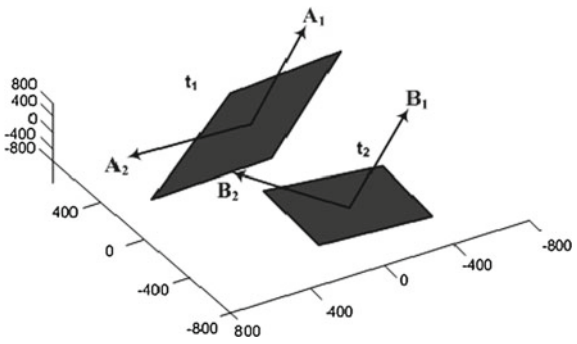


Fig. 4 Linear subspace with two sets of vectors, (A_1, B_1) and (A_2, B_2). However, different index pair vectors, i.e., (A_1, B_2), (A_2, B_1) are orthogonal behavior

characterize the variations of templates and subsequent feature selection, which is briefly as follows:

$$\text{mCCA} \rightarrow \begin{cases} \{X_1, X_2\}^{l_1}, \ldots, \{X_{j-1} X_j\}^{l_1}, & \text{for } h1 \\ \{X_1, X_1'\}^{l_1, l_2}, \ldots, \{X_j X_j'\}^{l_1, l_2}, & \text{for } h2 \end{cases}$$

where $h1$ and $h2$ stand for VM within- and inter-subject templates. Further, X_j and X_j' are multi-view vectors extracted from templates of two subject groups, l_1 and l_2, respectively. The statistical measures (i.e., r, A_i, B_i) between two views or multiple views signify the nature of variations [61]. The nature of projected vectors is relative to the correlations. Further, it indicates the degree of proximity between two views. Hereby, it is reasonable to analyze the variations in terms of r. Similar value of r s, i.e., $r_{12} \approx, \ldots, \approx r_{j-1j}$, indicates high correlation among the signal frames or multi-views (Mvs) and lesser variation, i.e., highly stable, and vice versa. In case of similar r, associated FVs will have similar nature. Additionally, the overall correlation (OC) ($\Gamma_{i-1j} = \sum_{i=1}^{d} r_i^2$) over an Mv pair to the next indicates quantitative nature. However, appropriate feature selection based on OCs is a difficult task. It is worth indicating that variations can also be characterized in terms rms values estimated over the views [56]. However, feature selection depending on the rms is a difficult task. Following the analysis, suitable dimension of FVs is extracted for learning.

To demonstrate the VM, the three subject groups, namely ALS, myopathy, and normal, are partitioned into a set of subgroups (Table 1), and subsequently, mean variations of views in term of correlation (i.e., r_μ) over different subgroups are estimated taking $q = k = 10$ templates from five subjects under each subgroup. Figure 5a shows the Mvs in A35MT5 and A35MV5 are highly correlated as compared to A56MAP0.5 and A67MB0.5 with different disease profiles (i.e., duration); however, the fluctuations of r_μ over the FVs in A67MB0.5 (d1) are lesser than other groups. Thus, it infers that signals are more stable in former cases, even under same disease group. Further, the variations r_μ tend to uniform for $d > 7$ in all the cases. It is worth mentioning that both A35MV5 and A35MT5 groups are

similar, except their recording sites, indicating that the consistency might not depend on the sites but on disease profile. Further investigations on different age subject groups with common disease profile confirm the fact. Consequently, in case of ALS group, under mild or less severe cases, the fluctuation of Mvs is more prominent than severe cases, irrespective of gender level.

Unlike ALS, the Mvs in myopathy group with higher age and disease duration are weakly correlated, and further, the nature of r_μ profiles are similar with no significant exception, which indicates that the Mvs are more asymmetrical in b2, c2, and d2 as compared to a2 (Fig. 5b). However, in normal groups, the lower correlations are observed with higher age subject groups (c3 and d3), whereas it is high with lower age subject groups (a3 and b3). Further, the variations within templates are significant toward the extreme age of subjects, but in case of the middle age, it is not so prominent (Fig. 5c). That is, in comparison, myopathy subject group shows asymmetrical behavior irrespective of their disease duration. Further, the extent of variation depends on the type and nature of diseases. This analysis is carried out by taking 10 to 20 templates; however, no significant exceptions are observed.

Further, this analysis is performed over inter-group templates, i.e., templates belonging to different subgroups of same groups, by pairwisely correlating among Mvs (i.e., iMvs) with the aid of second formulation as stated above. Figure 6a indicates that iMvs in A1 are smaller in comparison with other subject groups, especially for $d > 2$. Moderate level of correlations is observed in B1 and C1, whereas it is maximum in between two extreme age subject groups (D1 and E1). Thus, it infers that two extreme ages of subjective templates, although having different disease profiles, are highly symmetrical in behavior. Further, the subjective variations of templates are relatively small in two subject groups with same age and disease profile (Fig. 6b). However, it is insignificant for $d > 7$ for all the cases. Myopathy groups (i.e., Fig. 6c) show no significant remarks. However, the iMvs in D2 are highly correlated, and hence, the inter-subject profiles are more stable. In normal groups, the iMvs are highly correlated with A3 and B3; i.e., templates are more stable in comparison with the templates in C3 and D3, respectively. Thus, it shows that inter-subject templates in myopathy are more stable, although variation in Mvs. However, in normal, the lower age subjective templates are more stable as compared to higher age templates. Thus, it remains in favor of the strategy formulations for feature extraction in the method advocated in this article.

2.4 Feature Extraction and Dimension Reduction

To extract FVs via mCCA, each consecutive pair views are embedded to subspace for optimized learning. In brief, the adopted subspace learning for feature extraction may be written as follows:

Table 1 Study subject groups and its subgroups

Subject groups	Subgroup
ALS	A56MAP0.5, A35MT5, A35MV5, A67MB0.5[a]
Myopathy	M28MT12, M44MB26, M41FB2, M26MB1
Normal	H23MB0, H26MB0, H29MB0, H37MB0

Note [a]Class-age-sex-muscle-disease duration (year)

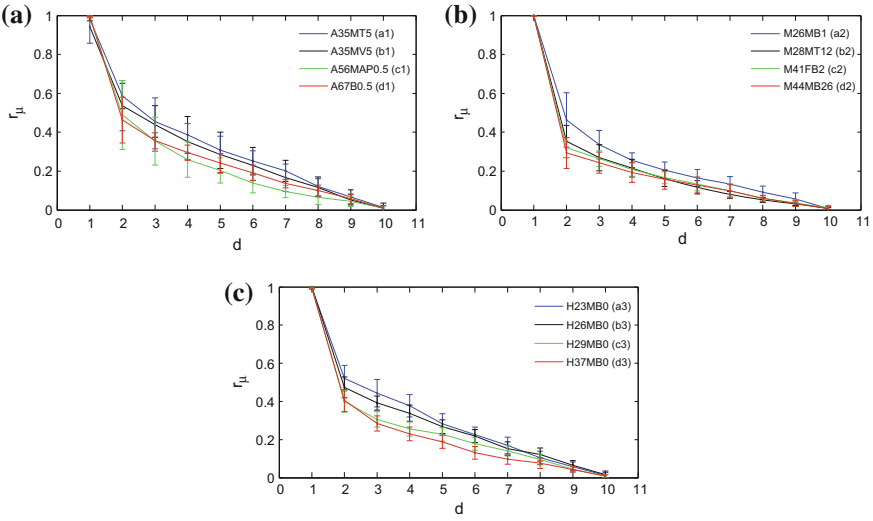


Fig. 5 VM within templates in terms of mean correlations (r_μ) with SDs **a** ALS; **b** myopathy; and **c** normal subject groups

$$f\left(x_{i,j}\right) \rightarrow \{\{\Omega_1, \Omega_2\}, \ldots, \{\Omega_{i-1}, \Omega_i\}\} \tag{8}$$

where $f(x_{i,j})$ is the multi-view feature space. In the adopted scheme, an equal number of projected FVs is extracted from each pair of subspace $\{\Omega_{i-1}, \Omega_i\}$.

The learned subspace through Eq. (8) does not appraise the uncorrelated information affiliated with them. A curative measure to overreach the limitations is the inclusion of statistically independent realizations (i.e., FVs) extracted via DWT, i.e., DWT + subspace, as synchronization elevates the generalized ability of learning models [14]. Keeping in view, next, second pair of FVs from orthogonal subspace is extracted, which are later integrated with its previous sets to enhance the learning.

Wavelet decomposes the signal into a set of basis vectors, called wavelets that are obtained by dilations, contractions, and shifts of a unique function called wavelet prototype. It offers temporal localization of signals' spectra; i.e., it provides the occurrence of frequency bands in particular time intervals. Basically, it analyzes the signal in different frequency bands by decomposing the signal into coarse

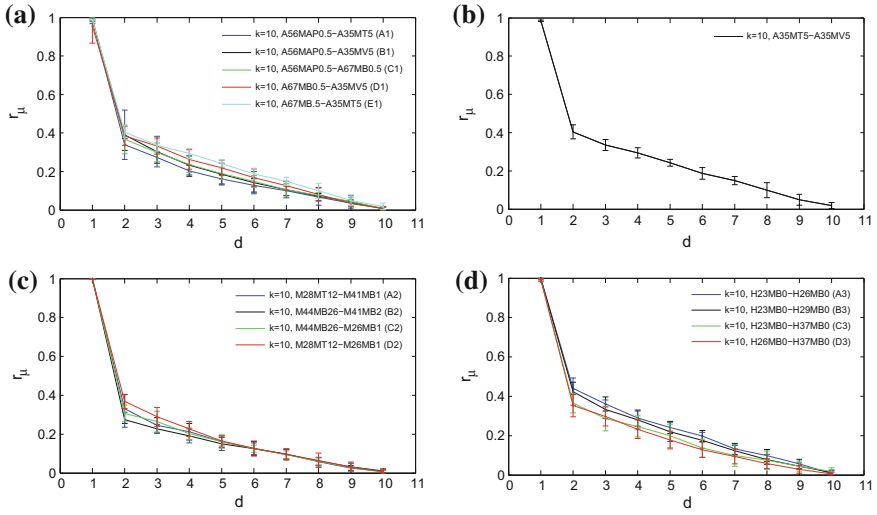


Fig. 6 VM in inter-subject templates in terms of r_μ with SDs **a** multiple ALS; **b** ALS (single); **c** multiple Myo; and **d** multiple nor subject groups (here, k represents number of observations (i.e., n)

approximation (low-frequency components) and detail coefficients (high-frequency components) [63]. Hence, it acts as filter; however, each level of decomposition is accomplished by downsampling two and a half frequency band spanned. The high-pass and low-pass filter outputs are given below:

$$y_{\text{high}}[k] = \sum X_j[n]h(2k - n) \tag{9}$$

$$y_{\text{low}}[k] = \sum X_j[n]g(2k - n) \tag{10}$$

The study has been performed with prototype Daubechies wavelet of order 2 (db2) which is efficient for EMG classification, and subsequently, low-frequency components (CAs) from each of two levels are fixed for analysis [56] as depicted in Fig. 7, wherein approximation and details are juxtaposed for each group to enable the understanding of DWT template selection mechanism for projection. Then, its consecutive pair projections are performed, and subsequently, FVs are extracted from both independent subspaces, i.e., one directly from subspace of Mv input space and other from DWT-transformed subspace for learning. The estimated feature representations are outlined in Table 2. Nonetheless, the dimensionality of the compact views is further reduced by imposing threshold selection criteria, following the VMs. Note that the adopted learning technique incorporates both optimization and singular value decomposition that notably reduce the *curse of dimensionality*. Additionally, regularization avoids the issue-related singularity.

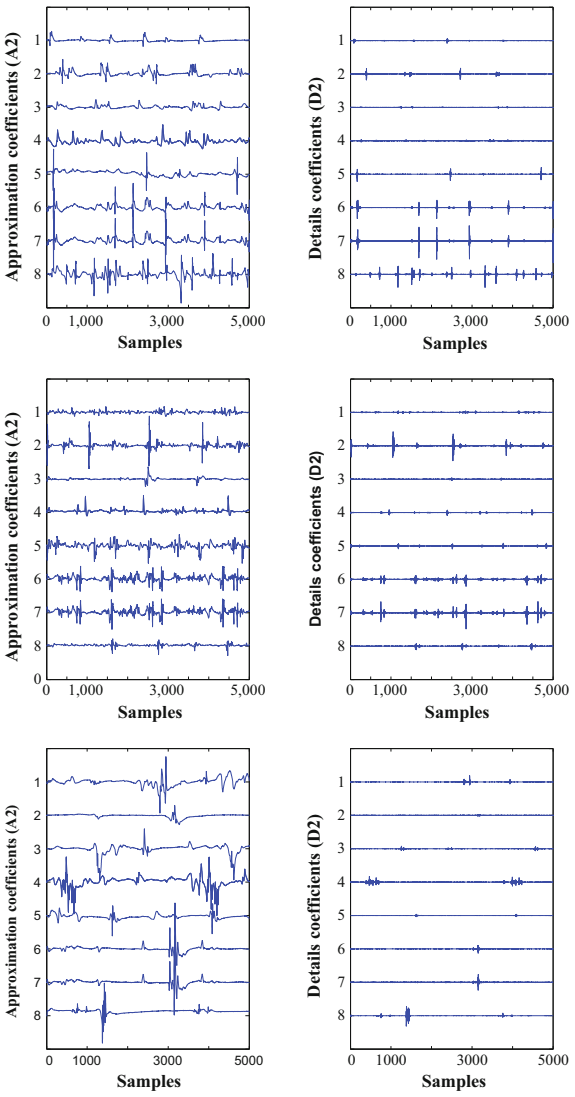


Fig. 7 For example, multi-view DWT strategy, *Left* approximation coefficients (A2s), when the DWT is applied on multi-template vectors (MTVs) with prototype db2 wavelet from three typical EMG pattern, normal (*top*), myopathy (*middle*), and ALS (*bottom*) subject. *Right* details (D2s) of respective MTVs

Table 2 Feature representations for learning

# Feature type	Feature representations
$d_i \times p(\lambda_1)$	SDmv
$R \times d_i \times p\ (\lambda_2)$	MDmv
$R \times N \times d_i \times p\ (\lambda_3)$	MDmv

Note: R Number of domain, N number of modality, and p number of views

Thus, it significantly reduces the complexity as well as computational overhead of model.

As could be seen from the analysis, most of the local information is confined within few FVs which capture characteristic information inherent in their respective templates. As a consequence, the learned pattern (Table 2.) reduces the bias in feature registration and complexity of the model. Note that for higher value of d, the FVs start acquiring orthogonal behavior which in turn increases the computational burden (Fig. 8). The termed feature pattern λ_1 promises to provide unique views in single domain space, while λ_2 gives twofold representation of views correlating both orthogonal and correlated components. Nonetheless, for multiple modality analysis, such as fMRI, sMRI, and EEG [11], λ_3 retains the quality of feature pace.

2.5 *Feature Fusion and Classification Strategy*

In supervised classification methods, the learning patterns are the members of a predefined set, and subsequently, FVs are searched in a supervised manner. The learned patterns are fused via parallel and serial fusions, which play significant roles in pattern recognition [13] (see appendix) to extract discriminant representation, A_p and B_p for respective groups. Afterward, patterns are re-labeled by introducing class indicator parameter γ_c to avoid feature overlapping due to their inherent similarities, as shown in Fig. 7

$$A_i(B_i) = A_\mathrm{p}\left(B_\mathrm{p}\right) + \gamma_\mathrm{c} I \& \forall i = 1, \ldots, C \tag{11}$$

with $\gamma_\mathrm{c} = 0$ for nor and $\gamma_\mathrm{c} \neq 0$ for others. Pattern vectors are further applied to supervised LDA, which transforms them to a well-defined coordinate system, preserving the class information. It helps finding the optimum decision surface by enhancing the separation margin among multi-group features exploiting the class information. Thereby, it boosts the learning while performing the classification task. That is, it minimizes the within-class variance and maximizes between-class variance, to attain optimal discrimination. Suitably, within-scatter matrix S_W and between-scatter matrix S_B can be defined as follows:

$$S_\mathrm{W} = \sum_\mathrm{c} N_\mathrm{c}(\mu_\mathrm{c} - \mu)(\mu_\mathrm{c} - \mu)^\mathrm{T} \tag{12}$$

$$S_\mathrm{B} = \sum_{x \in C_\mathrm{c}} (x - \mu_\mathrm{c})(x - \mu_\mathrm{c})^\mathrm{T} \tag{13}$$

where μ_c, μ, and N_c stand for class mean, overall mean of the entire sample set, and number of samples in class C, respectively. The learned patterns are applied to the classifier, i.e., labeled as LDA-kNN. Note that use of LDA in classification task avoids the iterative training as well as under or overfitting, which in turn leads to

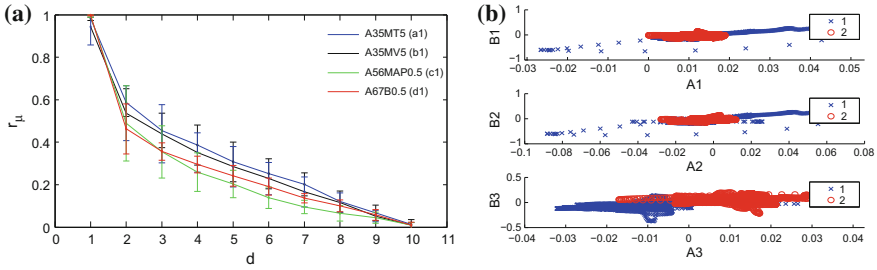


Fig. 8 **a** Mean correlations of views and **b** typical nature of FVs [A ($\circ$) and B ($\times$)]. Note that for higher value of d, multi-group features start overlapping at the decision surface. FVs $[A_1, B_1]$ and $[A_2, B_2]$ with $r = 0.99, 0.58$ are highly correlating, while in weekly paired case, such as in $[A_3, B_3]$ with $r = 0.43$ the data points start deviating from the mean, i.e., start acquiring the orthogonality character, indicating less informative

higher generalizing ability of the model [20]. It is worth noting that kNN finds the Euclidian distance function $D(x, x_i) = \sum_d (x_j - x_{ij})^2$ between the feature space in test set and neighboring healthy and pathological patterns in the training set. The pattern from the test set is categorized based on the class labels of closer patterns. In view of classification, the data set is divided into three subsets. The best sets of FV that correspond to the eigenvalue of matrix $S_W^{-1} S_B$ with significant group difference in paired t-test are applied to the classifier for segregating three classes—normal, myopathic, and ALS (Fig. 9).

2.6 *Classification Performance*

The performance of the classifier is measured in terms of biomedical markers—accuracy (Ac), sensitivity (Sn), specificity (Sp), negative predictive value (NPV), positive predictive value (PPV), and F-measure (F) by formulating 3×3 confusion matrix. Sensitivity (Sn) measures the degree of positive cases (e.g., ALS, myopathy), while specificity (Sp) indicates the degree of negative case, i.e., normal cases correctly classify. Other relative terms indicate the balance of classifier (appendix). The algorithm is implemented in MATLAB (*The MathWorks, Inc., Natick, USA*) on an Intel (R) Xeon (R) machine (Precision T3500) with processor 2.8 GHz and 8 GB of RAM.

The evaluation of proposed method is conducted on two-binary and three-level classification problems, including ALS versus normal classification, myopathy versus normal classification, and ALS versus myopathy versus normal classification. Further, k-fold cross-validation strategy is to evaluate the performance of our method, which is further compared with reported methods akin to our approach. Followed by VM on three subject groups, the dimensionality of feature spaces is

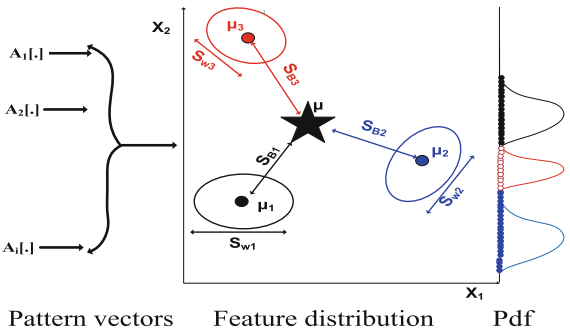


Fig. 9 Working principle LDA, including S_W and S_B for $C = 3$ with dimension 2, i.e., $C - 1$ (*pdf* probability density function)

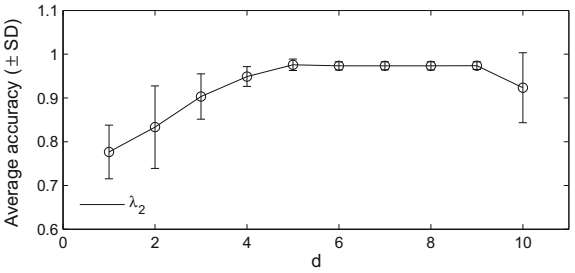


Fig. 10 Performance in terms of average accuracy (±SD) with variation of d (ALS vs. normal)

reduced by dint of threshold setting (i.e., 5). The use of common threshold avoids the issue related to dimensionality. The learned patterns are fused through two feature fusion strategies, followed by pattern transformation. Further, in estimating the transformed patterns, we set $\gamma_c = 0, 10, 20$ empirically for normal, myopathy, and ALS. However, in multi-class learning, wide range selection of γ_c is best suited. Later, pattern vectors are fed to LDA from where the best features corresponding to the higher eigenvalue of matrix $S_W^{-1}S_B$ are estimated by using learning criteria with significant group difference in paired *t*-test ($p = 0.05,\ 0.08$) through which the classifier can achieve the best average accuracies between k-fold cross-validation, *k* being two. We further ensure the incorporation of all FVs for training and validation. In sequel, the classifier aftermaths in segregating templates in two binary classes—ALS versus normal and myopathy versus normal and one ALS versus myopathy versus normal—are portrayed in terms of *average parameter values*[1] and from k-fold cross-validation. It is worth indicating that study mainly focuses on twofold MDmv feature, λ_2, due to non-stationary nature of the template. However, for better comparison, algorithm is also used λ_1. Further, results of cited scheme with two fusion strategies are narrated separately.

In first case, the algorithm intended for isolating the classes yields an accuracy up to 87.7, 97.6, and 87.9, 96.6%, with λ_1 and λ_2 in ALS versus normal and

[1]The average parameter values are estimated from confusion matrices in terms of mean μ and standard deviation σ from *k*—cross-validation techniques.

myopathy versus normal, respectively. As shown in Figs. 11 and 12, the proposed method with λ_2 achieves significant enhancement of parameter values. For instance, in terms of the classification accuracy, it achieves an improvement of $\sim 10\%$ in ALS versus normal classification and an improvement of $\sim 9\%$ in myopathy versus normal classification. The poor performances with λ_1 are presumably due to elimination of independent components which has been improved by incorporating second realization (i.e., FVs) in λ_2. The significant enhancement of relative parameters (i.e., Sp, Sn) which indicate the strength of classifier in identifying the individual class supports the efficacy of multi-view learning strategy. Nonetheless, the classification performance may vary with the dimensionality of feature space. For example, in Fig. 10, the classifier attains maximum value for particular values of d and then declines toward the lower end. With λ_2, the classifier avoids counterfeit error in classification that manifested from the accuracy label, i.e., twofold strategy significantly helps in extracting the unique class-specific information to make utilize for the class discrimination. The classifier achieves an accuracy up to 98.8% with λ_1 in case of three class subject groups. For visual appreciation, significant parameter values are outlined in Table 3. It is to be mentioned that the classifier attains promising individual accuracy in categorizing healthy subjects. Note that each binary group consists of one positive subject group (i.e., ALS or myopathy) and one negative subject group (i.e., normal). Thereby, single set of parameters is estimated for each binary group, while in three class group, two sets of parameter values are estimated [64]. For example, the sensitivity of ALS SnA $= 89.3 \pm 3.8^a$ with prefix 'a' and of that of myo SnM $= 87.7 \pm 7.9^b$ with prefix 'b' in λ_1 feature vector (Table 3). Thus, effectiveness of Mv learning is apparent from the significant parameters values.

Next, the performance of second fusion strategy (i.e., B_p) is depicted in terms of mean error (Fig. 12). From Fig. 12, it is clear that large error in second fusion strategy is due to high dimensionality. Nonetheless, it can be effectively resolved by partitioning the c-class larger subject groups into c_1 number of subgroups and then performing parallel classification task. Additionally, similar recognition rates are observed in small subject groups under non-regularization, i.e., $\beta_1 = \beta_2 = 0$. Nonetheless, degraded performance (i.e., 83.3% in λ_2 and 86.6% in λ_3) is observed in large scale groups. Thus, introduction of regularized parameters helps retaining the generalizing ability of the subsequent learning models, specifically in high-dimensional problems. However, choice of moderate value is an important requirement for effective learning without losing the information on scatter matrices $X_1X_1^T$ and $X_2X_2^T$ [29].

Thus, our subspace learning method provides a better way to make use of global and orthogonal information of template via MDmv feature representation to help promoting the ND classification. The advantage of our method could be due to the fact that our method utilizes the feature by integrating two views of multi-dimensional vectors in both domains followed by feature fusion strategy and by setting up of best decision surface by LDA projection.

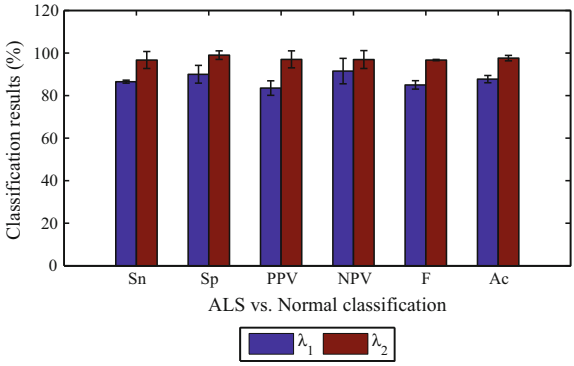


Fig. 11 Classification performance of the algorithm in terms of mean values with SD (±) as error bars in ALS versus normal classification

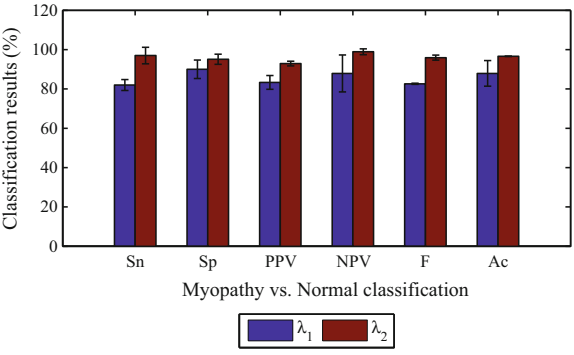


Fig. 12 Classification performance of the algorithm in terms of mean values with SD (±) as error bars in myopathy versus normal classification

Table 3 Average results in normal, myopathy, and ALS subject groups

# Feature	Sn in %		Sp in %		PPV in %		NPV in %		F in %		Ac in %	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
λ_1	89.3	3.8[a]	91.2	1.4	92.4	0.8[a]	88.7	1.2	90.9	1.6[a]	89.9	5.0
	87.7	7.9[b]			91.2	2.9[b]			89.4	5.6[b]		
λ_2	98.0	2.8[a]	99.0	1.4	97.1	4.2[a]	99.4	0.9	97.5	3.5[a]	98.8	2.0
	98.0	2.8[b]			98.9	1.5[b]			98.5	2.1[b]		

[a]Sensitivity of ALS and [b]sensitivity of myopathy

3 Method: Multi-view Spectrogram Image Analysis

CCA in images or character recognition is popular in machine learning community [2, 13, 22–25]. Further image-based clustering nowadays becomes popular in different domains, specifically in medical field due to its inherent capability of visual representation of information associated with the objects [65–71]. Here, we address clustering strategy by deploying multi-view images produced by time–frequency

representation of signal, such as spectrogram. Spectrogram-based recognition techniques are widely used in sound, speech, etc. [72–77]. It provides visual signature of spectral representation, from where feature can be extracted for pattern clustering. The motivation stems from the fact that different templates or segment of a timecoarse can produce particular spectrogram, which gives characteristic information for discrimination. The spectrograms are generated from signal, which are then pairwisely mapped to well-defined coordinate system to capture characteristic information in term of FVs based on their degree of proximity. Finally, extracted FVs are fused to form discriminant pattern vector. More, specifically, MvFS based on spectrograms (i.e., MvFS-sp) is applied for classification.

3.1 Spectrogram Generation

The steps involved in spectrogram generation from signal are illustrated in Fig. 13. The windowed discrete Fourier transformation (WDFT) of $x[n]$ is as follows [72]:

$$X_t(k) = \sum_{i=1}^{N-1} x[n]w[n] \exp\left(-\frac{2\pi i}{N} kn\right), \quad k = 0,\ldots,N-1; \tag{14}$$

where N is window length, $w[n]$ is the Hanning window function, and k corresponds to the frequency $f(k) = kf_s/N$, where f_s is the sampling frequency in Hertz. The linear and log-power representations to generate spectrograms are, respectively, as follows:

$$Z_{\text{Liner}}(k,t) = |X_t(k)| \tag{15}$$

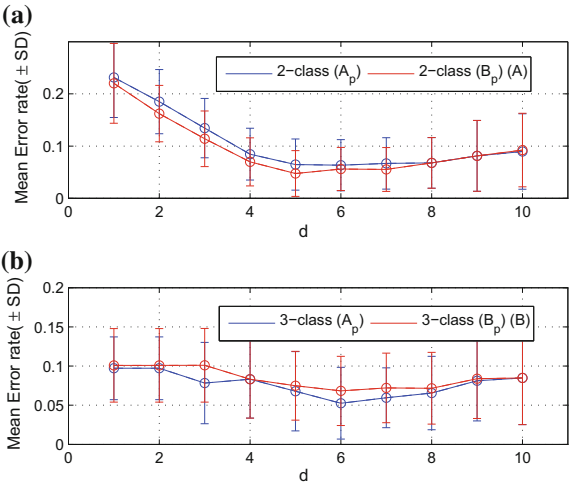


Fig. 13 Average error rates (±SD) of first fusion, i.e., A_p and second, i.e., B_p in two and three class subject groups

$$Z_{\mathrm{Log}}(k,t) = \log(Z_{\mathrm{Liner}}(k,t)), \quad \forall t; \tag{16}$$

The time–frequency matrix is normalized into a grayscale intensity image, with the range scaled between [0, 1] (Figs. 14 and 15).

$$\gamma(k,t) = \frac{Z(k,t) - \min(Z)}{\max(Z) - \min(Z)} \tag{17}$$

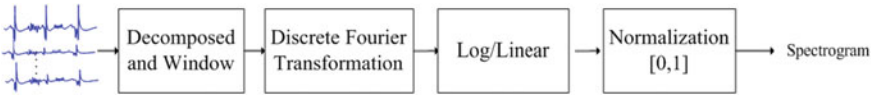


Fig. 14 Spectrogram image generation steps

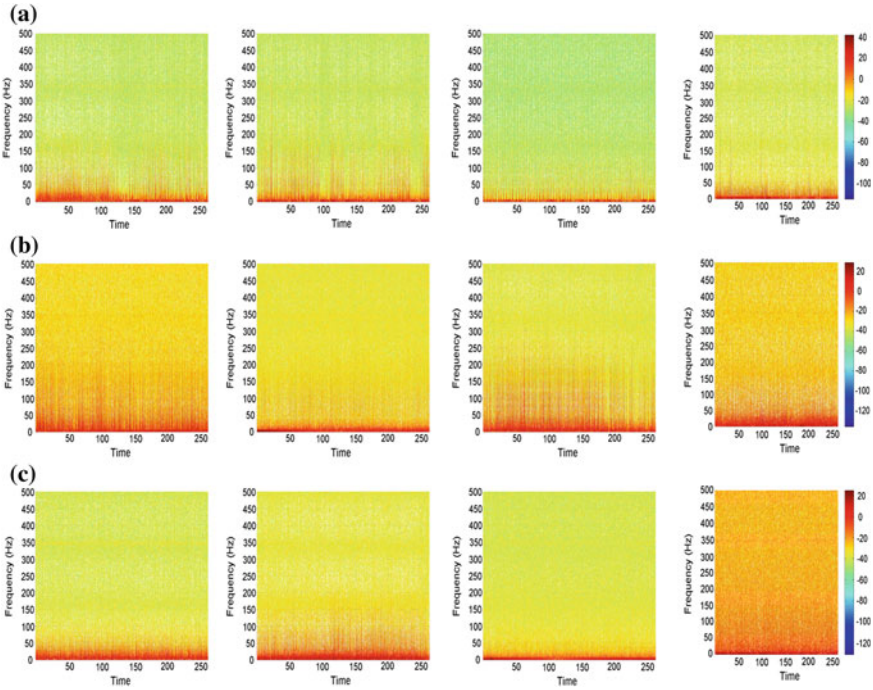


Fig. 15 Generated spectrograms for three typical patterns, **a** ALS, **b** myopathy, and **c** normal subjects. For each class, spectrograms are created from multiple templates considered, wherein each of image signifies single template of class pattern

3.2 Classification Performance

Each spectrogram has a particular statistical property and describes its variation over time and frequency regions associated with the class [72]. Thereby, it provides significant source of information for class discrimination. To extract image feature similar to aforementioned strategy, we transformed the feature space to subspace. Following the correlation measures, we set $d = 8$ as a common threshold for analysis, and subsequently, extracted feature space (i.e., FVs via λ_1) is concatenated through parallel fusion strategy as in Eq. (11) to derive the generalized pattern vectors for the classes. Afterward, model validations are performed with four cases of ALS, myopathy, and normal subject. In the first case, group-representative fused patterns are applied to the combined classifier, i.e., LDA-KNN. Introduction of LDA ensures the further enhancement of the separation margin among the multi-group FVs in the decision surface, and subsequently, extracted FVs that correspond to the significant eigenvalue of $S_W^{-1}S_B$ are applied KNN model, which yields an accuracy level up to 91.67%. On the other hand, pattern vectors that correspond to three different classes are trained with multilayer perceptron (MLP). The input layer of MLP is based on the pattern vectors. The network is constructed with two hidden layers, and the output layer has three neurons for three classes to be categorized. The network structure is determined by trial-and-error method, and further, selection criteria are based on the convergence of learning error, i.e., mean square error (MSE). Further, it is worth mentioning that the input pattern of MLP is the normalized value and initial weights and bias were selected from a uniform distribution with a mean and variance of 0 and 1, respectively. The learning process was ended when the performance, i.e., MSE per iteration, was sufficiently small. The trained classifier yields an accuracy level, i.e., 83.33% at optimum level of MSE. It is to be mentioned herein that the both classification strategies take larger computation time as compared to the previous approach. Further, low recognition rate may presumably due to large dimensionality of feature space and dissimilarity among the multi-view spectrograms that are relative to the morphological patterns. Nonetheless, adoption of parallel ensemble feature extraction technique can reduce the computation overhead of the method. Thus, it encourages further research pursuits for further enhancement of performance in multitask learning problems.

3.3 Comparison with State-of-the-Art Methods

Recalling the results are yielded by state-of-the-art techniques akin to our method, a comparison is made to explore the suitability of our approach, specifically the first approach that utilizes multi-view template information for class discrimination, with the rest. Further, details of reported methods, including classifiers, type of feature, and study groups, with subjects involved are outlined in Table 4.

Table 4 Comparison of performance with the state-of-the-art techniques

# Classifier	# Feature type	# Study groups/subject	# OAc in %
SVM [16]	F[a]	3/27	92.55
F-SVM [55]	DWT-SF	3/27	93.50
Ensemble-SVM [79]	DWT-SF	2/27	97.00
PSO-SVM [54]	DWT-SF	2/27	97.40
RF [78]	DWT-SF	3/25	96.67
MvFS-LDA-KNN	Mdmv-SF[b]	3/37	**98.80**

Bold face indicates the correct classification
Note [a]Frequency, [b]MDmv-statistical feature (*SF*)

Support vector machine (SVM) in [16], random forest (RF) in [78], and fuzzy-SVM (F-SVM) in [55] classified neuropathy, myopathy, and healthy subjects using statistical and DWT-based statistical features (i.e., mean, average power). However, in SVM, optimal choice of two radial basis kernel function parameters (C and γ) is tricky for accurate classification. Inaccurate setting leads to specious outcomes. Although RF is an efficient classifier, however, it takes large memory and also slows down the process in case of large number of trees, lacks explanation, and tends to overfit while classifying noise data. The algorithms associated with [79] and [54] that deal with binary classification problems, i.e., pair of neuropathy, myopathy, and normal subjects making use of DWT statistics, are derivative forms of evolutionary algorithms and particle swarm optimization (PSO). Two innate limitations associated with algorithms are computational burden and lack of theoretical guarantees. Further, use of DWT statistics reduces the *curse of dimensionality*; however, for wide variability of templates, it may not be feasible for analysis. Also, it requires human intervention, and choice of coefficients is impractical. It is also seen that choice of coefficients varies with problems, which may cause result diversity, e.g., DWT + AR-based method achieves 95.0 recognition rate in neu versus myo versus nor, while it is 71.4% in ALS versus myo versus nor which could be due to bias in FR or classifier [56]. Aside from reported works, this study is based on Mv learning, through which promising results are observed in comparison with reported methods. The aftermaths of probe endorse that the algorithm consisting of mCCA, fusion, transformation, and LDA-kNN leads to precise categorizations of subjects suffering from NDs or healthy subjects. Thus, it can help rendering the clinically relevant facts correlating to the structural and functional aspect of MUs of a muscle.

It is admitted that the advocated algorithm fails to categorize a few case of ALS and myopathy performance, which may likely due to inherent similarity of templates or extracted features that are significantly different from trained pattern. In addition, the order selection before fusion encourages further research pursuits. It is pertinent to mention that incorporation of DWT in the first approach to extract second view via subspace learning is significantly different from that used in the reported methods, which involves only decomposition of templates and extraction of components with the guidance of frequency contents in the templates.

The integrity of the proposed technique is due to well-defined feature extraction and fusion strategy which in turn abates the complexity. Further, unique FR can be easily captured in terms of compact forms, which plays a crucial role in multitask learning models. In sequel, it performs well in both sets of data. Additionally, the algorithm involves a number of steps which have been analyzed accordingly for different runs, average of all is found to be 78.72 s approximately. This could vary with machine.

3.4 Limitations

The efficacy of the proposed method is validated via two binary and one three-level classification task. Nonetheless, a few limitations of our methods are listed below.

First, the proposed method deployed CCA learning technique, while there are still many extensions of CCA techniques (i.e., LS-CCA [29], sparse CCA [29]) available in the literature which are also efficient for multi-class learning tasks. It is interesting to investigate whether they can provide consistent performance or enhance the performance in ND classification.

Second, we decomposed multi-views into set of vectors with visual investigation of MUAPs contained in that sequence with the help of EMGLAB, from where multiple feature representations are extracted. However, it is difficult to achieve symmetrical pattern vectors due to random variation of data patterns, which may cause feature biasing. Therefore, further investigation needs to done to overreach the limitations.

Third, we demonstrated the VM within- and inter-subject signals in terms of correlations among multi-template vectors (i.e., feature matrices) followed by feature estimation. However, it is interesting to investigate the nature of variation in combining multi-domain correlations.

Fourth, we reported the performance of spectrogram-based analysis with only small set of data. It remains to be investigated to find out the optimum settings for further enhancement of aftermaths as an efficient alternative replacement for automatic classification of EMG patterns. Further, our study mainly focuses on the classification of three subject groups; nonetheless, there are number of subclass of disease which are also needed to classify, which has a paramount impact on the supervision of an effective treatment of subjects at the early stage of disease.

4 Conclusion

In this article, we described two multi-view subspace learning feature fusion strategy for automatic classification of EMG templates. Specifically, we formulated multi-view representations using selected multiple templates from where their compact representations (i.e., λ_1, λ_2) are estimated out using adopted learning

followed by variability measures based on which dimensionality is further reduced. Then, feature vectors are concentrated through two fusion strategies, followed by class transformation, and then classify them using LDA-kNN and neural network. The efficacy of the proposed methodology is estimated with two sets of EMG data, one collected from online database and other collected from author institute, GNRC hospital, Guwahati, India, and achieved promising results. Admittedly, classification efficiency relies on multi-view features through which their compact representations are estimated for subsequent learning. Following the results, we envisage for the development of EMG analysis system with graphical user interference (GUI) that would be simple but accurate and reliable enough for clinical environment. Also, our future research pursuits are enhancement of recognition performance of spectrogram-based analysis and multimodality pattern analysis, i.e., investigation of λ_3 by considering multimodality data (e.g., fMRI, sMRI, and EEG).

Acknowledgements The authors would like to thank Mantoo Kaibarta, Meenakshi Boro, and Rajesh Barman for their help during signal analysis.

Appendix

Performance Evaluation

Let M be the 3×3 confusion matrix for three class problems, where the entries m_{ij} give the number of epochs that were predicted to be class i, while the actual class was j. However, it can also be generalized to a binary problem. Denoting TP, TN, FP, and FN as true positive, true negative, false positive, and false negative, respectively, we defined the parameters for each class c as follows [64, 80–83]:

↓→	Normal	ALS	Myopathy
Normal	m_{11}	m_{12}	m_{13}
ALS	m_{21}	m_{22}	m_{23}
Myopathy	m_{31}	m_{32}	m_{33}

→Actual class, ↓Predicted class

$$\mathrm{TP}_c = m_{ll};\ \ \mathrm{FP}_c = \sum_{i \neq c} m_{ci};\ \ \mathrm{TN}_c = \sum_{i,j \neq l} m_{ij};\ \ \mathrm{FN}_c = \sum_{i \neq c} m_{ij}$$

Based on these definitions seven evaluation metrics are defined as follows:

$$\mathrm{Sn}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_l} \tag{18}$$

$$\mathrm{Sp}_c = \frac{\mathrm{TN}_c}{\mathrm{TN}_c + \mathrm{FP}_c} \tag{19}$$

$$\mathrm{PPV}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c} \tag{20}$$

$$\mathrm{NPV}_c = \frac{\mathrm{TN}_c}{\mathrm{TN}_c + \mathrm{FN}_c} \tag{21}$$

$$F_c = \frac{\mathrm{PPV}_c \cdot \mathrm{Sn}_c}{\mathrm{PPV}_c + \mathrm{Sn}_c} \tag{22}$$

$$\mathrm{Ac} = \frac{\mathrm{Trace}(M)}{\sum_{i,j} m_{i,j}} \tag{23}$$

Fusion

Two widely used CCA-based fusions in pattern recognition are parallel and serial fusions [13, 24, 25]. In comparison, parallel fusion technique provides more robust solution due to its low dimensionality. Two projection vectors $\tilde{X}_1$ and $\tilde{X}_2$ are such that

$$\tilde{X}_1 = A^{\mathrm{T}} X_1 \tag{23}$$

$$\tilde{X}_2 = B^{\mathrm{T}} X_2 \tag{24}$$

The dimensions of A and B are determined from the rank of matrices X_1 and X_2. Further, dimension of projection vectors depends on the order selection of weight matrices. Similarly, second pair of weight vectors, i.e., C and D, and subsequent parallel and serial integration of both pairs for effective representation of discriminate pattern vectors are estimated as follows:

$$A_{\mathrm{p}}(p) = \mathrm{diag}[A \quad B \quad C \quad D]^{\mathrm{T}} \begin{bmatrix} X_1 \\ X_2 \\ X_{11} \\ X_{22} \end{bmatrix} \tag{25}$$

$$B_{\mathrm{p}}(s) = [A \quad B \quad C \quad D]^{\mathrm{T}} \begin{bmatrix} X_1 \\ X_2 \\ X_{11} \\ X_{22} \end{bmatrix} \tag{26}$$

where X_{11} and X_{22} are vectors associated with C and D.

References

1. Liu M, Zhang D, Shen D (2015) Inherent structure based multi-view learning with multi-template feature representation for Alzheimer's disease diagnosis. IEEE Trans Biomed Eng (in press)
2. Shen X, Sun Q (2014) A novel semi-supervised canonical correlation analysis and extensions for multi-view dimensionality reduction. J Vis Commun Image Represent 25(8):1894–1904
3. Xia T, Tao D, Mei T, Zhand Y (2010) Multiview spectral embedding. IEEE Trans Syst Man Cybern B Cybern 40(6):1436–1446
4. Ruping S, Scheffer T (2005) Workshop learning with multiple views. In: Proceedings 22nd ICML, Bonn, Germany
5. Zhou D, Burges C (2007) Spectral clustering and transductive learning with multiple views. In: Proceedings 24th international conference on machine learning, Corvallis, OR, pp 1159–1166
6. Bickel S, Scheffer T (2004) Multi-view clustering. In: Proceedings of international conference on data mining, pp 19–26
7. Hazarika A, Barthakur M, Dutta L, Bhuyan M (2016) Two-fold feature extraction technique for Biomedical signals classification. In: International IEEE conference on inventive computation technologies, IEEE press, India
8. Hazarika A, Barthakur M, Dutta L, Bhuyan M (2016) Fusion of projected feature for classification of EMG patterns. In: International conference on IEEE recent advances and innovations in engineering, IEEE press, India
9. Chu JU, Moon I, Lee YJ, Kim SK, Mun MS (2007) A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control. IEEE/ASME Trans Mechatron 12(3):282–290
10. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 22(1):4–37
11. Correa NM, Adali T, Yi-Ou L, Calhoun VD (2010) Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. IEEE J Sel Topics Signal Process 2(6):39–50
12. Xu L, Zhang JQ, Yan Y (2004) A wavelet-based multisensor data fusion algorithm. IEEE Trans Instrum Meas 53(6):1539–1545
13. Sun QS, Zeng SG, Liu Y, Heng PA, Xia DS (2005) A new method of feature fusion and its application in image recognition. Pattern Recog 38(12):2437–2448
14. Sargin ME, Ycel Y, Engin E (2007) Audiovisual synchronization and fusion using canonical correlation analysis. IEEE Trans Multimed 9(7):1396–1403
15. Sun BY, Zhang XM, Li J, Mao XM (2010) Feature fusion using locally linear embedding for classification. IEEE Trans Neural Netw 21(1):163–168
16. Gokgoz E, Subasi A (2014) Effect of multiscale PCA de-noising on EMG signal classification for diagnosis of neuromuscular disorders. J Med Syst 38(4):1–10
17. Dutta L, Hazarika A, Bhuyan M (2016) Comparison of direct interfacing and ADC based system for gas identification using E-Nose. In: International IEEE conference on inventive computation technologies, IEEE press, India
18. Kim TK, Kittler J, Cipolla R (2007) Discriminative learning and recognition of image set classes using canonical correlations. IEEE Trans Pattern Anal Mach Intell 29(6):1005–1018
19. Dutta L, Hazarika A, Bhuyan M (2016) Microcontroller based E-Nose for gas classification without using ADC. Sens Transducers 202(7):38–45
20. Naik G, Selvan S, Nguyen H (2016) Single-channel EMG classification with ensemble-empirical-mode-decomposition-based ICA for diagnosing neuromuscular disorders. IEEE Trans Neural Syst Rehabil Eng 24(7):734–743
21. Hotelling H (1936) Relations between two sets of variates. Biometrika 28(3/4):321–377

22. Fu Y, Cao L, Guo G, Huang TS (2009) Multiple feature fusion by subspace learning. In: Proceedings ACM international conference on content-based image and video retrieval, New York, NY, USA, pp 127–134
23. Peng Y, Daoqiang Z, Zhang J (2010) A new canonical correlation analysis algorithm with local discrimination. Neural Process Lett 31(1):1–15
24. Yuan YH, Sun QS, Zhou Q, Xia DS (2011) A novel multiset integrated canonical correlation analysis framework and its application in feature fusion. Pattern Recog 44(5):1031–1040
25. Shen X, Sun Q (2015) Orthogonal multiset canonical correlation analysis based on fractional-order and its application in multiple feature extraction and recognition. Neural Process Lett 42(2):301–316
26. Khushaba RN (2014) Correlation analysis of electromyogram signals for multiuser myoelectric interfaces. IEEE Trans Neural Syst Rehabil Eng 22(4):745–755
27. DeClercq W, Vergult A, Vanrumste B, Paesschen WV, Huffel SV (2006) Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. IEEE Trans Biomed Eng 53(12):2583–2587
28. Lin Z, Zhang C, Wu W, Gao X (2006) Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Trans Biomed Eng 53(12):2610–2614
29. Sun L, Shuiwang J, Jieping Y (2011) Canonical correlation analysis for multilabel classification: a least-squares formulation, extensions, and analysis. IEEE Trans Pattern Anal Mach Intell 33(1):194–200
30. Arikidis NS, Abel EW, Forster A (2002) Interscale wavelet maximum—a fine to coarse algorithm for wavelet analysis of the EMG interference pattern. IEEE Trans Biomed Eng 49 (4):337–344
31. Waclawik AJ (2004) Neurodegenerative disorders: amyotrophic lateral sclerosis and inclusion body myositis, neurology board review manual. Neurology 8(3):1–15
32. Kamali T, Boostani R, Parsaei H (2014) A multi-classifier approach to MUAP classification for diagnosis of neuromuscular disorders. IEEE Trans Neural Syst Rehabil Eng 22(1): 192–200
33. Paganoni S, Amato A (2013) Electrodiagnostic evaluation of myopathies. Phys Med Rehabil Clin N Am 24:193–207
34. Abel EW, Meng H, Forster A, Holder D (2006) Singularity characteristics of needle EMG IP signals. IEEE Trans Biomed Eng 53(2):219–225
35. Fuglsang-Frederiksen A (2000) The utility of interference pattern analysis. Muscle Nerve 23(1):18–36
36. Fukuda TY, Echeimberg JO, Pompeu JE, Lucareli PRG, Garbelotti S, Gimenes RO, Apolinario A (2010) Root mean square value of the electromyographic signal in the isometric torque of the quadriceps, hamstrings and brachial biceps muscles in female subjects. J Appl Res 10(1):32–39
37. Tohru K (1992) Investigation of parametric analysis of dynamic EMG signals by a muscle-structured stimulation study. IEEE Trans Biomed Eng 39(3):280–288
38. Gler NF, Kocer S (2005) Classification of EMG signals using PCA and FFT. J Med Syst 29(3):15–25
39. Yousefi J, Wright AH (2014) Characterizing EMG data using machine-learning tools. Comput Biol Med 51(1):1–13
40. Subasi A, Yilmaz M, Ozcalik HR (2006) Classification of EMG signals using wavelet neural network. J Neurosci methods 156(1):360–367
41. Podrug E, Subasi A (2015) Surface EMG pattern recognition by using DWT feature extraction and SVM classifier. In: 1st conference of medical and biological engineering in Bosnia and Herzegovina, CMBEBIH
42. Pattichis CS, Schizas CN, Middleton LT (1995) Neural network models in EMG diagnosis. IEEE Trans Biomed Eng 42(5):486–496
43. Pattichis CS, Pattichis MS (1999) Time-scale analysis of motor unit action potentials. IEEE Trans Biomed Eng 46(11):1320–1329

44. Katsis CD, Exarchos T, Papaloukas C, Goletsis Y, Fotiadis DI, Sarmas I (2007) A two-stage method for MUAP classification based on EMG decomposition. Comput Biol Med 31(9): 1232–1240
45. Kaur G, Arora AS, Jain VK (2009) Multi-class support vector machine classifier in EMG diagnosis. WSEAS Trans Signal Process 5(12):379–389
46. Hassan AR, Mohammed IHB (2016) A decision support system for automatic sleep staging from EEG signals using tunable q-factor wavelet transform and spectral features. J Neurosci Meth 271:107–118
47. Hassan AR, Siuly S, Yanchun Z (2016) Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating. Comput Methods Prog Biomed 137:247–259
48. Englehart K, Hudgins B, Philip A (2001) A wavelet-based continuous classification scheme for multifunction myoelectric control. IEEE Trans Biomed Eng 48(3):302–311
49. Dobrowolski AP, Wierzbowski M, Tomczykiewicz K (2012) Multiresolution MUAPs decomposition and SVM-based analysis in the classification of neuromuscular disorders. Comput Methods Prog Bio-med 107(3):393–403
50. Subasi A (2012) Classification of EMG signals using combined features and soft computing techniques. Appl Soft Comput 12(8):2188–2198
51. Alickovic E, Subasi A (2016) Medical decision support system for diagnosis of heart arrhythmia using dwt and random forests classifier. J Med Syst 40(4):1–12
52. Podrug E, Subasi A (2015) Surface EMG pattern recognition by using DWT feature extraction and SVM classifier. In: Conference of medical and biological engineering, Bosnia and Herzegovina, CMBEBIH
53. Bentaleb Y (2010) An algorithm of wavelets for the pretreatment of EMG biomedical signals. Int J Contemp Eng Sci 3(6):285–294
54. Subasi A (2013) Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. Comput Biol Med 43(5):576–586
55. Subasi A (2012) Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines. Comput Biol Med 42(8):806–815
56. Doulah ABMSU, Fattah SA, Zhu WP, Ahmad MO (2014) Wavelet domain feature extraction scheme based on dominant motor unit action potential of EMG signal for neuromuscular disease classification. IEEE Trans Biomed Circuits Syst 8(2):155–164
57. Nikolic M (2001) Detailed analysis of clinical electromyography signals: EMG decomposition, findings and firing pattern analysis in controls and patients with myopathy and amytrophic lateral sclerosis. PhD thesis, University of Copenhagen, Copenhagen, Denmark
58. Arya R, Jaiswal S (2015) Design of low pass FIR filters using Kaiser window function with variable parameter beta. Int J Multidiscip Curr Res 3
59. McGill KC, Lateva ZC, Marateb HR (2005) EMGLAB: an interactive EMG decomposition program. J Neurosci Meth 149(2):121–133
60. Kakarala R, Philip OO (2001) Signal analysis using a multiresolution form of the singular value decomposition. IEEE Trans Image Process 10(5):724–735
61. Johnson RA, Dean WW (1992) Applied multivariate statistical analysis. Prentice hall, NJ
62. Hardoon DR, Sandor S, John ST (2004) Canonical correlation analysis: an overview with application to learning methods. Neural Comput 16(12):2639–2664
63. Zhang C, Wang H, Fu R (2014) Automated detection of driver fatigue based on entropy and complexity measures. IEEE Trans Intell Transp Syst 15(1):168–177
64. Bruser C, Diesel J, Zink MDH, Winter S, Schauerte P, Leonhardt S (2013) Automatic detection of atrial fibrillation in cardiac vibration signals biomedical and health informatics. IEEE J Biomed Health Inform 17(1):162–171
65. Hemanth DJ, Valentina EB (2016) Special issue on decision support systems for medical applications. Intelligent decision technologies. IOP Press, pp 1–2 (Preprint)
66. Hemanth DJ, Valentina EB, Anitha J (2016) Hybrid neuro-fuzzy approaches for abnormality detection in retinal images. In: Soft computing applications. Springer International Publishing, pp 295–305

67. Hemanth DJ, Valentina EB (2015) Performance improved hybrid intelligent system for medical image classification. In: Proceedings ACM on informatics conference, p 8
68. Maheswari SU, Hemanth JD (2015) Image steganography using hybrid edge detector and ridgelet transform. Def Sci J 65(3):214–219
69. Maheswari SU, Hemanth JD (2015) Frequency domain QR code based image steganography using Fresnelet transform. AEU-Int J Electron Commun 69(2):539–544
70. Hemanth DJ, Vijila CKS, Selvakumar AI, Anitha J (2014) Performance improved iteration-free artificial neural networks for abnormal magnetic resonance brain image classification. Neurocomputing 130:98–107
71. Hemanth DJ, Vijila CKS, Selvakumar AI, Anitha J (2013) Distance metric-based time-efficient fuzzy algorithm for abnormal magnetic resonance brain image segmentation. Neural Comput Appl 22(5):1013–1022
72. Dennis J, Tran HD, Li H (2011) Spectrogram image feature for sound event classification in mismatched conditions. IEEE Signal Process Lett 18(2):130–133
73. Typke R, Wiering F, Veltkamp R (2005) A survey of music information retrieval systems. In ISMIR, pp 153160
74. Clavel C, Ehrette T, Richard G (2005) Event detection for an audiobased surveillance system. In: International IEEE ICME, Amsterdam, The Netherlands
75. Gerosa L, Valenzise G, Tagliasacchi M, Antonacci F, Sarti A (2007) Scream and gunshot detection in noisy environments. In: EURASIP, Poznan, Poland
76. Temko A, Nadeu C (2005) Classification of meeting-room acoustic events with support vector machines and confusion-based clustering. In: International proceedings IEEE ICASSP 05, pp 505508
77. Zue V (1985) Notes on spectrogram reading. Department EECS, Massachusetts Institute of Technology, Cambridge
78. Gokgoz E, Subasi A (2015) Comparison of decision tree algorithms for EMG signal classification using DWT. Biomed Signal Process Control 18:138–144
79. Subasi A (2015) A decision support system for diagnosis of neuromuscular disorders using evolutionary support vector machines. Signal Image Video Process 9(2):399–408
80. Barthakur M, Hazarika A, Bhuyan M (2014) Classification of peripheral neuropathy by using ANN based nerve conduction study (NCS) protocol. ACEEE Int J Commun 5(1):31
81. Barthakur M, Hazarika A, Bhuyan M (2014) Rule based fuzzy approach for peripheral motor neuropathy (PMN) diagnosis based on NCS data. In: International IEEE conference on recent advances and innovations in engineering, pp 1–9, India
82. Barthakur M, Hazarika A, Bhuyan M (2013) A novel technique of neuropathy detection and classification by using artificial neural network (ANN). In: Proceedings ACEEE international conference advances in signal processing and communication, pp 706–713, India
83. Barthakur M, Hazarika A, Bhuyan M (2013) A computer-assisted technique for nerve conduction study in early detection of peripheral neuropathy using ANN. Int J Electron Commun Eng Tech 4(5):47–65

Automatic Detection of Brain Strokes in CT Images Using Soft Computing Techniques

B.S. Maya and T. Asha

Abstract Stroke is the cerebrovascular issue influencing blood supply to the mind that predominantly influences individuals over 65 years old. This article proposes an automatic technique to perceive and orchestrate the sorts of strokes starting with 2D cerebrum CT images. The methodology is divided into four steps. In the introductory step, preprocessing may be performed on the image to expel unwanted disturbance by applying median filtering. In second step, different texture-based features are extricated utilizing wavelet packet transform (WPT) for classification. In the following step, Linear Discriminant Analysis (LDA) is utilized to diminish the dimensionality of the features. Finally, the diminished group of feature is connected to the supervised learning techniques for classification of normal and infected region. The goal of the proposed work is to build up a framework that accurately extracts the stroke region from CT images that helps doctors in their diagnosis decisions. The performance of the proposed scheme has fundamentally enhanced the stroke classification precision contrasted with other neural system-based classifier.

Keywords Computed tomography (CT) · Wavelet packet transform (WPT) · Linear discriminate analysis (LDA) · Classification

1 Introduction

The cerebrovascular chance event influencing the blood supply to the cerebrum is the stroke. For every 40 s, stroke happens. According to World Health Organization (WHO), stroke hazard extremely increments after the age of 65 and also in developed countries it will increment from 10 to 23% [1]. The survey says stroke

B.S. Maya (✉) · T. Asha
Department of Computer Science & Engineering, BIT, Bangalore, Karnataka, India
e-mail: bs1985.maya@gmail.com

T. Asha
e-mail: asha.masthi@gmail.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_5

occurs more among female compared to male because of stress [2, 3]. Up to 80% instances of stroke can be prevented. 35% of people affected by stroke are suffering from various diseases such as permanent disability, sight, mobility, and losing their speech. For proper function of brain, there must be continuous supply of blood, glucose, and oxygen to the nerve cells. On the off chance that the supply is weakened incidentally, the parts of cerebrum will quit working. If it persists for longer time, the brain cells die and are damaged permanently.

1.1 Symptoms of Stroke

The blood vessels carrying blood to the brain is blocked due to thrombus, leads to ischemic stroke. Due to rupturing of weakened blood vessels in the brain, hemorrhagic stroke occurs. There may be co-occurrence of both ischemic and hemorrhagic strokes at the same time. The symptoms of stroke include paralysis, difficulty in speech, blurred vision, confusion, and severe headache [4, 5].

1.2 Risk Factors of Stroke

The most essential hazard considered for the stroke to happen is hypertension. Other hazard variables of stroke incorporate diabetes, seniority, tobacco smoking, high cholesterol, transient ischemic assault, and so forth.

1.3 Imaging Modalities

Stroke is diagnosed through various imaging technologies such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) as shown in Fig. 1.

Compared to MRI, CT imaging is preferred due to good recognition of calcification, hemorrhage, and hard detail [6, 7]. In addition to that, CT images offer lower cost, wider availability, and sensitiveness to early stroke. For these reasons, our proposed work aims to explore the methods for classification of the brain strokes into normal, ischemic, and hemorrhagic stroke and consequently diminish the danger of misdiagnosis and botch.

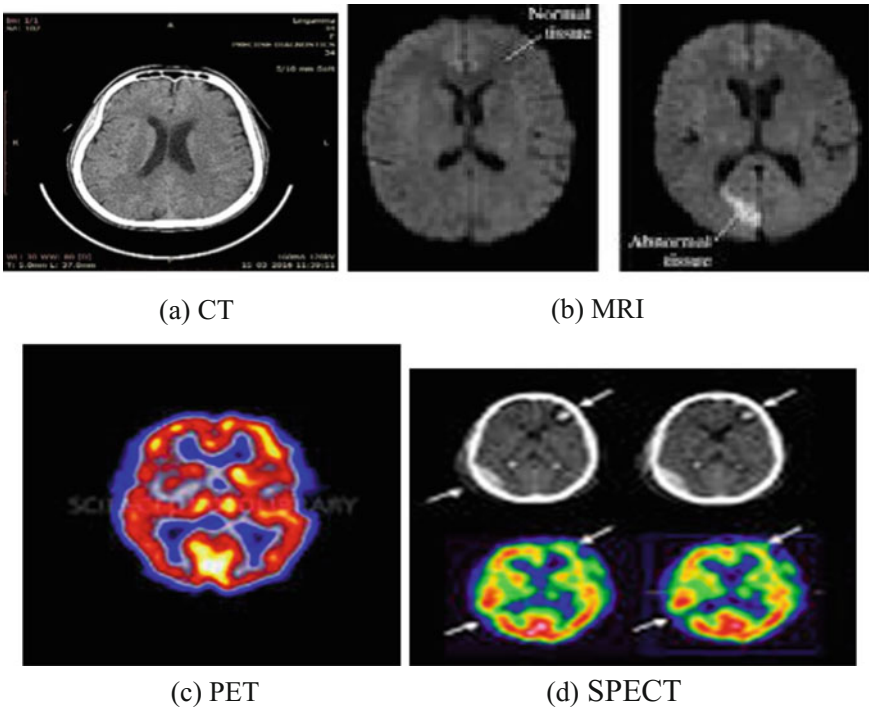


Fig. 1 Different types of imaging modalities

1.4 Proposed Scheme

For correct detection and diagnosis of stroke-affected regions, an accurate segmentation and classification is essential, which is done by well-experienced radiologists. Yet at the same time, it is a testing errand for them in view of time utilization, inter- and intra-subject variability, and expert's domain knowledge. A radiologist uses the automated segmentation methods to make a quicker decision [8].

In this work, the study is about the automatic classification of CT images into stroke and non-stroke images. The proposed method uses median filter for preprocessing the images and extracts features from CT brain images using discrete wavelet transform (DWT). Feature reduction is accomplished by Linear Discriminate Analysis (LDA). To classify the images as normal and abnormal CT brain images, the support vector machine (SVM) is used.

The ease of the paper is sorted as follows: Sect. 2 contains the related work. Section 3 presents the proposed technique for segmentation of CT images and it also includes preprocessing, pre-segmentation, feature extraction, and feature reduction techniques. Section 4 describes the supervised learning classifier for

classification of strokes. Section 5 deals with experimental results and discussions, and finally, Sect. 6 concludes the paper.

2 Related Work

Numerous scientists have been attempting to discover the appropriate calculation for distinguishing stroke of CT cerebrum images in the previous three decades. Regarding stroke hazards, a few computer-aided diagnosis frameworks have been proposed to assist the doctor in the analysis and treatment of stroke patients [9–12]. These frameworks have contributed in the identification of early stroke signs and to the improved diagnostic accuracy for acute stroke. A conspicuous work around there was control parallel symmetry based recommendation to recognize strokes in CT scan images volume [13]. A few unique advances have been examined with the expectation to build up a minimized framework that can be utilized to recognize ischemic from hemorrhagic stroke in a pre-hospital setting. In [14], ultrasonography can be used to identify large vessel occlusions in acute stroke care, but cannotexclude hemorrhagic transformation of the resulting ischemic regions. A more exhaustive audit of different strategies for drain examination can be found in [15, 16]. The strategies will check the symmetry for every cut. In the event that the midline of the cerebrum moves, a discharge is thought to be available. These strategies accept the neurotic cerebrum truant through the symmetric property. In this way, if some anomalous area seems just without a moment's delay side, it is considered as stroke.

The existing work done on stroke discovery focuses mainly on hemorrhagic stroke location. The methodology in [17] exploits the fact that hemorrhagic tissues are brighter than the normal tissues and hence uses a histogram based k-means initial clustering followed by final segmentation using 3D morphological binary dilation of the initial clusters. Learning-based methodologies have been recommended previously in [18, 19]. Zhang and Wang presented a strategy in view of the elements of surface, dim scale, shape and the symmetric parts human mind CT picture is ordered [20]. The morphological operation based technique for feature extraction was utilized to distinguish between normal and abnormal CT images [21]. More recently, wavelet based texture analysis has been utilized to first eliminate all the nasal cavity slices followed by intensity based thresholding [22–24] to recognize the stroke-influenced locales. Picture deterioration is that it empowers the information into segregate and control with particular properties [25]. A research done for data denoising and local contrast enhancement in multi scaled domain improves infarction perception. The wavelet based method upgrades the sublets sign of hypo density of CT images [26].The stroke-influenced locale from the irregular images is sectioned utilizing the force distinction of the image and the sore district [27, 28]. In [29], k-means clustering strategy is employed along with classification technique such as SVM and ANN. The fuzzy and watershed algorithms are used with neural system to characterize the strokes [30, 31].

3 Methodology

The proposed technique contains four stages such as image preprocessing, image segmentation, feature extraction, and classification as illustrated in Fig. 2.

3.1 Information Acquisition

Continuous CT information sets have been gathered from different sources such as medicinal focus managing brain determination, Premier Healthcare, and KIMS Research Center and Hospital, Bangalore, Karnataka. All collected images are stored in a database. The images considered are of the size 512 X 512 reconstruction matrix, 2.5–5-mm slice thickness, X-ray source voltage is 120 kV and maximum X-ray tube current is 63 mA.

3.2 Image Preprocessing

CT images need to be preprocessed because of unorganized nature of the brain tissue. Basically, image preprocessing is used to expel commotion and intensity enhancement. The different steps involved in image preprocessing are conversion of

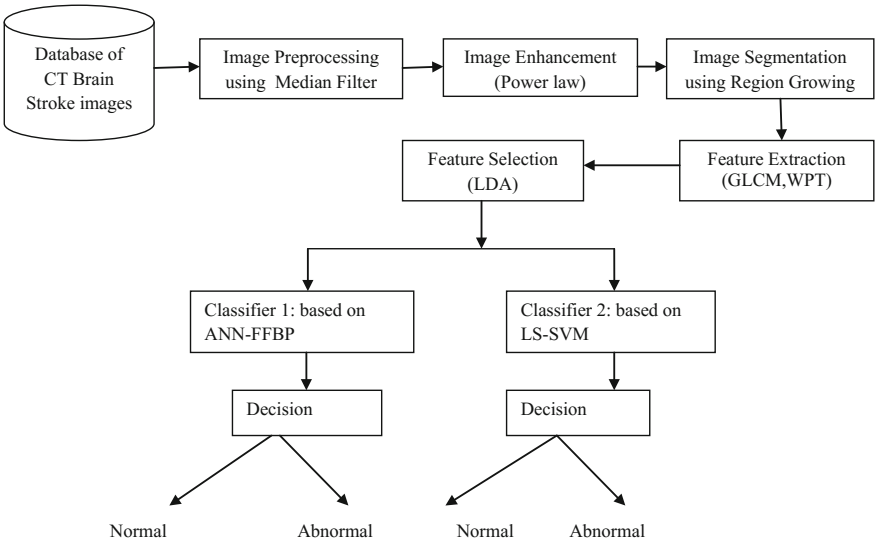


Fig. 2 Block diagram of proposed system

CT images to grayscale image, enhancement of the gray image, region growing and brain insulation, and noise removal.

Conversion of CT images to Grayscale image: The original CT image which is shown in Fig. 3a is cropped and converted into grayscale images by eliminating brightness information, thus converting the image format from 512 × 512 × 3 color RGB to 512 × 512 gray image as shown in Fig. 3b.

Enhancement of the Gray image: The difference of the image is lessened by the low circulation of the gray levels of the first image. Along these lines, differentiate extending is connected to expand the dynamic scope of the gray levels in the images. Image enhancement is only evolving an image f under image g using T, where T is the change function, and r and s are utilized to mean the estimations of pixels in images f and g individually. The pixel values r and s are connected by the expression (1)

$$s = T(r) \tag{1}$$

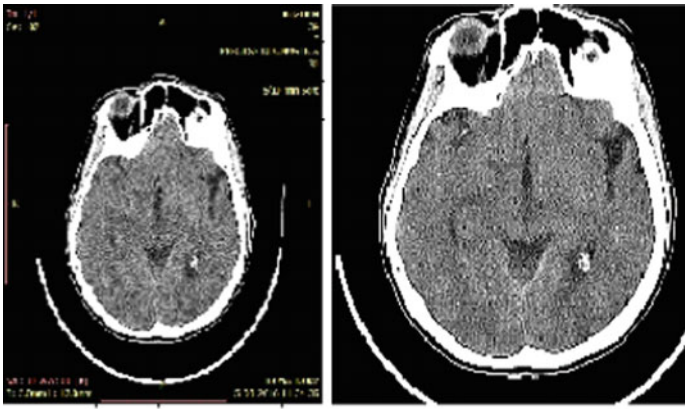
The aftereffects of this change are mapped into the grayscale-computerized images. So the outcomes are mapped once again into the range [0, $L - 1$], where $L = 2^k$, k is the quantity of bits in the image being considered [32].

Linear Contrast Enhancement: It is used to enhance the images. Using a function $g(f)$ produces a new image B from a given image A by means of Eq. (2).

$$B(i,j) = g(A(i,j)) \tag{2}$$

where $i = 0, \ldots, N - 1$, $j = 0, \ldots, M - 1$.

The function $g(f)$ works on each image pixel freely. All pixels with unique gray level f are changed to have gray level $g(f)$ by the Eq. (3).



(a) Original CT image (b) Cropped and grayed image

Fig. 3 Converting RGB to grayscale images

$$g(f) = af + b \tag{3}$$

The original CT image and its histogram which are shown in Fig. 4 are constructed first. The developed histogram contains a few pinnacles. Basically, the background pixels and delicate tissues are contributed by the left peak and the furthest right, respectively. The useful information is obtained by the rightmost peak contributed by the delicate tissues [33]. Linear contrast stretching method is applied over the required region to enhance the visibility of the image as shown in Fig. 5. In linear contrast stretching, the image contrast is insufficient to make its texture clear. To defeat this, we have proposed a technique using power law enhancement technique which improves the image contrast.

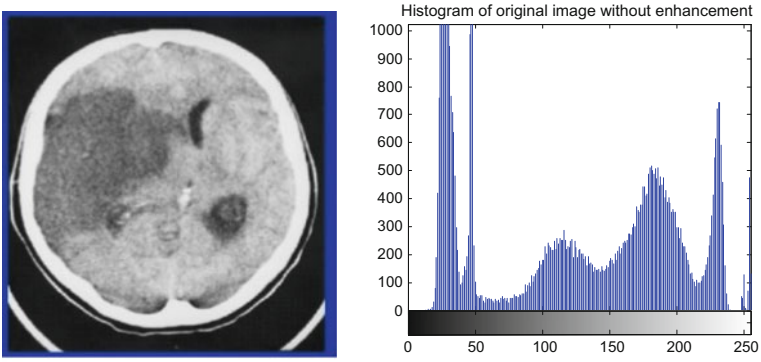
Power Law Enhancement: This change capacity is likewise called as gamma amendment. For different estimations of γ, diverse levels of upgrades can be acquired by Eq. (4).

$$s = cr^{\gamma} \tag{4}$$

Diverse screens show images at various intensities and clarity by setting distinctive gamma revision values. For the given values gamma = 0.6 and constant = 1, power law enhancement gives clear information when compared to linear enhancement as shown in Fig. 6.

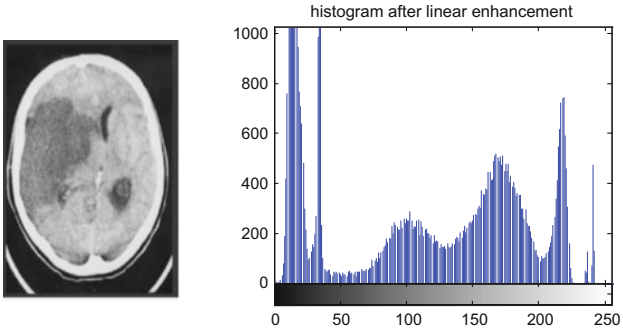
Region Growing and Brain Insulation: Based on predefined criteria, it assembles the pixels or sub-locales into larger areas. To compute the zone of white matter in the skull, the intensity level is used [34] as shown in Fig. 7.

After visualization, the next step is to identify and to expel the skull parts. Since it is the brightest part of the image, the intensity of its pixels is above 230 and it can be effectively evacuated utilizing MATLAB work. The morphological operations such as logical OR and logical AND are applied to remove the brain skull tissue [34]. The pseudocode 1 depicts the removal of skull part from the original image.



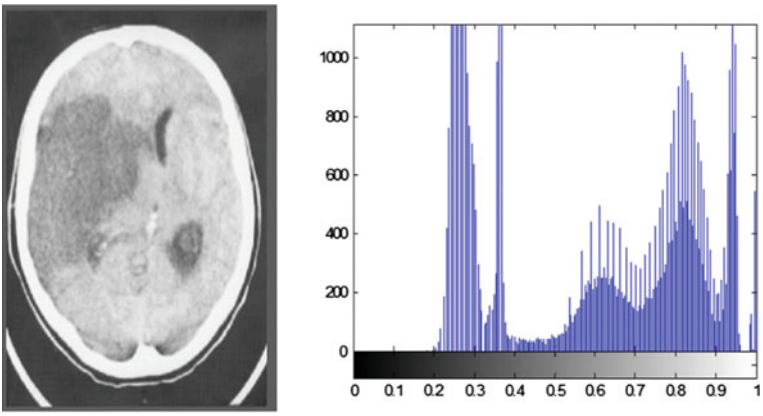
(a) Original CT Image (b) Histogram of the CT Image

Fig. 4 Original image and its histogram



(a) Enhanced CT Image (b) Histogram of the enhanced CT Image

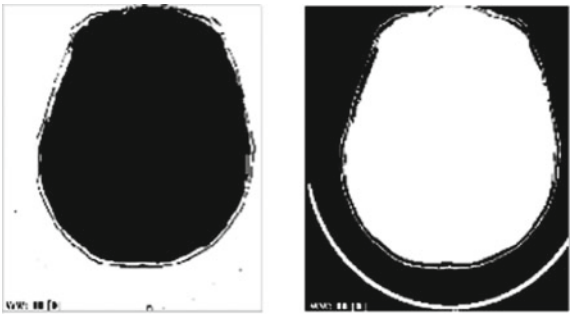
Fig. 5 Linear contrast enhanced image and its histogram



(a) Enhanced CT image (b) Histogram of the enhanced CT image

Fig. 6 Power law enhancement and its histogram levels

Fig. 7 Removed skull part



(a) Isolated Skull (b) Masked Image

pseudocode 1: Removal of skull part from original image

Stage 1: Read input image as *O* (p, q) and measure size of given image.
Stage 2: Assess the skull image as S(k, l) from image O (p, q) utilizing the following step

Assume m = p, n = q
E(i,j) = O(i,j) for S (i, j) > T

Stage 3: E_Re(i,j) = difference [O (i,j) and S(i, j)]

Assume m = p, n = q
where E_Re is the removal of skull part from image.

Noise Removal: The median filtering method reduces the noise and relics in the image in preprocessing step [35]. To improve the quality of normal images and eliminating the anomalies without decreasing sharpness of the image median filter is chosen. A median filter of window [3 * 3] is applied on the image to remove the impulse noise in the CT image as shown in Fig. 8. Apply median filtering four times to remove unwanted noise and to get clear area of the stroke area as shown in Fig. 9.

3.3 Pre-segmentation Method

If the images are having similar characteristics, it can be segmented into various regions by the region-based segmentation methods. In the proposed system, we are using region splitting and merging methods. The basic steps for region growing and merging are shown in pseudocode 2 [36].

pseudocode 2:

Let 'P' be a chance to be the first image and 'T' be the specific predicate, flag0 = accurate, flag1 = inaccurate

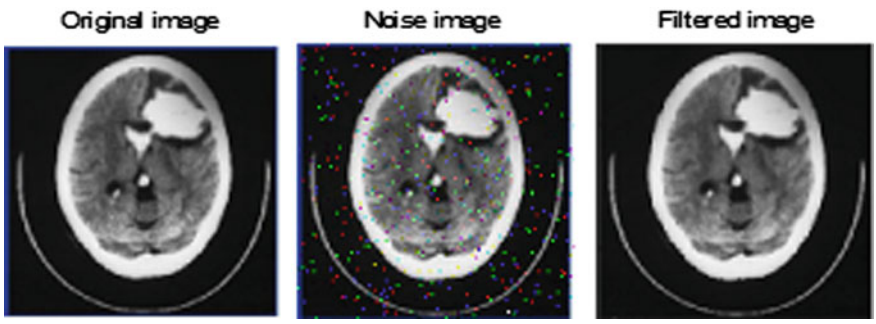


Fig. 8 Resulted images using median filter

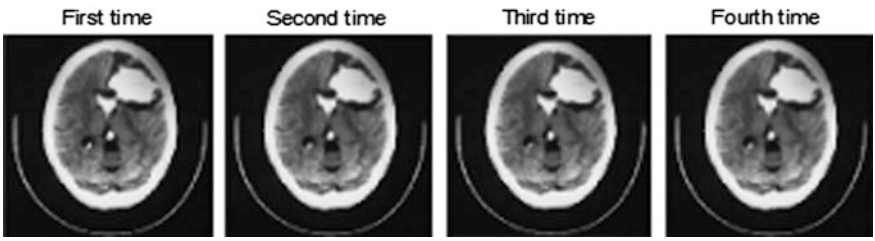


Fig. 9 Resulted image after applying four times median filter

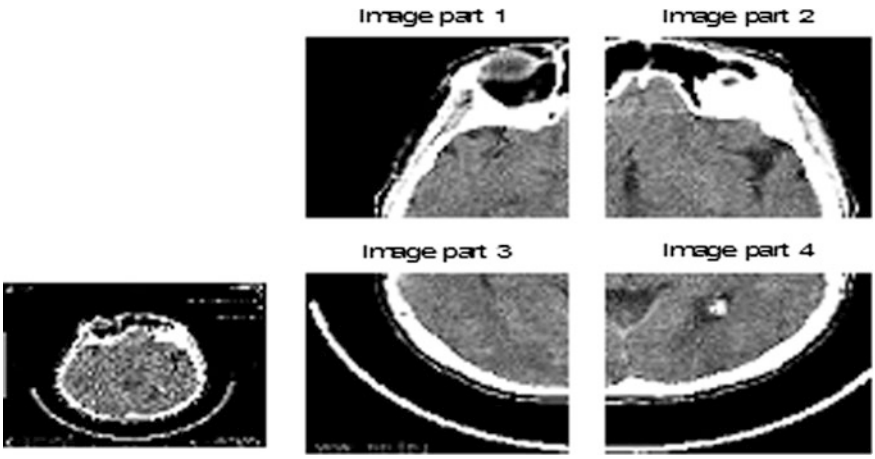


Fig. 10 Quad tree decomposed image

Step 1: Initially region is equivalent to P.
Step 2: Every region is partitioned into quadrants for which T(Region $_{i)}$ = flag1.
Step 3: Every region, T(Region$_j$) = flag0, then union adjacent regions Region$_i$ and Region$_j$ such that T(Region$_i$ U Region$_j$) = flag0
Step 4: Rehash Step 3 until merging is unimaginable.

The entire image is broken into four parts as shown in Fig. 10 and its histogram [28] is shown in Fig. 11. One part among the four parts from the resulted image that has the likelihood of abnormal areas is partitioned into four more sub-parts utilizing statistical feature before grouping is done as shown in Fig. 12 and its histogram is shown in Fig. 13. The proposed pre-segmentation avoids the search in unnecessary regions as shown in pseudocode 3. So we can conclude that stroke area has a mean value which is higher than the other regions [37]. The statistical features that are frequently used are mean and standard deviation, and they are used to distinguish between the normal and abnormal regions. Mean is the sum of all pixels divided by the total number of pixels in an image. Standard deviation is the variety from its mean. They can be characterized by Eqs. (5–7).

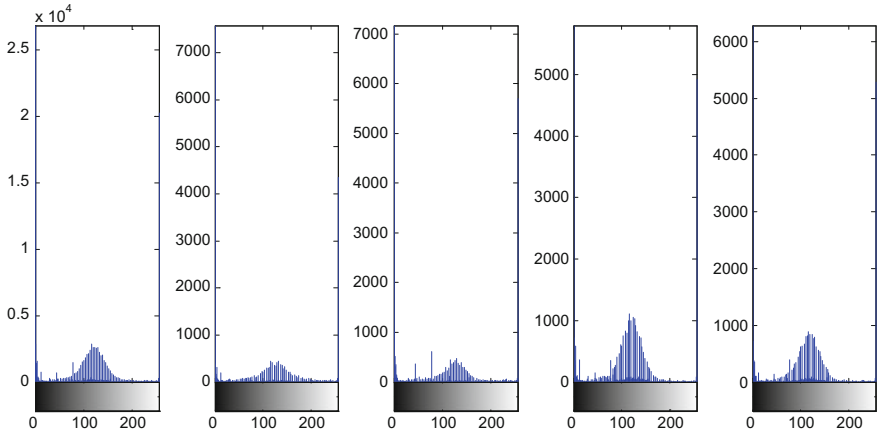


Fig. 11 Histogram of original and decomposed

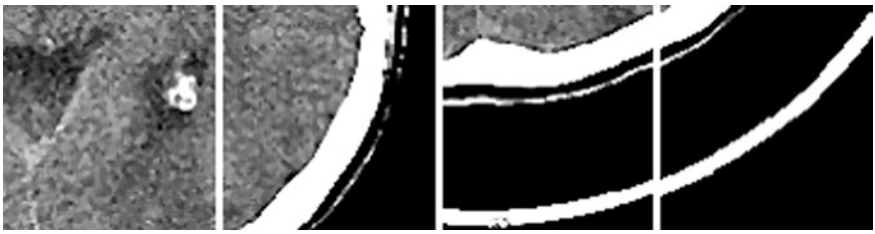


Fig. 12 Decomposed image for the region 4 at second level

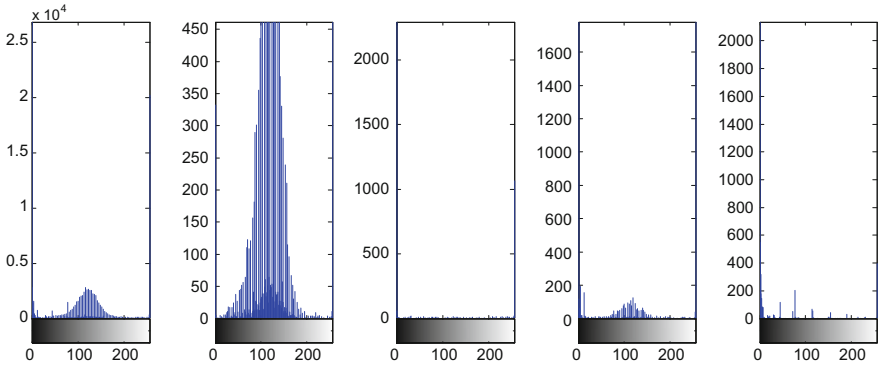


Fig. 13 Histogram for second-level segmented image for part 4

$$P(i) = \frac{H(\mathrm{i})}{NM} \tag{5}$$

$$\text{Mean: } \mu = \sum_{i=1}^{G-1} iP(i) \tag{6}$$

$$\text{Standard Deviation: } \sigma^2 = \sqrt{\sum_{i=1}^{G-1} (i-\mu)^2 P(i)} \tag{7}$$

where 'i' is the irregular variable representing the gray levels of the image extending from $i = 0, 1, 2 \ldots G - 1$. G is gray level of an image (255), N is the quantity of cells in the horizontal space, M is the quantity of cells in vertical space, and $P(i)$ is histogram of i.

pseudocode 3:

The proposed steps for pre-segmentation are as follows:

Step 1: Obtain the CT image and store it in 2D matrix
Step 2: The image is divided into four parts
Step 3: Calculate the histogram and pixel value for each parts
Step 4: Compare histograms and pixel value with each part to get the histogram with the highest pixel value.
Step 5: For the detection of abnormality region, consider the highest histogram pixel value, mean, and standard deviation of the decomposed image.
Step 6: Now, actual abnormality area can be determined by repeating the above steps (2–5).

The features are extracted from each associated image as recorded in Table 1, and it represents mean and standard deviation for the four parts. Figure 14 shows graphical representation of statistical features of mean and standard deviation. In our work, the histogram with pixel value 450 is considered as abnormal regions. Figure 15 and Table 2 represent the statistical features of mean and standard deviation for second-level decomposed image. Only the identified abnormal image part will be given for the further analysis.

This pre-segmentation process makes the segmentation process quicker than conventional process. The final segmentation results tested upon the region resulted from pre-segmentation and precision rates of the proposed framework will be more. The framework helps the radiologists to recognize the hemorrhage and ischemic

Table 1 Mean and standard deviation for decomposed image

Image 1	Part 1	Part 2	Part 3	Part 4
Mean	39.2090	43.7209	53.112	*57.6100*
SD	81.1022	82.0176	80.6726	*77.2007*

Italics indicates that mean of the abnormal region is higher than the normal ones and standard deviation of the abnormal region is lower than the normal ones

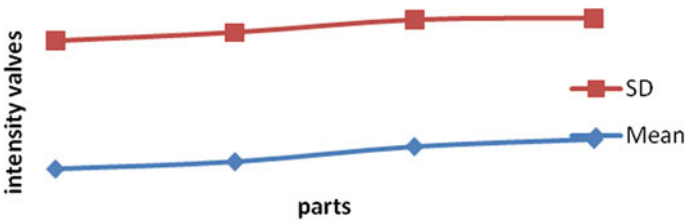


Fig. 14 Statistical features of mean and standard deviation

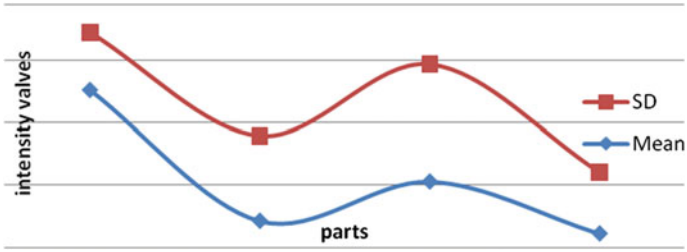


Fig. 15 Statistical features of mean and standard deviation for second-level decomposed image

Table 2 Mean and standard deviation for second-level decomposed image

Image 1	Part 1	Part 2	Part 3	Part 4
Mean	*126.3472*	21.5217	53.1893	11.3315
SD	*45.5955*	68.1509	93.4810	49.3669

Italics indicates that mean of the abnormal region is higher than the normal ones and standard deviation of the abnormal region is lower than the normal ones

stroke in human brain and to arrive at an exact choice quicker. Radiologists utilize the outcomes from the PC as a moment supposition to settle on ultimate conclusions.

3.4 Feature Extraction

The reason for feature extraction is to reduce the original data set by measuring certain properties, or features, that distinguish one input pattern from another. There are a few surface-based element extraction techniques; however, gray-level co-event matrix (GLCM) is extremely normal and effective. Despite high capability of GLCM for feature extraction, it is so time consuming [38]. Keeping in mind the end goal to beat this downside, in the proposed strategy, wavelet change (i.e., DWT and WPT, on the other hand) is used to break down information image into 4 sub-images and after that GLCM technique is connected on every sub-picture.

3.4.1 Discrete Wavelet Transform

The discrete wavelet transform is indistinguishable to a various leveled sub-band framework, where the subgroups are logarithmically divided into recurrence and speak to an octave-band decay [39]. By applying DWT, the image is really separated into four subgroups as shown in Fig. 16a.

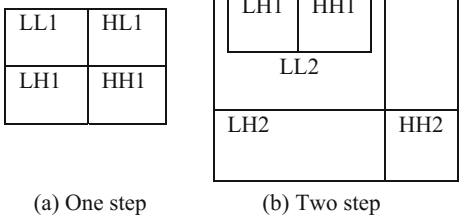
The subgroups speak to the finest scale wavelet coefficients. The sub-band LL1 corresponds to coarse level co-efficient. The sub-band LL1 is further deteriorated to deliver the following disintegration level. The outcomes from second level wavelet decomposition as shown in Fig. 16b. This process continues until some final scale is reached. In like manner, the proposed strategy uses Daubechies wavelet of request 4 as mother wavelet. Consistently, in order to achieve the desired decomposition level, an examination is done and classification accuracies of different decomposition levels are compared together.

3.4.2 Wavelet Packet Transform (WPT)

In wavelet packet transform (WPT), the subtle elements and additionally the approximations would be part and consequently it creates a full paired tree as shown in Fig. 17. Wavelet Packet Decomposition has 3 steps. The low-pass guess coefficients and the high-pass detail coefficients are utilized for choice. In the proposed strategy, db4 is utilized for WPT. Same examination as DWT is additionally done to decide ideal decay level and level 3 is picked. Therefore, eight subgroups are gotten in the proposed technique utilizing WPT per every CT image as shown in Fig. 18.

3.4.3 Gray-Level Co-occurrence Matrix (GLCM)

The GLCM component $C(i, j, d, \Theta)$ speaks to likelihood of the match of pixels, which are situated with a bury test remove d and a bearing Θ, having a gray level i and gray level j. Parameter d values as 1 to most extreme size of picture, and Θ can



(a) One step (b) Two step

Fig. 16 Decomposition of image

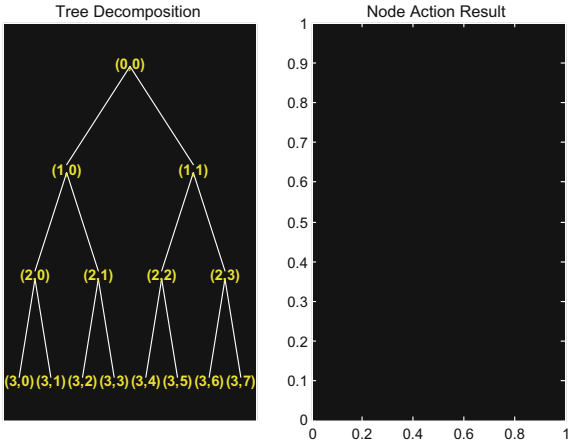


Fig. 17 Wavelet packet decomposition tree

```
==================================
Size of initial data   : [391  361]
Order                  : 2
Depth                  : 3
Terminal nodes         : [7   8   9  10  11  12  13  14]
----------------------------------------------------
Wavelet Name           : db3
Low Decomposition filter  : [0.03523  -0.08544   -0.135    0.4599    0.8069    0.3327]
High Decomposition filter : [-0.3327   0.8069  -0.4599   -0.135  0.08544  0.03523]
Low Reconstruction filter : [ 0.3327   0.8069   0.4599   -0.135 -0.08544  0.03523]
High Reconstruction filter: [0.03523  0.08544   -0.135  -0.4599   0.8069  -0.3327]
----------------------------------------------------
Entropy Name           : shannon
Entropy Parameter      : 0
----------------------------------------------------
```

Fig. 18 Wavelet packet object structure

be esteemed as four normal points: 0°, 45°, 90°, and 135° [40, 41]. In the proposed strategy, *d* is set to 1 and one grid is acquired per every sub-image. After calculating of GLCM matrix, Haralick method [42], is used to extract texture features from Eqs. (8–12) shown in Table 3 and graph shown in Fig. 19.

Table 3 Features and its weights

Features name	Weights abnormal	Weights normal
Contrast	2.6544	1.6904
Correlation	0.8158	0.8746
Energy	0.1755	0.1712
Homogeneity	0.7563	0.8122
Solidity	0.0851	0.01725
Entropy	0.7044	0.7564
Variance	2.55E + 04	1.25E + 03
Skewness	10.4981	8.321
Correlation coefficients	1	1
Mean	126.3472	110.2
Standard deviation	45.5955	75.32

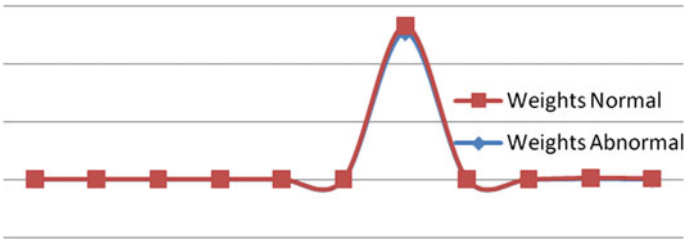


Fig. 19 Graph for feature vector

$$\text{Entropy: } H = -\sum_{i=1}^{G-1} p(i)\log 2[P(i)] \tag{8}$$

$$\text{Energy: } E = \sum_{i=1}^{G-1} (p(i))2 \tag{9}$$

$$\text{Variance} = \sigma^2 = \sum_{i=1}^{G-1} (i-\mu)^2 P(i) \tag{10}$$

$$\text{Skewness} = \sigma^{-3} = \sum_{i=1}^{G-1} (i-\mu)^3 P(i) \tag{11}$$

$$\text{Contrast} = \sum_{i=1}^{G} \sum_{j=1}^{G} P(i,j) \tag{12}$$

3.5 *Feature Selection Using LDA*

Feature selection is a dimensionality reduction technique widely used for data mining and knowledge discovery and it allow elimination of irrelevant or redundant features. It enhances the nature of the informational collection and the execution of learning frameworks [42–44]. There are diverse component determination systems, yet in this paper, LDA is utilized. LDA includes the determination of a subset of '*d*' components from a sum of '*D*' elements. Restricting the component vectors by such a joining procedure prompts an expansion in exactness rates and a lessening in unpredictability and computational time. The *D* elements are meant extraordinarily by particular numbers from 1 to *D*, so that the aggregate arrangement of *D* elements can be composed as $S = \{1, 2, \ldots, D\}$ as shown in Fig. 20. *X* signifies the subset of chosen components, and *Y* means the arrangement of residual elements. Hence, *S* can be represented as following Eq. (13) at any instant of time. $K(X)$ denotes a function evaluating the performance of subset selected from features *X*.

$$S = X \cup Y \tag{13}$$

LDA attempts to express less one dependent variable as a linear combination of other features or measurements. LDA is also closely related to PCA and factor analysis in that, they both look for linear combination of variables which gives best explanations about the data. PCA then again does not assess any distinction in class, and element investigation constructs the component. Combination depends on contrasts instead of likenesses. LDA scans for those vectors in the fundamental space that are best discriminable among classes. More formally given a number of independent features relative to which the data is described, LDA creates a linear combination of those which yields the largest mean differences between the desired classes. We characterize two measures:

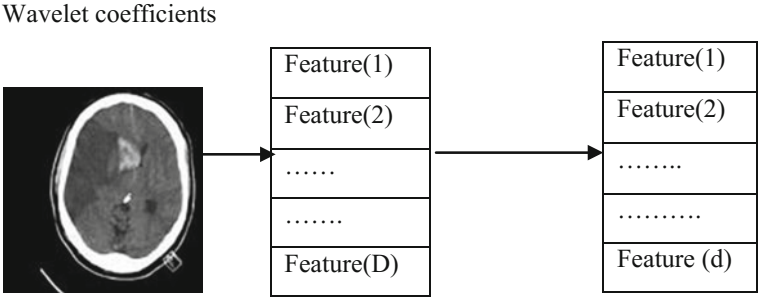


Fig. 20 Schematic diagram for feature reduction scheme using LDA

1. One is called within-class scatter matrix as given by in Eq. (14)

$$\mathrm{Sw} = \sum_{j=1}^{c} \sum_{i=1}^{Nj} \left(x_i^j - \mu_j\right)\left(x_i^j - \mu_j\right)^{\mathrm{T}} \tag{14}$$

where x_i^j is the ith sample of class j, μ_j is the mean of class j, c is the number of classes, and μ_j is the number of samples in class j and

2. Between-class scatter matrix as given by in Eq. (15)

$$\mathrm{Sb} = \sum_{j=1}^{c} (\mu j - \mu)(\mu j - \mu)^{\mathrm{T}} \tag{15}$$

where μ represents the mean of all classes.

4 Soft Computing

Soft computing is not exactly characterized. It comprises of particular ideas and procedures which mean to beat the troubles experienced in genuine issues. These issues result from the way that our reality is by all accounts loose, dubious, and hard to arrange. The essential strategies incorporated into subjective processing are fuzzy logic (FL), neural network (NN), and genetic algorithms (GA).

4.1 Neural Network

The investigation of neural systems began by the publication of McCulloch and Pitts [45]. Neural system makes an endeavor to mimic human cerebrum. The operation of cerebrum is accepted to be found on basic essential components called neurons which are associated with each other with transmission lines called axons and responsive lines called dendrites.

4.2 Artificial Neural Network Classifier

Artificial neural network (ANN) is a mathematical model consisting of a number of highly interconnected processing elements organized into layers, geometry and functionality. The number of neurons used in the proposed work hidden layer is six. The single neuron in the yield layer is utilized to address to ordinary and irregular human brain stroke images. The neural system has been prepared to modify the association weights and inclinations with a specific end goal to create the craved mapping.

In this proposed method neural network concepts are used for characterization issues is feed forward neural network with associated error back propagation learning algorithm. The nodes are connected to such that each node is connected to all nodes of the previous and the successive layer. The input layer is just associated forward to the first hidden layer and the output layer just in reverse to the last concealed layer. The error is then back spread through the networks, and weights are balanced as the system endeavors to perish the blunder by improving the weights.

Back Propagation (BP) algorithms have been regularly utilized for the training of ANN applications comprises of two steps: In forward pass in which the system is initiated on one case and the blunder of yield layer is computed and in backward pass, the system mistake is utilized for over hauling the weights. These weights are iteratively balanced so that the general blunder measure is limited. This can be implemented in Eq. (16), where 'η' is the learning rate, 'n' is the number of iterations, '$\dot{\alpha}$' is the momentum, and δpj is the error signal.

$$\Delta Wji\,(n+1) = \eta\left(\delta_{pj} O_{pi}\right) + \alpha\,\Delta Wji\,(n) \tag{16}$$

The system execution and merging rely on numerous parameters such as introductory weights, inclining rate, momentum utilized, and number of hubs in the shrouded layer amid the training. All input associations are allocated a weight. The training of the system happens through the many cycles of information examples introduced to the network. These cycles are called ages or emphasess. The informational collection is partitioned into an inclining and approval set. Learning informational collection is utilized to prepare the system in regulated mode, and approval informational collection is utilized to test the system execution.

4.3 Support Vector Machine (SVM) Classifier Using LS-SVM

SVM is an effective regulated classifier and exact learning strategy. SVM depends on the basic hazard minimization rule from the statistical learning hypothesis [46]. To overcome the computational overhead, a least squares versions of SVM (LS-SVM) is exploited constraints in he LS-SVM formulation, a set of linear equations has to be solved, instead of a quadratic programming problem for standard SVM [47]. Input to the LS-SVM algorithm is features extracted from wavelet packet coefficients method. The basic step for classification using LS-SVM is shown in pseudocode 4.

pseudocode 4:

Step 1: Create a training dataset of N data points $\{x_k, d_k\}$ k = 1 with input data $x_k \in X$ and class label $d_k \in D = \{-1, +1\}$
Step 2: Apply n-fold (n = 7) cross-validation procedure to find training and testing samples.
Step 3: Choose a kernel function K and train the LS-SVM classifier.
Step 4: Classify test images as normal or abnormal.

Use 7-fold rather than seven-fold.

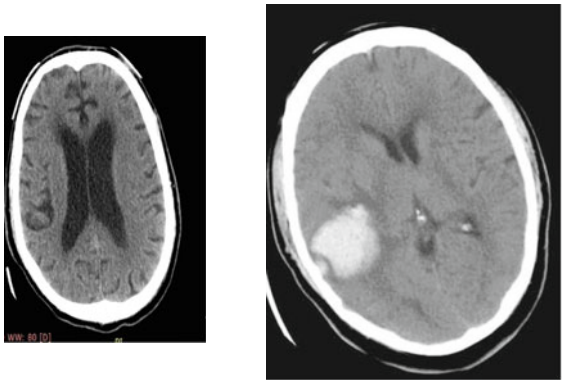
The setting of training and the validation images of the datasets for both classification (ANN-FFBP and LS_SVM) is shown in Table 4.

5 Experimental Results and Discussion

The procured CT real brain images are in JPG format. The CT brain image of 45 patients is considered in the information data set, out of which 90 are abnormal brain images and 25 are normal brain images. This data set is collected from Premier Health Care and KIMS Research center and Hospital, Bangalore. Each patient image consists of 40–60 slices. Slices contain different circumstances, for example, removals, movement obscure, and turn. Not all slices are considered for our study. The input data utilized for anomaly recognition comprises of non-improved 512 * 512 CT brain images. The normal and abnormal images of ischemic and

Table 4 Different classifier datasets

Classifiers	Total number of images (115)		Training images (70)		Validation images (45)	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Feed forward with BP	25	90	15	55	10	35
LS-SVM	30	85	12	58	7	38



(a) Normal Brain Image (b) Abnormal Image

Fig. 21 Original CT brain images. **a** Normal brain image, **b** Abnormal image

hemorrhagic stroke were taken for classification. The significant feature of normal human brain is the symmetry. A normal CT brain image and an abnormal CT brain image are shown in Fig. 21a, b individually.

In the final stage, supervised learning technique SVM is used for CT normal/abnormal human brain classification.

The results of the proposed method are evaluated using the following measures: sensitivity, specificity, and accuracy. The outcomes are quantitatively assessed by a human master. The performance of the proposed classifier technique is analyzed over CT data set. We employed three quantitative measures to evaluate the performance of proposed method in terms of accuracy, sensitivity, and specificity. Every pixel belongs to one of the following four classes. True Positive (TP) is accurately delegated positive pixels, False Positive (FP) is mistakenly named negative pixels, True Negative (TN) is effectively named negative pixels and False Negative (FN) is erroneously named positive pixels. They are characterized in the accompanying Eqs. (17)–(19).

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{17}$$

The likelihood an analytic test is adverse, given that individual does not have the illness.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

The likelihood an indicative test is certain, given that individual has the infection.

Table 5 Mean and standard deviation for second-level decomposed image

Classifier	Sensitivity	Specificity	Accuracy
ANN	98.2	91.5	93
SVM	98.5	98	97.6

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{19}$$

The likelihood an analytic test is effectively performed.

Table 5 demonstrates the characterization rates for the proposed approach. In this analysis, two classifiers based on supervised learning classifiers are presented for CT normal and abnormal human brain classifications. The classification accuracy of ANN is 93%, and SVM is 97.6% obtained in testing phase.

6 Conclusions and Future Work

This review proposes a technique for the arrangement of CT brain images as stroke and non-stroke. The proposed strategy separates features from CT brain images using pre-segmentation and wavelet packet coefficients. LDA is efficiently used for feature reduction. The supervised learning methods give extremely encouraging results in classifying the healthy and brain patient having a stroke. This review utilized a small sample set to prepare the neural system. Reduced image quality and image distortion may result in misinterpretation, henceforth to influence the neural system effectiveness, it is required to use larger data sets.

Future work aims to apply different feature extractions and feature selection procedures that can be utilized to enhance the execution of the framework. Finally, ensemble classifiers for classification will be considered as a future work to fulfill higher exactness. This will permit the future framework to achieve a level that will permit it to be a huge resource for any therapeutic foundation managing brain hemorrhages to identify and characterize diverse sorts of strokes, for example, subdural, epidural, and chronic infarcts.

References

1. The Atlas of Heart Disease and Stroke. World Health Organization, 2004
2. Ministry of Public Health (2002) Burden of disease and injuries in Thailand Priority setting for policy, pp A14–A16
3. Jalivand M, Li X, Zwick T, Wiesbeck W, Pancera, E (2011) Hemorrhagic stroke detection via UWB medical imaging. In: Antennas and propagation (EUCAP), proceeding of the 5th European conference on, pp 2911–2914, April (2011)
4. Neethu S, Venkataraman D (2015) Stroke detection in brain using CT images. J Artif Intell Evol Algorithms Eng Syst 324:379–386 (Springer, 2015)

5. Roy S, Chatterjee K, Bandyopadhyay SK (2014) Segmentation of acute brain stroke from MRI of brain image using power law transformation with accuracy estimation. J Adv Comput, Netw Inf 27, 453–461 (Springer International Publishing)
6. Wysoki MG et al (1998) Head trauma: CT scan interpretation by radiology residents versus Staff Radiologists. Radiology 208(1):8–125
7. Erickson BJ, Bartholmai B (2002) Computer aided detection and diagnosis at the start of the third millennium. J Digit Imag 15(2):59–68
8. Sharma N, Aggarwal LM (2010) Automated medical image segmentation techniques. J Med Phys, Assoc Med Physicists India 3, 35(1)
9. Vymazal J, Rulseh AM, Keller J, Janouskova L (2012) Comparison of CT and MR imaging in Ischemic Stroke. In: Insight into imaging 3(6), 619–627
10. Praveen R Mirajakar, Arun Vikas Singh, Dr. Kishan Asok Bhagwat, Ashalatha M E.: Acute ischemic stroke detection using wavelet based fusion of CT and MRI images. In: International conference on advances in computing, communication and informatics, IEEE (2015)
11. Doi K (2007) Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 31(4–5):198–211
12. Yoshida H, Nappi J (2001) Three dimensional computer-aided diagnosis scheme for detection of colonic polyps. IEEE Trans Med Imaging 20(12):1261–1274
13. Yoshida H, Nappi J, MacEneaney P, Rubin DT, Dachman AH (2000) Computer aided diagnosis scheme for detection of polyps at CT colonography. Radiographics 22(4):963–979
14. Schlachetzki F, Herzberg M, Hlscher T, Ertl M, Zimmermann M, Ittner KP et al (2012) Transcranial ultrasound from diagnosis to early stroke treatment: part 2: prehospital neurosonography in patients with acute stroke: the Regensburg stoke mobile project. Cerebrovasc Dis 33:262–271
15. Perez N, Valdes J,Guevara M, Silva A (2009) Spontaneous intracerebral hemorrhage image analysis methods: a survey. In: Advances in computational vision and medical image processing (2009)
16. Liu Y, Rothfus WE, Kanade T (1997) Content based 3D neuoradiologic image retrieval: preliminary results. In: IEEE content-based video and image retrieval workshop associated with CVPR97
17. Dhawan AP, Loncaric S, Hitt K, Broderick J, Brott T (1993) Image analysis and 3D visualization of intracerebral brain hemorrhage. In: IEEE Symposium on Computer Based Medical Systems, pp 140–145
18. Cosic D, Loncaric S (1997) Computer system for quantitative analysis of ischemic from ct head images. In: 19th Annual international conference of the IEEE (1997)
19. Chan T (2007) Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain. Compute Med Imaging Graph 31(4), 285–298
20. Zhang WL, Wang XZ (2007) Feature extraction and classification for human brain CT images. Proc IEEE Int Conf Mach Learn Cybern 2:19–22
21. Fallahi AR, Pooyan M, Khotanlou H A new approach for classification of human brain CT images based on morphological operations. J Biomed Sci Eng 3, 78–82
22. Chawla M, Sharma S, Sivaswamy J, Kishore L (2009) A method for automatic detection and classification of stroke from brain CT images. In: Proceedings of the Annual International conference IEEE Engineering in Medicine and Biology Society (EMBC 09)
23. Liu R, Tan CL, Leong TY, Lee CK, Pang BC, Lim CT, Tian Q, Tang S, Zhang Z (2008) Hemorrhage slices detection in brain ct images. In: International conference on pattern recognition, pp 1–4
24. Hara T, Matoba N, Zhou X, Yokoi S, Aizawa H, Fujita H, Sakashita K, Matsuoka T (2007) Automated detection of extradural and subdural hematoma for content-enhanced CT images in emergency medical care. In: Proceeding of SPIE (2007)
25. Das DS, Rani GU, Moorthy GS (2012) Analysis of PSNR for Different 3D DWT. J Int J Inf Technol Secur 1, ISSN 2279–008X

26. Przelaskowski A et al (2007) Improved early stroke detection: wavelet based perception enhancement of computerized tomography exams. J Comput Biol Med 37, pp 524–533, Science Direct
27. Seemann T (2012) Digital image processing using local segmentation book. Monash University, Australia
28. Badioze Zaman H et al (2009) Automated segmentation and retrieval system for CT head images. In: IVIC 2009. LNCS, 5857, pp 97–109. Springer, Berlin
29. Rajini NH, Bhavani R (2013) Computer aided detection of ischemic stroke using segmentation and texture features. J Meas 46, 1865–1874 (Science Direct)
30. Nagalkar V, Agrawal S (2012) Ischemic stroke detection using digital image processing by fuzzy methods. J Int J Latest Res Sci Technol 1(4), 345–347
31. Balasooriya U, Perera MUS (2012) Intelligent brain hemorrhage diagnosis using artificial neural networks. In: IEEE business, engineering & industrial applications colloquium (BEIAC)
32. http://funnotes.net/tofpages/TopicOfFortnight.php?tofTpcFl=topicoffortnight2
33. Lee TH (2009) Segmentation of CT brain images using unsupervised clustering. J Vis 12(2), 131–138
34. Ramos OE, Rezaei B Scene segmentation and interpretation image segmentation using region growing. M.Sc. thesis, Computer
35. Gonzalez CR, Woods ER (2000) Digital image processing, 2nd edn. Prentice Hall, New Jersey
36. Kaganami HG, Beij Z (2009) Region based detection versus edge detection. In: IEEE transaction on intelligent information hiding and multimedia signal processing, pp 1217–1221
37. Kyaw MM (2013) Computer aided detection system for hemorrhage contained region. J Int J Comput Sci Inf Technol 1(1)
38. Pauline J, Hitesh Sh (2012) Brain tumor classification using wavelet and texture based neural network. J Int J Sci Eng Res 3(10). ISSN 2229–5518
39. Coifman R, Meyer Y, Quake S, Wickerhauser MV (1993) Signal processing and compression with wavelet packets. In: Progress in wavelet analysis and applications (Toulouse, 1992), pp 77–93, Frontiers, Gif
40. Licciardi G, Pacifici F, Tuia D, Prasad S, West T, Giacco F, Thiel C, Inglada J, Christophe E, Chanussot J, Gamba P (2009) Decision fusion for the classification of hyper spectral data: outcome of the 2008 GRS-S data fusion contest. In: IEEE transaction on geo science and remote sensing, vol 47, no 11, pp 3857–3865
41. Arivazhagan S, Ganesan L (2003) Texture segmentation using wavelet transform, Elsevier. Pattern Recog Lett 24:3197–3203
42. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst, Man Cybern 3(6):610–621
43. Hema Rajini N, Bhavani R (2013) Automatic classification of computed tomography brain images using ANN, k-NN and SVM. J. AI & Society 29, 97–102 (Springer)
44. Sundararaj GK, Balamurugan V (2014) Robust classification of primary brain tumor in computer tomography images using K-NN and linear SVM. In: International conference on contemporary and informatics
45. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5
46. Cristianini N, Shawe T, Taylor J (2000) An introduction to support vector machines and other Kernel-based learning methods, 1st edn. Cambidge University Press, New York
47. Suykens JAK, Vandewalle J (1999) Least squares support vector machines classifiers . Neural Process Lett 9(3), 293–300
48. Olesen J, Gustavsson A, Svensson M, Wittchen HU, Jonsson B (2012) CDBE2010 study group: European brain council the economic cost of brain disorders in Europe. J Eur J Neurol 19:155–162

49. Padma Nanthagopal A, Sukanesh Rajamony R (2012) Automatic classification of brain computed tomography images using wavelet based statistical texture features. J Vis Soc Jpn 1–10
50. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J (2013) Artificial neural network in medical diagnosis. J Appl Biomed 111, 47–58
51. irtskhulava L, Wong J, Al-Majeed S, Pearce G (2015) Artificial neural network model in stroke diagnosis. In: UKSIM-AMSS International conference on modeling and simulation
52. Li GZ, Yang J, Liu GP, Xue L (2004) Feature selection for multi-class problems using support vector machines. Image Process Pattern Recogn 109–111

Survey on the Classification of Intelligence-Based Biometric Techniques

K. Martin Sagayam, J. Felix Jacob Edwin, J. Sujith Christopher, Gowru Vamsidhar Reddy, Robert Bestak and Lim Chot Hun

Abstract Over the past few decades due to the advancement of technology, biometrics has evolved into a key factor of security for societal needs. Biometrics started as a basic ID verification system and has evolved into a major factor of authentication by taking several biological parameters as references. There are certain issues that researchers have yet been facing against hack or overriding. This chapter covers a comparative study of different biometrics and its methods used for safe authentication that improves security management for complicated scenarios.

Keywords Biometrics · ID verification · Authentication · Security · Hacking

K. Martin Sagayam (✉) · J. Felix Jacob Edwin · J. Sujith Christopher · G.V. Reddy
Department of ECE, Karunya University, Coimbatore, Tamil Nadu, India
e-mail: martinsagayam.k@gmail.com

J. Felix Jacob Edwin
e-mail: felixedwin.j@gmail.com

J. Sujith Christopher
e-mail: chrisbon95@gmail.com

G.V. Reddy
e-mail: vamsidharreddy@karunya.edu.in

R. Bestak (✉)
Department of Telecommunication Engineering,
Czech Technical University, Prague, Czech Republic
e-mail: robert.bestak@fel.cvut.cz

L.C. Hun
Department of Engineering and Technology,
Multimedia University, Cyberjaya, Malaysia
e-mail: chlim@mmu.edu.my

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_6

1 Introduction

Biometrics are automated ways of recognising a person with regard to their physiological or behavioural characteristic. The features of biometrics are fingerprint, iris, face, retinal, voice, signature and earprint. Today, biometrics has become the base of great opinions of highly secured identification and verification, i.e. security solutions. As transaction fraud increases and complexity in security breaches, the need for technology is becoming clearly visible. Biometric-based solutions are capable of providing person data privacy and private financial transactions. The necessity of biometrics is widely arising which can be observed in federal, state and local government, and commercial applications in the defence sector. Extensive networking enterprise for industrial infrastructures in the domain of security, investing, safe electronic banking, financial retail sales, governmental identities, health services and law enforcement are already being developed using these technologies. Confiding in these innovation exchanges is fundamental for a healthy growth in the worldwide economy. Using biometric systems for all objects is getting to be plainly advantageous and impressively it is more precise than techniques for security fields.

The three most vital variables utilised as a part of the security field for verification are as follows:

- To remember something such as—password, person's information such as birth date or PIN,
- To carry something such as—a smart key, card key or pass token,
- Something that you are—a biometric.

Biometrics can be classified into two basic types: physiological and behavioural model as shown in Fig. 1. It is very much predominant to secure and authenticate in an intelligent system. It cannot be shared information, overlooked or stolen, and forging is impossible practically in a biometric system.

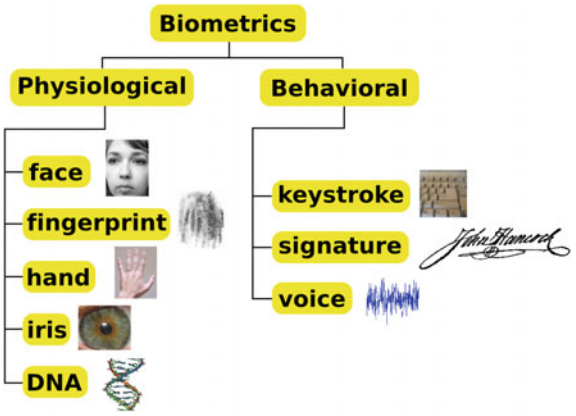


Fig. 1 Dynamic representation of biometric security system

2 Literature Survey

Biometrics has evolved in various aspects in different applications by using facial recognition, fingerprint recognition, retina, signature, voice and earprint [1]. This section describes deeply about the types of biometrics and its applications.

Facial recognition Face pattern analyses the characteristics of a face having certain points. Having a digital camera with transducer is needed which was suggested by Roberto in order to develop a facial pattern for recognition and authentication [2]. This method of authenticating had attracted the market and the society considerably. This is even compatible with a basic android device for the security of the same.

Fingerprint biometrics is obtained by various methods. While some imitate the traditional methods of obtaining the print, there are others which are a little more different and distinct, which also includes ultrasonic and moire patterns. Some methods of verification are capable of detecting a live finger while others are not. Various models of fingerprint modules are available comparatively to any other as suggested by Jain [3]. The cost of the FPR modules and processing system is less while taking the number of enrolments in count. FPR biometric security is a reasonable yet efficient system for a house security system.

Iris biometrics is a method of authentication process that provides a very high level of security and nevertheless requires very less effort for an eye-related biometrics, which uses only a conventional camera for obtaining the iris pattern to analyse the feature of the colour ring tissue surrounding the pupil of the eye. The data acquisition module does not require a very close contact with the user as it has been demonstrated by Ma et al. [4]. The technique uses image processing tools to identify each person's uniqueness from the pattern iris tissue.

Retinal biometrics is a procedure where the layer of veins arranged at the back portion of the eye is analysed. A low concentration light source is involved for this method to scan the patterns which are unique, that of the retinal through an optocoupler. Scanning of the retina does not require the person to look into the receptor focal point. It can be quite accurate, but it is not particularly convenient if the person wears glasses or has issues due to close contact with the scanning device have been explained by Cofta et al. [5] and Nixon [6]. Although the technique is very functional Because of the varying criterion, retinal scan process is not accepted widely by the end users.

Signature Biometric authentication is a method in which the characteristics of a signature from a user are captured through the parameters such as the name, speed, pressure and velocity in certain points [7]. The society was used to traditional thinking towards signature which was used for transaction verification, and many of them would not be surprised that in today's technology this technique is used for person's identification. The verification devices nominally uses an electronic pad which is used for signature employ precise methods to extract the vital data of authenticating and can lend themselves to any applications that requires a signature identity verification as presented by Cofta et al. [5].

Voice recognition analysis is actually a voice to text converted for biometrics and processed. This biometrics does not require a dedicated hardware; most PCs already consists of a quality inbuilt mic as experimented by Mallat [8]. The verification and enrolment are reflected on factors such as poor-quality audio and noise. The enrolment procedure is more complicated when compared with other biometrics.

Earprint biometrics This method employs image ray transform, a technique that is used for observing structures of different shapes such as eardrum in ears, making it possible to identify them, structures such as the helix, which is used as the base method for enrolment for ear biometrics as suggested by Meraayan [9]. And having said that, ears have an advantage over other biometrics which are being used traditionally and are emerging in today's technology because of the structure of the ear which is used for identification will remain stable from birth to the upper age band, and instead of ageing, it just gets bigger. However, the fact that hair can conceal ears, research is constantly moving forward for their use as a biometrics in the future and striving hard to come up with new techniques and algorithms to make it possible as proposed by Rutty et al. [10].

2.1 Problem Statement

Based on the performance metrics of different parameters involved in the security management system, the problems that arose due to the various parameters such as uniqueness, permanence, collectability, acceptability and circumvention for various models have been listed in Table 1.

The metric units represented in Table 1 are H—high, M—medium and L—low. The sample references for biological features have been displayed in Fig. 2. The data set consists of various templates in different orientation and backgrounds. The entire data will be preprocessed by using suitable image processing techniques before proceeding to next level.

Based on the various parameters which are considered for evaluation of each biometrics, Table 1 shows the comparative states of each biometrics. Figure 2 shows the representation of biometric system (a) fingerprint, (b) voice, (c) retina, (d) iris, (e) signature and (f) face. Sabanci University biometric research group has a

Table 1 Problem statement in various biometric system

Model	Circumvention	Permanence	Uniqueness	Acceptability	Collectability
Finger	H	H	M	H	M
Voice	L	L	H	L	M
Iris	H	H	L	H	M
Retina	H	H	L	M	L
Signature	L	L	H	L	H
Face	L	L	H	M	H

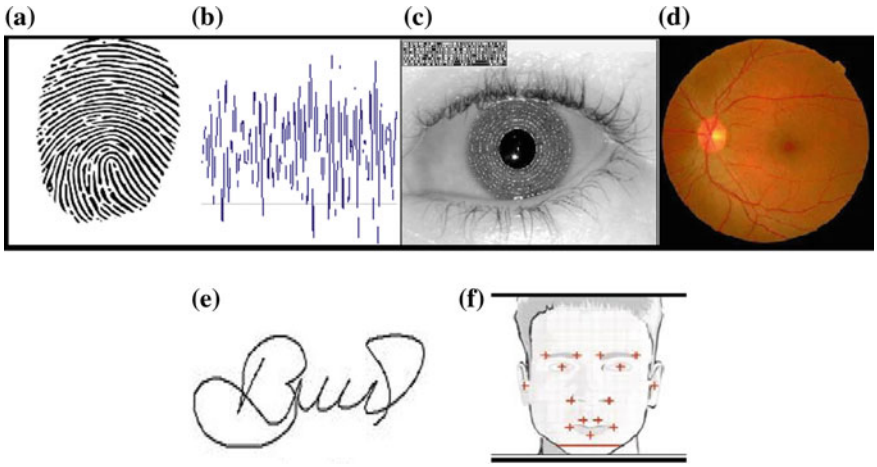


Fig. 2 Representation of biometric system. **a** Finger, **b** Voice, **c** Retina, **d** Iris, **e** Signature and **f** Face

new online signature data set used for training and testing for various biometric applications as shown in Fig. 2e [11]. The states represent the weight of each parameter to that particular biometric data [12].

3 Framework

The biometric data is initially acquired from the user, and the critical features are extracted from the data. By using robust technique, the person authentication has been processed. It can be adopted by any suitable classifier for identifying the respective feature vector points, for better performance and accuracy various techniques plays a significant role.

This chapter will be a reference guide for researchers and people who want to deploy the biometric security for several scenarios since the work covers comparative study in different biometric methods in further sections.

4 Various Methods of Biometric Security System

4.1 Facial Biometrics

Over the past 30 years, various techniques are used for system recognition of people's faces. A recorded obstacle for face recognition was taken into consideration of the early stages of computer vision, which is currently on process of

development in an established manner [13]. Different technological proposals which are updated recently, where classified mainly into two types that inducing a point to point nodes corresponding in-between the stimuli which results a facial pattern to be recognised and facial pattern stored into data set [2].

(i) **Geometrical feature-based matching**—Even when the featured parameters (which are nose, mouth and the eyes) are not resolved, yet facial pattern is recognised. The rest of the information is purely based on geometrical representations, of what has remained in a true coarse resolution. The goal is to obtain the relevant spots of other parameters having a unique feature which includes nose, mouth, chin and eyes.

(ii) **Template Matching**—Basic model for templet matching for the image that was drawn into bi-dimensional array for intrinsic values is compared employing an appropriate measurement with one model representing the complete face [2].

(iii) **3D Face information acquisition**—3D facial recognition or other ways knew 3D FR-based technique must possess certain properties such as robustness on light varying in rotation, scaling and position of the true model [14]. It is quite unfortunate that the 3D FR technique does not attain all the suggested goals entirely. Generating a 3D pattern of a human face, there are two main ways: the first method is by using a 3D Scanner, and it is acquired by a sequence of steps as given below:

   a. 3D data are compiled respective to the predetermined referral system for image capturing module.
   b. 3D data from various point of computer vision undergo a fusion procedure.
   c. 3D data are optimised. Lastly, the 3D polygonal connections are built.

The alternate ways to obtain a 3D facial template can be utilised as a morphable template.

(iv) **Optimisation of feature sets**—For a face recognition task to be completed, the geometric features are enough. This will require time for the computations and deficit memory due to their low dimension. The facial geometric feature set consists of about 26 features, out of which 14 features are obtained by the facial feature projections. Whereas the remaining 12 are from the face detector [15, 16].

4.2 Fingerprint Biometrics

A fingerprint is formed through the impression of a finger in any opaque surface due to the friction ridges present in all fingers of the person. The recovery of the same from a crime spot is an essential technique for forensic science that would be

useful in capturing the person who is guilty of the crime. The significance of fingerprint which can be deposited easily on surfaces due to the natural secretions in the epidermis with the help of sweat glands in finger ridges. This phenomenon is referred as 'Chanced Impressions' [17, 18]. In other terms, fingerprints are the traces of a person's impression on a surface [3]. Hence, mark of their impression may remain unchanged for several hours because of their sweat glands making it affable for an impression on the surface.

(a) **Uniqueness and obstacles**—Human fingerprints are detailed, unique for both hands that are difficult to meddle with, and they remain unchanged throughout his/her lifetime of every person and thereby considering the fingerprint parameter for authentication of human identity will be effective for a longer period. These techniques have been empowered in national identification cards, smart cards and other forensic purposes to identify the particular person, or to find out a person who is not well or debilitated thus being not able to recognise their own, in a scenario related to natural calamities [16]. Analysing fingerprint data provides us with a moderate security; it is ranked highly effective among other biometrics and mobile way to authenticate a person's identity. Recent technology trends have involved in transferring funds from bank accounts through smartphones procured with fingerprint sensors [17]. With an error rate of 1–500, fingerprint is practically employed in mobile electronic devices.

(b) **Algorithm used**—In order to compare and match two fingerprint images obtained through sensors, Short-Time Fourier Transform (STFT) [18, 19] has been used to analyse and for the overall clarity improvement of the fingerprint image and also for providing its binary format. From the following steps

Step 1: Finding common points,
Step 2: Tuples in base model (BM),
Step 3: Tuples in input image (IM),
Step 4: Matching Phase,
Step 5: Finding confirmed common points list.

Figure 3 shows a flow diagram of fingerprint enrolment and authentication of every individual and transformed into binary codes. The finger pattern is stored and processed which is encrypted between servers. A performance score has been

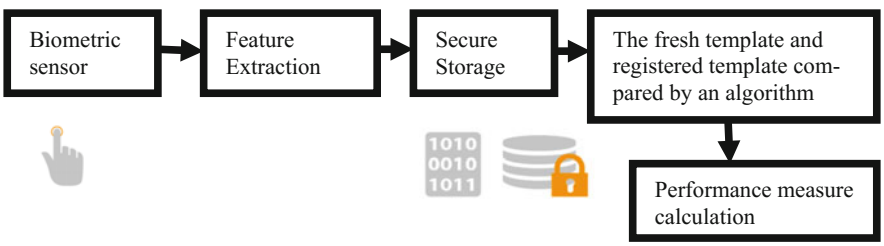


Fig. 3 Acquisition and processing of fingerprint verification

calculated between two templates of the fingerprint, to verify one person's identity [3]. Squares in print represent minutiae points which are potential information for obtaining finger patterns, and the circles are ignored because they contain standard data similar to other patterns [16, 17].

4.2.1 Method of Evaluation

Step 1: Taking each feature point $x = 1$ to N_1, the five consecutive feature points are determined. A Euclidean distance is obtained by calculating from the xth feature point to all the other subset of features in the data set 'N'. The standard feature points are denoted in rectangle and ellipse as in Fig. 4 [20].

Step 2: If $x_1 \ldots x_5$ are the nearest feature points of x, then $Z(x)$ is calculated as shown below:

(a) Calculating the distances $x - x_1$, $x - x_2$, $x - x_3$, $x - x_4$ and $x - x_5$. Observe that distance '$x - x_N$' represents the Euclidean distance between the feature points x and x_N. Therefore, the Euclidean distance has been obtained for the feature points in the subsets.

(b) To find the following 10 ratio points such as $(x - x_1):(x - x_2)$, $(x - x_1):(x - x_3)$, $(x - x_1):(x - x_4)$, $(x - x_1):(x - x_5)$, $(x - x_2):(x - x_3)$, $(x - x_2):(x - x_4)$, $(x - x_2):(x - x_5)$, $(x - x_3):(x - x_4)$, $(x - x_3):(x - x_5)$, $(x - x_4):(x - x_5)$.

In general, the above equation can be rewritten as follows:

$$(a-b):(a-c) = Z\max\{(a-b),(a-c)\}/Z\min\{(a-b),(a-c)\}.$$

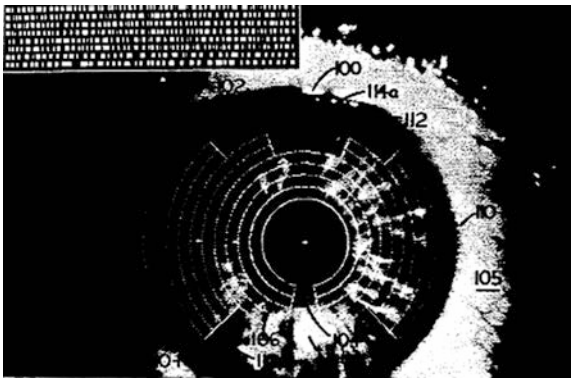


Fig. 4 Infrared image of an iris using greyscale reduction

4.3 Iris Biometrics

Iris is a thin muscular layer which is located above the pupil of the eye which is circle in shape when exposed to light; it is responsible for regulating the pupil's diameter. Iris forms a very complex structure of blood vessels and muscular tissues, and this complex structure can be used to identify every person's identity with remarkably low false match rates [8]. Iris is proven to be an extreme level of biometric authentication technique that promises a high security level for protecting confidential premises.

(i) **Uniqueness and obstacles**

Examined by both ophthalmologist and anatomist, it has been found that iris possesses extreme specificity from person to person in both structure and colour, and considering a large number of eyes, it has been found that either side of every person's eyes shows distinctness in their iris patterns [4]. It is even said that the patterns appeal to be varying little after child age. While extracting the general iris pattern, some of the dependent futures for obtaining the geometry of the iris on embryonic precursor and the size of the pupil is only matured after crossing the period of childhood to the term of adolescence [21]. However, once out of the teen years, the person's iris is likely to be unaltered throughout his lifespan.

Iris has different colours so obtaining a pattern for colours which is not corporative to light is a common factor that we face, in order to reduce the effect many iris recognition devices use the video imaging technique combined with the illumination of near infrared (NIR) which gives more detail to the pattern when compared with the images produced at visible wavelength (VW) [16, 22] as shown in Fig. 4.

The pattern focuses on two bands of muscles from the iris which is responsible for controlling the diameter of the pupil to be exposed: the dilator contracting the muscles to enlarge the diameter pupil and sphincter which will contract in order to shrink the size of the pupil [21]. After obtaining video image frames of the iris, it can be optimised for biometric identification using pattern-recognition techniques.

A digital image is produced by a camera by scanning a person's eye [4].

1. The software for processing the image will extract the image by drawing two circles; they are placed in two boundary levels at the internal and external limits of the iris (limbus, between the white sclera and iris). The inner part is relatively easy to detect because of the circular shape with a change in brightness instantly where the pupil will give way to the iris. A similar process is used to find the boundary of the outer eye, though it has to allow for the likeness of the eyelids blocking a part of the iris.
2. To define individual areas of analysis, the polar coordinates are added to the image. The way the iris changes with respect to the pupil dilating, the module will allow it to get narrower in different conditions with respect to light. Hence, the key feature of iris could be determined accurately and contrasted in 2D space.

3. The dark and light areas in the iris pattern are then transformed in the form of digital using bandpass filters (If the luminosity in a specific area is more than minimum threshold, filters will register 1. If not, it would register 0) and with a bit of mathematical juggling, this generates a pattern named iris code.

(ii) **Algorithm**

From 1990s, numerous scientists and researchers have been working on this algorithm and have classified four consecutive steps to obtain the iris biometrics.

1. Localisation—To determine the position of the boundaries, both the interior and exterior of the iris.
2. Normalisation—The captured iris of varied people may be in distinct size, so as to reduce the error factor. And even in a specimen, the size tends to change because of variation in lighting while capturing the image.
3. Feature extraction—When an iris is captured, it exhibits many features and information about the texture. Sequential ordered features are obtained from different iris representation, which are also called as a feature vector.
4. Matching—The captured iris image's feature vectors are evaluated by different threshold matching techniques such as distance of hamming, selection of winner and weight vector and function of dissimilarity [23].

(iii) **Comparison chart of different algorithms**

The performance measures of various algorithms for iris recognition are analysed in Table 2.

4.4 Retinal Biometrics

Retinal biometrics is one of its kind the most secure and an alternative to DNA-based identification, which has a very less possibility of an error to a ratio of 1 in 10,000,000, [24] compared with fingerprinting can be as high as 1 out of 500, retinal scanning is very unique which will be essentially used on daily bases. Retinal scanning is not to be confused with iris scanning which involves in observing the pattern of human iris. A thin layer of tissue comprising of neural cells

Table 2 Performance measures of various algorithms

Algorithm	FAR	FRR	Overall accuracy
Avila [1]	0.033	2.088	98.899
Li Ma [2]	0.022	1.988	98.8
Tisse [3]	1.844	8.799	88.377
Daugman [4]	0.011	0.099	98.99

Where *FAR* false acceptance rate and *FRR* false rejection rate are compared with respect to its overall accuracy.

that are located beneath the iris is called retina and is made up of complex connections of blood vessels, and even identical twin does not have the same blood vessel configuration.

(i) **Uniqueness and obstacles**

There is a higher possibility that retinal patterns may alter in cases of various disorders such as diabetes, glaucoma or other eye retinal disorders, but typically the retina remains without any change from birth until death [25, 26]. Since the method uses expensive equipment and methods to obtain the pattern of the blood vessels, retinal scanning is the least form of biometrics deployed in the market [5]. There are definite advantages among which resist the error rate due to false matching and also it is static for the retina, similar to the fingerprint which remains stable physiologically throughout one's life. Technology that has certain characteristics like robust matching is usually programmed to initiate one against many templates in the identifying a particular user in the database. This technology also has certain drawback in which the user cannot enrol him to the system unless the image of his/her retina captured is of a very high resolution. The area in which the retina is located has a default resistance towards environment for any accidental conditions that may alter the structural nodes.

An estimated error rate done by state courts of national government which shows the error rate for false matching in scanning is one for ten million enrolments, and this technique requires prolonged concentration. It also requires a motivated and well-trained person to use the device. The scanning for the retinal enrolments of every pattern takes time more than both iris scanning and fingerprinting technique [25]. The scanning of the retina is done by bursting out a beam of unperceived light source which is a very low power infrared illumination into the person's eyes when he looks through the device eyepiece. The fact, that it is hard to use the retinal scanning technology causing discomfort and requires a quite amount of time to enrol every retinal pattern is a very great disadvantage for this high level of security providing devices. The drawback incorporates the way that the innovation is difficult to utilise. After the beam is passed, it traces a standardised path on the retinal blood capillaries that are present in the retina absorbs light much rapidly than the other neighbouring muscular tissues, the reflection of infrared light will be corresponding to the person's retina during the scanning progress. Variation of the pattern is digitised and stored in a database as shown in Fig. 5.

(ii) **Operation**

Retinal technology is an image acquisition challenge in a row of scanty biometrics. The retinal layers are small and inbound, demanding quite peculiar hardware and software. The person must place his/her eyes nearer to the device's protruded lens. While the client looks directly into the lens and remains still, a retinal image is obtained, and development or any aggravation will result in another attempt of the acquisition process. The acquired image is deduced to 192 points of reference before it is extracted into a digitised 96-byte templet and pushed into memory

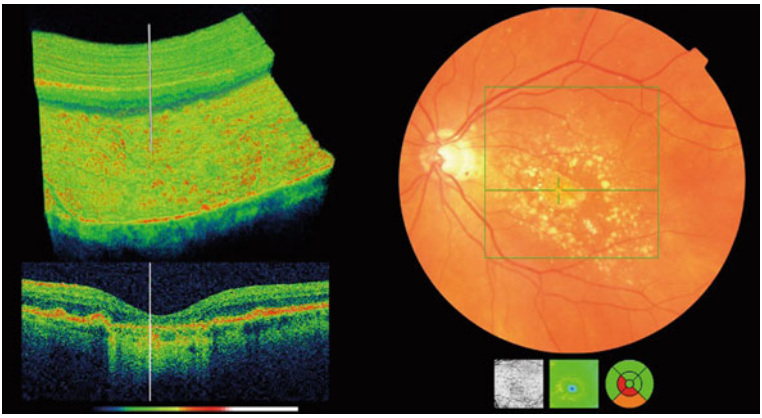


Fig. 5 Retinal biometric acquisition

storage for sequential authentication. Usually, it takes a few couple of proper images to ensure enrolment. Hence, the process is lengthy by taking many still images of the retina. Enrolment procedure takes more than 1 min and yet some may not be able to enrol themselves because of some factors faced due to disorders, errors in the environment or even retinal infections. Scanning the vascular structure of a retina, a low-intensity light source is utilised. A scan of 360° circular motion is necessary for the retinal area for over 400 readings step by step to determine the pattern of blood vessels [21, 27]. It seems that the more a user understands and get experienced to process, sooner the process is executed. After the acquisition of image, a compilation of the features with the uniqueness of the retinal blood vessels into a templet by using software.

4.5 Signature Biometrics

A handwritten signature is widely studied for verification and usage as a biometrics [6]. The general classification of handwritten signature is divided into two types: those are the online and offline verification approaches. Offline verification is the regular way of dealing with signature information in a statically pattern, while online method may require information or feature about the signature which includes acceleration, writing speed, pressure points, strikes and also the static characteristics of signatures. The online extraction requires a module that can be pen input method connected to an electronic tablet which will observe the signature of the user; these dynamic ways of extractions are difficult to duplicate. Application of this includes credit card validation, banking and security [28].

A significant advantage of this biometric system is the fact that the person need not necessarily remember the password or any other sensitive data that is defined

for authentication and identification purpose. Signature verification has the advantage of a very high user acceptance because people are used to signing in the day-to-day life [7]. Signature authentication systems are said to be static or dynamic. Static authentication uses a static digitalised image of the signature. The manual verification says the person's signature is often unrealistic. Thus, having certain unrealistic conditions thereby not much into the advancement.

4.6 Voice Recognition

The technique which is used to analyse the speech recognition is discussed in detail. Speaker that enables the recognition and transition speaking languages into text (into digital codes) understanding by computer languages (i.e. machine languages) [9]. The framework of voice recognition system is shown in Fig. 6.

To frame signal, the electronic device is used. Fig. 7 shows the speaker acquires the voice (or) any input speech by person.

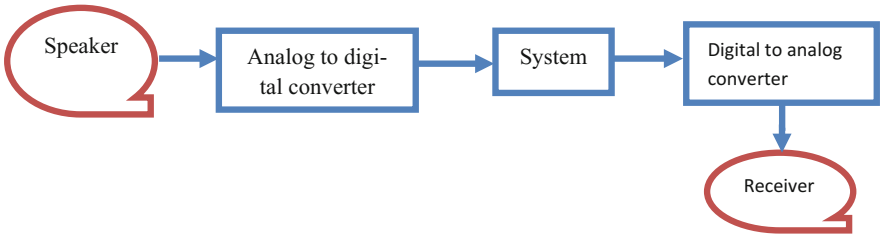


Fig. 6 Flow diagram of natural voice recognition

Fig. 7 Structure and features of the human ear

In the year of 1932, Bells Labs researchers have invested in speech recognition laboratories, but a system for digit signal recognization and as implement the formats of codes. In the middle of 1950, the data are stored started with first ten words with single speaker system with vocabularies [4]. Due to some reasons, the Bells Labs researchers stopped again. Raj Reddy was the first person to start the coding of the continuous signals. In 2000, the DTW (dynamic time warping) was formed, and algorithm of this DTW was divided into short signals (significant signals are segregated into smaller signals, and they are named as frames) [25, 26].

Classification methods used

By surviving different methods of classification of intelligence based on biometric techniques, there are four major classifiers which are used frequently in biometric system for better accuracy and recognition rates as suggested by Ashbourn [25] and Leonidou [26]: hidden Markov model (HMM), neural network (NN), dynamic time warping (DTW), deep feed-forward and recurrent neural networks.

4.6.1 Hidden Markov Model (HMM)

There are some statistical techniques that the outputs of the symbols HMM are used in SR because of a speech signals. It is viewed as a peak stationary signal (or) a long-time stable signals (or) short-time stationary signals, and also it is dependent hidden models. HMM is a popular method because it is very easy and feasible to use in SR. The hidden Markov models output a sequence of n-dimensional for real value factors that provides output every 10 ms. The factors consist of deep real coefficients and mathematical equations by Fourier's transformations [12, 15].

4.6.2 Dynamic Time Warping (DTW)

Historically used in SR but it is used to stay for long time displaced and DTW is backbone of the HMM method [12, 15]. HMM and DTW both have some similarities, and for example, if one video walking is fast and another video walking is slow the acceleration and de-acceleration during observation, DTW can apply voice music and audio graphic, the output of the linearity and nonlinearity can be compared by the DTW.

4.7 Earprint Biometrics

The earprint biometrics is barely attended to with relevancy different fashionable modern techniques of the automated face, eye or unique fingerprint impression. In any case, earprint assumes an amazing part in crime investigation and rhetorical

science. Since the development of non-automated measurements (manual) by Iannarelli, ear organisation was supporting the cause for about forty years, particularly within USA [27].

Ear structure certainly has an edge advantage over the other established biometrics; the structure of the ear is vibrant and stable structure which undergoes little change with respect to age. This structure (ear) of the body as shown in Fig. 7, will not suffer changes because of facial expression variation and is steadily bound on the either side of the head so that the immediate sound of the extra facial surface is predictable, whereas face recognition or other established biometrics requires the face to be captured against a defined background [29]. Any relevant hygiene issue is not in the picture while acquiring the image or data would be the condition of contact biometrics. The ear is bigger compared to the fingerprint, retina and iris without more effort captured at a distance.

The interior structure of the ear includes the ear lobe and helix. The prominent feature of the structure such as the antihelix and the helix that is just parallel of the helix and distinguishable shape just above the lobe of the ear that is similar to a hairpin bend called inter tragic notch [29]. The central area is named concha, shown in Fig. 8.

As seen of the established biometrics, earprint biometrics is seemingly a rising star among the others. The earprint is obtained by two types of acquisition: they are image and sonic, respectively, as shown in Fig. 9.

For the image, the projection of the ear is made, and it is mathematically derived as

$$P = DS\left(D^{\mathrm{T}}D\right)$$

Let p be any of the sample picture,
Let m be the picture's average,
Let $d = p$ difference between 'p' and 'm',
The array obtained by the columns of difference picture 'd' is D.

Fig. 8 Various ear samples for study

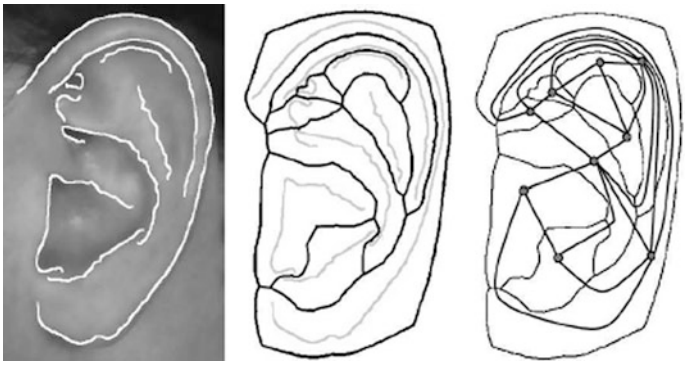


Fig. 9 Image tracking points in an earprint

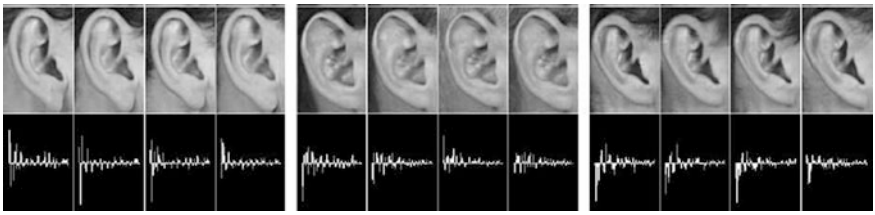


Fig. 10 Sonic response creating a signature wave

It seems that the ear signal plays a major role based on its distinctive shape. It behaves like a filter so as to be turned as a modified form as shown in Fig. 10. This acoustic or audible transfer function provides the acoustic effect of ear signature [6, 10].

5 Performance Measures

Performance measurement is the process of acquiring, analysing and reporting information to the system. Biometric performance is calculated for the environment of the user. The performance portfolio for the biometric security management is shown below:

(a) **Level of security** of the biometrics is evaluated according to the highest gain output it provides, such as the performance stats such as uniqueness, resistivity to external factors and durability.
(b) **Level of accessibility** of the biometric can be defined as the ease of usage and availability of the product in market.

Table 3 Formula of performance metrics in confusion matrix

Measure	Derivations
Sensitivity	TPR = TP/(TP + FN)
Specificity	SPC = TN/(FP + TN)
Precision	PPV = TP/(TP + FP)
Negative predictive value	NPV = TN/(TN + FN)
False positive rate	FPR = FP/(FP + TN)
False discovery rate	FDR = FP/(FP + TP)
False negative rate	FNR = FN/(FN + TP)
Accuracy	ACC = (TP + TN)/(P + N)
F1 score	F1 = 2TP/(2TP + FP + FN)
Matthews correlation coefficient	((TP * TN) − (FP * FN))/sqrt ((TP + FN) * (TP + FP) * (TN + FP) * (TN + FN))

(c) **Confusion matrix** can be defined as a table that can be used to define the performance values of a classifier's model based on a group of analysing data, where the true values are pertained already. The list of formulae of performance metric in confusion matrix is shown in Table 3.

Measure parameters:

TPR Sensitivity or True Positive Rate,
SPC Specificity or TNR-True Negative Rate,
PPV Precision or Positive Predictive Value,
NPV Negative Predictive Value,
FPR Fall-out or False Positive Rate,
FDR False Discovery Rate,
FNR Miss Rate or False Negative Rate,
ACC Accuracy,
F1 F1 Score,
MCC Matthews Correlation Coefficient.

6 Result Analysis

The various established biometric systems show the viability of it in some or the other parameter in this vastly upgrading society that paved the way for newer inventions and technologies. We have come across facial, finger, eye, signature and earprint as shown in Fig. 11.

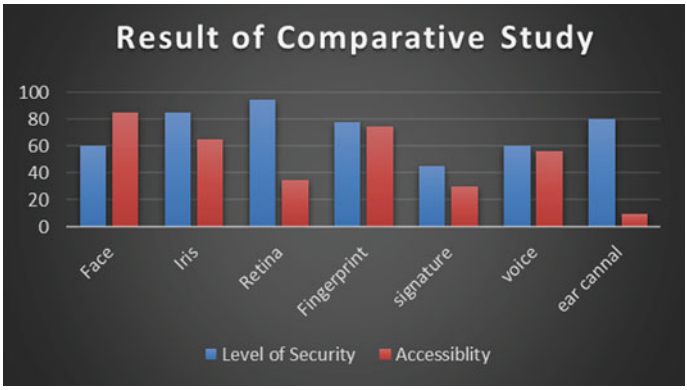


Fig. 11 Comparative analysis: level of security versus accessibility

Table 4 Performance values of facial recognition

Measure	Obtained value
Sensitivity (TPR)	0.5055
Specificity (SPC)	0.6563
Precision (PPV)	0.7360
Negative prediction value (NPV)	0.4118
False positive rate (FPR)	0.3438
False discovery rate (FDR)	0.2640
False negative rate (FNR)	0.4945
Accuracy (ACC)	0.5576
F1 score (F1)	0.5993
Matthews correlation coefficient (MCC)	0.1546

The ear lobe structure, when taken as biometric data, its no greater in its infancy, and it has been showing continually pushing forward the progress so far, its appears will not change with respect to age and also in terms of individuality than other biometrics, it is a booming interest specifically motivated by researchers and scientists into the three-dimensional modelling, thus prominently producing a wide range of research opportunities and inventions (Tables 4, 5, 6 and 7).

The above chart has been evaluated through market analysis of security system. From Figs. 12 and 13 have been understood that the availability of individual biometrics has to be improved regarding their performance measures (Tables 8 and 9).

Table 5 Performance values of fingerprint recognition

Measure	Obtained value
Sensitivity (TPR)	0.5043
Specificity (SPC)	0.4972
Precision (PPV)	0.3919
Negative prediction value (NPV)	0.6096
False positive rate (FPR)	0.5028
False discovery rate (FDR)	0.6081
False negative rate (FNR)	0.4957
Accuracy (ACC)	0.5000
F1 score (F1)	0.4411
Matthews correlation coefficient (MCC)	0.0015

Table 6 Performance values of iris recognition

Measure	Obtained value
Sensitivity (TPR)	0.3600
Specificity (SPC)	0.4946
Precision (PPV)	0.2769
Negative prediction value (NPV)	0.5897
False positive rate (FPR)	0.5054
False discovery rate (FDR)	0.7231
False negative rate (FNR)	0.6400
Accuracy (ACC)	0.4476
F1 score (F1)	0.3130
Matthews correlation coefficient (MCC)	−0.1392

Table 7 Performance values of retina

Measure	Obtained value
Sensitivity (TPR)	0.4928
Specificity (SPC)	0.3816
Precision (PPV)	0.2656
Negative prediction value (NPV)	0.6237
False positive rate (FPR)	0.6184
False discovery rate (FDR)	0.7344
False negative rate (FNR)	0.5072
Accuracy (ACC)	0.4163
F1 score (F1)	0.3452
Matthews correlation coefficient (MCC)	−0.1180

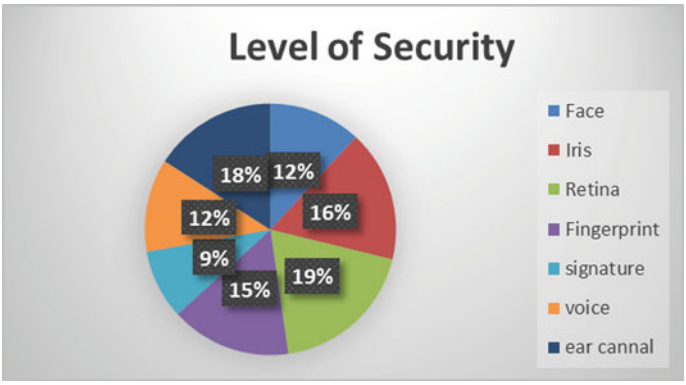


Fig. 12 Comparative analysis: level of security

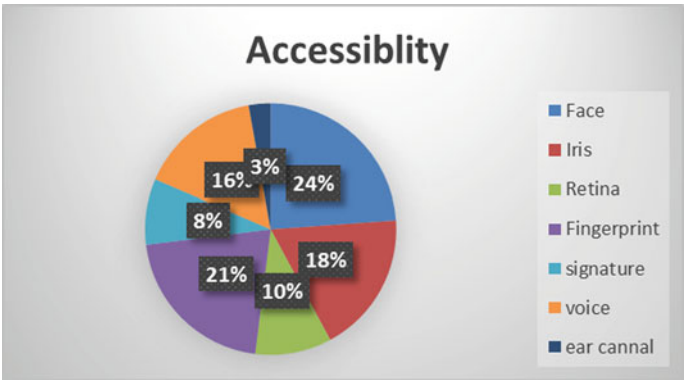


Fig. 13 Comparative analysis of accessibility

Table 8 Performance values of signature

Measure	Obtained value
Sensitivity (TPR)	0.5056
Specificity (SPC)	0.5075
Precision (PPV)	0.7339
Negative prediction value (NPV)	0.2764
False positive rate (FPR)	0.4925
False discovery rate (FDR)	0.2661
False negative rate (FNR)	0.4944
Accuracy (ACC)	0.5061
F1 score (F1)	0.5987
Matthews correlation coefficient (MCC)	0.0116

Table 9 Performance values of voice

Measure	Obtained value
Sensitivity (TPR)	0.5974
Specificity (SPC)	0.4667
Precision (PPV)	0.7419
Negative prediction value (NPV)	0.3111
False positive rate (FPR)	0.5333
False discovery rate (FDR)	0.2581
False negative rate (FNR)	0.4026
Accuracy (ACC)	0.5607
F1 score (F1)	0.6619
Matthews correlation coefficient (MCC)	0.0583

7 Conclusions

Through the analysis of different types of biometrics based on their uniqueness and their importance for authenticating a person's identity, it has been clearly understood that the biometrics paves a path for the future security management system and can be used in our day-to-day applications. Thus, concluding that from the measure of high security and accessibility analysis in the market, retinal scanning and face recognition have marked its predominance in the security system and management, respectively.

References

1. Yeung B. A San Francisco Bay area solution provider for IT knowledge management: webcache.googleusercontent.com/search?q=cache:http://www.tns.com/biometrics.asp&gws_rd=cr&ei=btGNWNyCF4rxvgSXmonYCw
2. Brunelli R, Poggio R (1993) Face recognition: templates versus features. IEEE Trans Pattern Anal Mach Intell 15(10)
3. Jain AK, Prabhakar S, Chen S (1999) Combining multiple matchers for a high-security fingerprint verification system. Pattern Recogn Lett 20(11–13)
4. Ma L, Tan T, Wang Y (2002) Iris recognition using circular symmetric filters. In: Proceedings of the16th international conference on pattern recognition, vol 2, pp 414–417
5. Cofta P, Furnell S (2008) Use of biometric data. In: Understanding public perceptions: trust and engagement in ICT-mediated services. International Engineering Consortium, p 153
6. Nixon M (2010) A novel ray analogy for enrolment of ear biometrics, IEEE
7. Iannarelli A (1989) Ear identification. Paramount Publishing Company, Freemont, California
8. Mallat SG (1991) Zero-crossing of a wavelet transform. IEEE Trans Inf Theory 37(14):1019–1033
9. Meraayan A (2003) Proximity authentication. Sans Institute
10. Rutty GN, Abbas A, Crossing D (2005) Could earprint identification be computerised? an illustrated proof of concept paper. Int J Legal Med 6:335–343
11. http://biometrics.sabanciuniv.edu/susig.html
12. Bhatia R (2013) Biometrics and face recognition techniques. Int J Adv Res Comput Sci Softw Eng 3(5)

13. Hiremath PS, Ajit D, Prabhakar CJ (2007) Modelling of uncertainty in representation of facial features for face recognition and detection. Gulbarga University, Gulbarga, JNN College of Engineering, Shimoga & Kuvempu University, Shimoga, India June 2007, pp 189–194
14. Nanvati S (2002) Biometrics: identity verification in a networked world. Wiley, New York
15. Ashbaugh, DR (1999) Quantitative-qualitative friction ridge analysis: introduction to basic and advanced ridgeology 27 Oct 1999
16. Gold S, Rangarajan A (1996) A graduated assessment on algorithm for graph matching. IEEE Trans Pattern Anal Mach Intell 18(4):377–388
17. Beavan C (2001) Fingerprints: the origins of crime detection and the murder case that launched forensic science. Hyperion, New York
18. Cowger JC (1992) Friction ridge skin: comparison and identification of fingerprints. CRC Press, Boca Raton, Florida
19. Scheibert J, Leurent S, Provost A, Debregeas G (2009) The role of fingerprints in the coding of tactile information probed with a biomimetic sensor. Science 323:1503–1506
20. Marchant Z (2003) Biometrics: fingerprint authentication. Sans Institute
21. Jain AK, Ross A, Prabhakar S (2004) An introduction to biometric recognition. IEEE Trans Circ Syst Video Technol 14(1)
22. Quinche N, Margot P (2010) Coulier, Paul-Jean (1824–1890): a precursor in the history of fingerprint detection and their potential use for identifying their source (1863). J Forensic Ident (California) 60 (2):129–134
23. Tisse C-L, Torres L, Robert M (2002) Person identification technique using human iris recognition. In: Proceedings of the 15th international conference on vision interface
24. Daugman JG (2001) Statistical richness of visual phase information: update on recognising persons by Iris Patterns. Int J Comput Vis 45(1)
25. Ashbourn J (2002) Biometrics: advanced identity verification. Springer, London
26. Leonidou M (2002) Iris recognition: closer than we think? Sans Institute
27. Jain A et al (2004) Handbook of biometrics. In: Man convicted of murder by ear print is freed. www.timesonline.co.uk/article/0,1-973291,00.html, 22 Jan 2004
28. Zimmer A, Ling LL (2003) A hybrid on/off-line handwritten signature verification system. In: Seventh international conference on document analysis and recognition, vol 1, Aug 2003, pp 424–428
29. Hurley DJ, Arbab-Zavar B, Nixon MS (2008) The ear as a biometric: handbook of biometrics. Springer US, pp 131–150. doi:10.1007/978-0-387-71041-9_7

Spatial and Spectral Quality Assessment of Fused Hyperspectral and Multispectral Data

Somdatta Chakravortty and Anil Bhondekar

Abstract Hyperspectral sensors collect images in hundreds of narrow, continuous spectral channels, whereas multispectral sensors collect images in relatively broader wavelength bands. The spatial resolution of a hyperspectral image is, however, lower than that of a multispectral image. This study has integrated the high spectral and high spatial information of hyperspectral (Hyperion) and multispectral (LISS-IV) imagery of the Henry Island of Sundarbans, India. An integrated image has been successfully generated that has 5 m spatial resolution of LISS-IV image and 10 nm spectral resolution of Hyperion image. The prime objective of this study is to obtain an image of enhanced spectral and spatial resolution that would facilitate accurate interpretation and mapping of mangrove types, sea and creek water, pisciculture water, agricultural soil and saline soil of the study area. The methodology adopted considers band remapping of different spectral regions covered by multispectral and hyperspectral images. This study has applied algorithms to restore spatial information of hyperspectral data by integrating hyperspectral bands with those multispectral bands that fall within its range. After comparison of the spectral profile of fused and original hyperspectral image, similarity is found in the spectral curves across the electronic spectrum indicating that the spectral signature of the fused image is maintained. High values of crosscorrelation between the images show that both the spectral and the spatial information are well preserved. The quality evaluation of the fused image is based on quantitative criteria which include spatial and spectral properties that are defined in the image. This study assesses the quality of fused image through utilization of various statistical indicators. Spectral quality of integrated image is assessed using parameters such as spectral discrepancy, correlation coefficient, values of root mean square error (RMSE), Spectral Angle Mapper (SAM), standard deviation and signal-to-noise ratio (SNR) are used for spectral quality assessment of the fused image. Spatial quality assessment has been done using bias and Edge Detection techniques. After analysis of application of these indicators on the fused and the original hyperspectral image, it is observed that the spectral discrepancy shows low discrepancy values in the visible and

S. Chakravortty (✉) · A. Bhondekar
Government College of Engineering and Ceramic Technology, Kolkata, India
e-mail: csomdatta@rediffmail.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_7

near-infrared (NIR) part of the electromagnetic spectrum. The correlation coefficient has given good results in the NIR region as compared to visible region. The results obtained from RMSE show that its values in the visible region are lower than that in the NIR region. SAM performs well in the NIR region as compared to the visible region. The standard deviation of fused and hyperspectral images is very similar and their differences are either close to zero or negative. The values of SNR calculated between fused and hyperspectral images vary randomly with the spectral wavelength. The random variability of the signal over the image (i.e. noise) is very less, thus giving a good SNR value. The result obtained from bias shows a good-quality fused and hyperspectral image with the ideal bias values varying with the spectral wavelength of the fused image. The results thus obtained after application of the indices show that the spectral and spatial property of fused image is close to its ideal value. Neural network has been used to assign classes to pixels of the new integrated image.

Keywords Bias · Entropy · Average gradient · Spectral · Discrepancy · Hyperspectral

1 Introduction

Hyperspectral images with high spectral resolution provide better characterization of the physical structure of ground constituent data. The hyperspectral image contains more than hundred continuous spectral bands from blue to shortwave infrared wavelengths ranging from 0.4 to 2.5 mm with 242 bands. Each pixel in hyperspectral data covers an area of 900 m^2 (30 m spatial resolution) signifying low spatial resolution. The coarse spatial resolution of hyperspectral data gives rise to the occurrence of mixed pixels where multiple endmember types may coexist within the same pixel. This results in inaccurate pixel representation of endmember types. Alternatively, multispectral image has a good spatial resolution of 5.8 m. However, it has only 3 broad wavelength bands, namely green, red and near-infrared band of the electromagnetic range signifying low spectral resolution. Low spectral resolution of multispectral data does not allow characterization of subtle absorption and reflectance peaks of signature spectra of objects that hampers discrimination between objects having similar characteristics. Integration of fine spectral characteristics of hyperspectral data and good spatial characteristics of multispectral data are expected to enable more accurate target detection and classification of its endmembers [1].

Literature survey indicates that there has been a lot of research in integrating multispectral images captured by different sensors with varying spatial resolution. Multiple techniques exist and have been applied for data fusion, specifically, for integrating multispectral and panchromatic images. Multispectral and hyperspectral data sets have also been integrated in [2] that has provided more reliable mapping of road, building, water and grass. Hyperspectral image has also been fused with panchromatic data with good spatial resolution using algorithms such as intensity

hue saturation (IHS), principal component analysis (PCA) and fast Fourier transform (FFT) and their performance evaluated [3]. PCA [3–5], the most commonly used technique, performs a transform of principal components on multispectral data by replacing principal components with high-resolution bands. The Gram–Schmidt pan-sharpening process simulates a panchromatic band from the bands of lower spatial resolution [5, 6]. The Colour Normalized (Brovey) transform [3] is a sharpening technique that uses an amalgamation of the coloured image and high-resolution image data. The HSV (hue saturation value) pan-sharpening algorithm transforms an RGB-coloured image to HSV colour space [7, 8].

A good fusion method preserves the spatial and spectral quality of the input multispectral and hyperspectral images, respectively. In [7], the quality of spectra of integrated hyperspectral image is assessed by root mean square error, spectral discrepancy, correlation coefficient and Spectral Angle Mapper; spatial quality is assessed by high-pass filter, entropy, edge detection and average gradient. In remote sensing, there are many fusion methods such as mean matching, regression methods, correlation modelling used for producing high-resolution multispectral images from panchromatic and multispectral images, respectively. Algorithms such as standard deviation, entropy, correlation coefficient, signal-to-noise ratio and root mean square error have been used to analyse the quality of information of integrated hyperspectral and multispectral images [9]. Global-local image quality analysis (GLIQA) is very useful for analysing the fusion quality of the entire image. The similarity between the fused image and source image is compared using Universal Image Quality Index (UIQI), which is the extended version of GLIQA that shows important information of source image and fused and hyperspectral image [10]. For successful identification of the ground-level data, it is desirable to have high spatial and spectral resolution. In the new generation of high-resolution satellite sensors, panchromatic images are available in less than 1 m spatial resolution. Algorithms such as PCA and Brovey and Ethlers transforms have been applied to integrate panchromatic and multispectral data of Quickbird sensor [11]. Quantitative indicators such as gradient, correlation coefficient and standard deviation have been used for evaluation and comparison of fused and original input image [12]. Integration of hyperspectral and high spatial images has also been attempted for improving extraction of biophysical parameters and assessment of physiological characteristics of agriculture crops [13–17].

This study has obtained an integrated image that has the spatial information of multispectral data and spectral information of hyperspectral data. The methodology in this study has resampled spectral bands of endmembers in the green, red and infrared regions of hyperspectral data to their respective wavelength regions in multispectral data. New fused pixels are generated after applying linear spectral unmixing on hyperspectral data with the estimated abundances. The fused image contains spatial and spectral details of multispectral and hyperspectral data, respectively, enabling more accurate target identification. Results showing high values of normalized crosscorrelation between the images indicate that both the spectral and the spatial characteristics have been preserved. The objective of this study is also to evaluate the spatial and spectral quality of a fused and hyperspectral image. The spatial quality of the fused image is measured by bias, UIQIs, entropy,

average gradient and ERGAS and spectral quality is measured by spectral discrepancy, correlation coefficient, RMSE, Spectral Angle Mapper, standard deviation and signal-to-noise ratio. Artificial neural network has been used to classify the target pixels of the new integrated high spatial and spectral image. The use of artificial neural networks (ANNs), particularly multiple-layer perceptrons, is proposed as a method to classify pixels in the fused image [18, 19].

2 Study Area

The Henry Island located in the Sundarbans bio-geographical province of the State of West Bengal, India, is selected for this study purpose (Fig. 1). This island extends between 21°35′49.00″N to 21°33′30.00″N latitude and 88°15′40.00″E to 88°18′09.00″E longitude which contains wide variety of mangrove species.

3 Methodology

3.1 Data Acquisition

The hyperspectral data has been acquired by the Hyperion sensor which was procured through Data Acquisition Request on the 27th May 2011. The IRS P6 LISS-IV

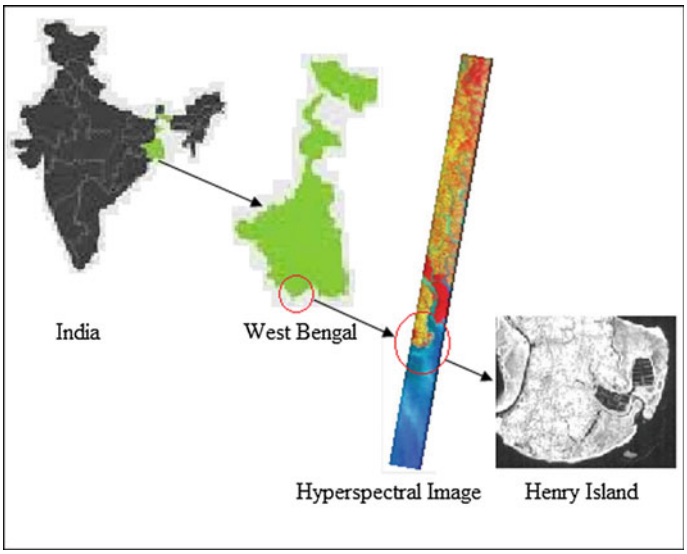


Fig. 1 Henry Island, Sundarbans, India

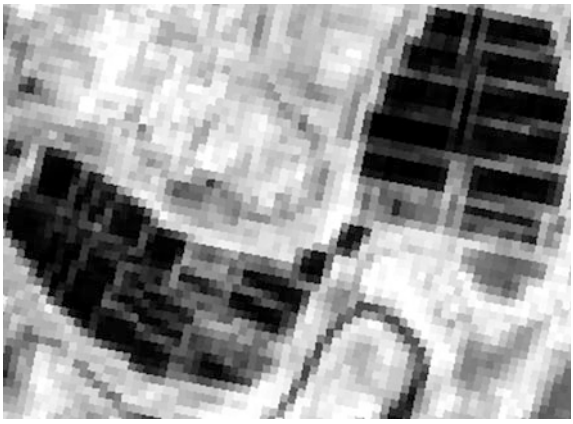


Fig. 2 Original hyperspectral image of study area

Fig. 3 Multispectral image of the study area

multispectral data has been procured from National Remote Sensing Centre (NRSC), ISRO. Figures 2 and 3 display the original hyperspectral image and LISS-IV multispectral image, respectively, that has been integrated in this study.

3.2 Pre-processing of Data

FLAASH is an atmospheric correction tool that needs to be procured as an add-on tool to ENVI software. ENVI (Environment for Visualizing Images) is a software that enables advanced spectral image processing and geospatial analysis with a modern, user-friendly interface. ENVI has the latest processing and analysis tools that help extract meaningful information from multispectral and hyperspectral imagery to arrive at better decisions. FLAASH corrects wavelengths along the visible, near-infrared and shortwave infrared regions of the electromagnetic

spectrum. This algorithm incorporates MODTRAN 4 energy transfer code. This model retrieves water vapour and aerosol with wavelength bands in proper positions. FLAASH also corrects images procured in either nadir or off nadir geometries. Factors such as central location of scene, type of sensor, date of acquisition, altitude of sensor, average elevation of study area, time of acquisition are required as input for processing of the pixel radiance data. For more precise correction, the 'tropical atmospheric model' and 'maritime aerosol model' present in ENVI software has been used for water vapour content extraction from the water absorption bands of hyperspectral data [20].

Geometric correction helps to locate the accurate position of a pixel in terms of geographical coordinates. The image-to-image registration procedure has been adopted in this study for geocorrection. This is a transformation technique by which two images of same geometry and same geographical area coincide over one another so that the corresponding objects of the two images appear at the same location [9]. A georeferenced hyperspectral image of Henry Island acquired by USGS has been used as base image [21].

3.3 Selection of Hyperspectral Bands

The hyperspectral image covers a wider range of the electromagnetic spectrum as compared to the multispectral image that covers only few broad wavelength regions of the spectrum. During the process of resampling, if the entire set of hyperspectral bands is mapped to the limited band set of multispectral data, an unwanted overlapping condition may arise leading to erroneous target detection results.

Thus, a proper band selection process must be applied to the hyperspectral image before the fusion model is applied on it. Out of the total hyperspectral band set, only those bands covering the spectral range of multispectral data are selected and the data set is spectrally resized.

3.4 Endmember Detection

Endmember detection helps to identify the set of unique spectral signatures that comprise the mixed pixels [22]. The image obtained after pre-processing and band selection is used for detection of target endmembers. This study applies the NFINDR algorithm to detect pure endmember pixels from the hyperspectral data. NFINDR attempts to estimate endmembers by finding the set of pixels with the largest possible volume by 'inflating' a simplex inside the data [23]. The procedure begins with a random set of vectors. In order to refine the estimate of the endmember, every pixel in the image is evaluated as to its likelihood of being a pure or

nearly pure pixel. To do this, the volume is calculated with each pixel in place of each endmember. A trial volume is calculated for every pixel in each endmember position by replacing that endmember and finding the volume. If the replacement results in an increase in volume, the pixel replaces the endmember. This procedure is repeated until there are no more replacements of endmembers.

3.5 Resampling of Hyperspectral Bands to Multispectral Bands

The bands of hyperspectral image are selected and resampled to the broad wavelengths of the multispectral image bands and a resampled response of spectra is obtained and given by $S(\omega_i)$:

$$S(\omega_i) = \frac{\int s(\omega) R_i(\omega) \mathrm{d}\omega}{\int R_i(\omega) \mathrm{d}\omega} \tag{1}$$

$s(\omega)$ is the response of spectra acquired by hyperspectral sensor at wavelength ω, $R_i(\omega)$ is the response of spectra of ith band of the image acquired by multispectral sensor. 'ω' is the wavelength of the hyperspectral image band and 'ω_i' is the wavelength of multispectral image band. The spectral response function $R_i(\omega)$ is assumed to be a Gaussian distribution for the multispectral sensor and is modelled as

$$R_i(\omega) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{\frac{-(\omega-\mu_i)^2}{2\sigma_i^2}} \tag{2}$$

where 'ω' represents continuous spectra of image acquired by hyperspectral sensor, and 'μ_i' and 'σ' represent central wavelength and standard deviation, respectively, for the ith band of the image acquired by multispectral sensor. We estimate $R_i(\omega)$ for every ith band of the multispectral sensor.

3.6 Linear Spectral Unmixing (LSU)

The model follows the principle that if the entire area is separated proportionally on the basis of fractional abundance estimates of comprising endmembers, the reflected light bears the properties of these endmembers in similar proportions [24, 25]. The relationship envisaged between the fractional abundance of entities comprising image and the spectrum of reflected radiation is linear.

The spectral signature of an image pixel r is modelled as

$$r = M\alpha + n \tag{3}$$

where n is noise or can be interpreted as model error.

LSU is an unconstrained model that calculates the fractional abundance values ($\alpha 1$, $\alpha 2$, $\alpha 3$,…, αp) of the endmembers in the pixel vector 'r'. In this study, the resampled spectrum is subjected to LSU for unmixing the mixed pixels and calculating their fractional abundances. Initially, LSU is applied on the multispectral image with NFINDR extracted hyperspectral endmembers as input. The fractional abundance values thus generated represent the endmember distribution of the extracted endmembers in the multispectral image [26–29].

3.7 Integration of Hyperspectral and Multispectral Data

We have a hyperspectral image scene h with p number of bands and a multispectral scene m with q number of bands. The output of linear spectral unmixing from the above subsection is an abundance matrix α. These abundance values generated from the multispectral image data are again fed into the LSU model along with the NFINDR extracted endmember signatures. This generates a fused pixel containing integrated information of endmembers and their distribution in a spatially enhanced image. The fusion model is written as

$$r_f = M\alpha \tag{4}$$

where M is the $k \times p$ resampled endmember matrix and k in the number of endmembers. M is multiplied by the abundance matrix generated by the multispectral image, and we get a fused pixel r_f with p bands [30].

3.8 Classification with Artificial Neural Networks

An artificial neural network is a collection of conceptually interconnected neurons simulating a human brain at work. Each neuron receives input from the several *upstream* neurons in the network, generates a scalar output and sends it *downstream* to another group of neurons. A weight is associated with each input and the output is then a function of the weighted sum of inputs in the form of discrete or continuous function depending on the variety of network in use. Multilayer perceptrons (MLP) are neural networks of the general class known as feedforward networks. MLP has at least three layers of neurons, the input layer, the hidden layer and the output layer. There can be more than one hidden layer in multilayer feedforward neural networks but only one input and output layers. The number of neurons in the

input layer is specified by the type and amount of data which will be available for input neural network. The number of output neurons corresponds to the type of answer of the network.

4 Quality Assessment Criteria

The evaluation of quality of the fused image is based on quantitative criteria of spatial and spectral properties defining image [7]. In this study, ten evaluation criteria have been used for quality assessment of the fused and hyperspectral image. Spectral discrepancy, correlation coefficient, RMSE, Spectral Angle Mapper and standard deviation are used for spectral quality evolution whereas bias, UIQI, entropy, average gradient and ERGAS are used for spatial quality assessment.

4.1 Spectral Quality Assessment

The integrated image has to preserve the spectral content of hyperspectral image. The evaluation techniques are applied on the integrated and hyperspectral image of $M \times N \times B$ dimensions where M and s are the numbers of pixels along x-axis and y-axis, and B is the number of bands with same spectral wavelength in fused (F) and hyperspectral (H) images.

4.1.1 Spectral Discrepancy

The spectral quality of an integrated image with $M \times N$ dimensions is analysed by D_k for every band [7].

$$D_k = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |F_k(x, y) - H_k(x, y)|$$

where $F_k(x, y)$ and $H_k(x, y)$ are the pixel values of the integrated and hyperspectral images at location (x, y) at kth band [7].

4.1.2 Correlation Coefficient (CC)

An integrated image of good quality has high correlation with the original image. High CC signifies similarity between fused and original image. The ideal value of the CC is 1 and the lowest correlation shows a zero value [3, 7, 9, 12, 31].

$$\mathrm{CC}(H,F) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(H_{i,j}-\bar{H})(F_{i,j}-\bar{F})}{\sqrt{\left(\sum_{i=1}^{M}\sum_{j=1}^{N}(H_{i,j}-\bar{H})^2\right)\left(\sum_{i=1}^{M}\sum_{j=1}^{N}(F_{i,j}-\bar{F})^2\right)}}$$

where $\bar{H}$ and $\bar{F}$ stand for means of the corresponding data set and CC is estimated for the entire data set [7, 12].

4.1.3 Normalized Root Mean Square Error

Normalized RMSE is calculated between the fused and hyperspectral images. The RMSE value should be as low as possible [3, 7, 9].

$$\mathrm{NRMSE}_k = \sqrt{\frac{1}{M\times N\times B^2}\sum_{i}^{M}\sum_{j}^{N}(F_k(i,j)-H_k(i,j))^2}$$

NRMSE is calculated between (i, j)th pixel of fused and hyperspectral images of the kth band.

4.1.4 Spectral Angle Mapper (SAM)

This indicator estimates the spectral angle between fused and hyperspectral images on same coordinates as a spectral vector. SAM shows the direction of the vectors. SAM with a value of zero signifies the absence of spectral distortion [3].

$$\mathrm{SAM} = \arccos\left\{\frac{\sum_{i=1}^{P}H_i * F_i}{\sqrt{\sum_{i=1}^{P}(H_i)^2\sum_{i=1}^{P}(F_i)^2}}\right\}$$

where P stands for number of pixel per band.

4.1.5 Standard Deviation (SDD)

The spectral closeness between hyperspectral and fused images is shown by this indicator. The value of SDD between hyperspectral and fused images should be close to zero [3, 9].

$$\mathrm{SDD} = \sqrt{\frac{1}{M*N}\sum_{i=1}^{M}\sum_{j=1}^{N}(P_{i,j}-\mu)^2}$$

where $P_{i,j}$ is pixel brightness value at coordinate (i, j) and μ is the mean value of the pixels.

4.1.6 Signal-to-Noise Ratio (SNR)

SNR analyses the spectral properties of hyperspectral image in the fused image [9].

$$\mathrm{SNR}_k = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N} F_k(i,j)^2}{\sum_{i=1}^{M}\sum_{j=1}^{N} \left(F_k(i,j) - H_k(i,j)\right)^2}}$$

where k is the number of bands of fused and hyperspectral image.

4.2 Spatial Quality Assessment

The enhancement of spatial resolution of imagery depends upon the increase in spatial information after integration of multispectral and hyperspectral data [7].

4.2.1 Bias

Bias is the difference between mean of the original hyperspectral and of fused image. A good-quality fused image shows a value of zero for bias [3].

$$\mathrm{Bias} = \bar{H}_{\mathrm{Hyper}} - \bar{F}_{\mathrm{Fused}}$$

where $\bar{H}$ and $\bar{F}$ are the mean values of pixels in hyperspectral and fused images.

4.2.2 Edge Detection

The edge detection of multispectral fused and hyperspectral images is performed by the Canny operator [7].

5 Results and Discussion

Image Fusion

The acquired hyperspectral data of Henry Island has been atmospherically and geometrically corrected for pre-processing. Automated endmember detection algorithm, NFINDR, has been applied on the image that has successfully extracted

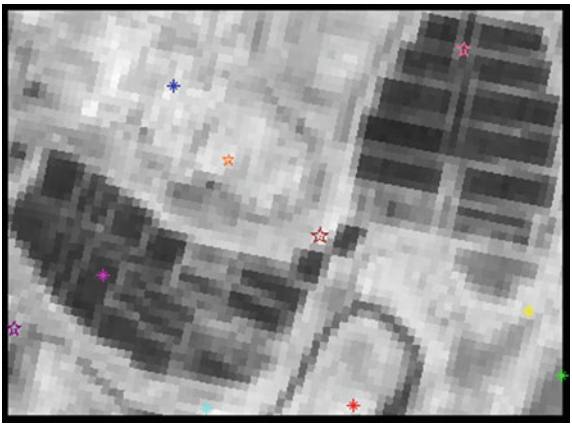


Fig. 4 Location of endmembers obtained from NFINDR algorithm

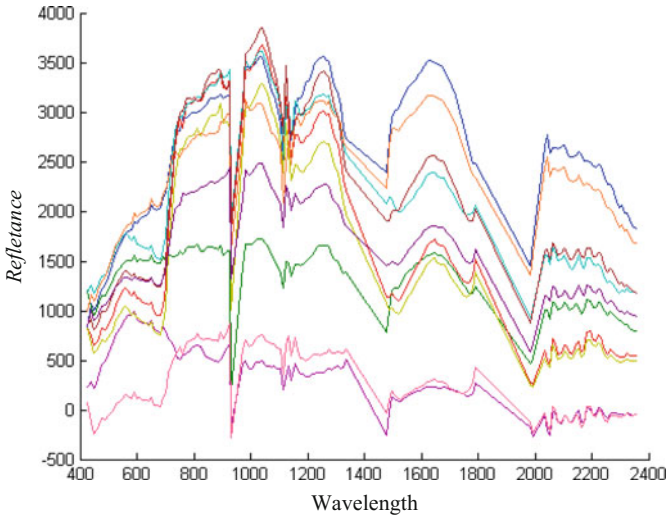


Fig. 5 Spectral profile of endmembers extracted from NFINDR considering entire hyperspectral band range

10 dominant endmembers from the image data. The endmembers mostly represent different mangrove species, soil types, aquaculture and sea water. Figures 4 and 5 display the location of endmembers extracted and their spectral signatures, respectively.

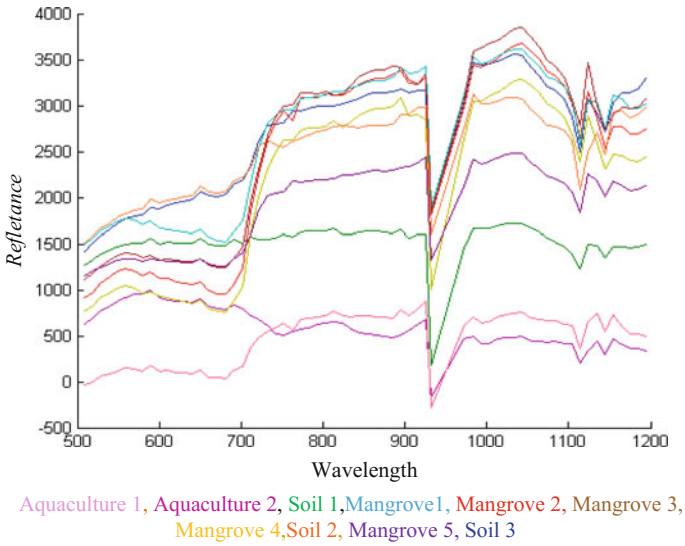


Fig. 6 Spectral profile of endmembers in resized hyperspectral image with bands ranging from 500 to 1200 nm

The original 242 bands of the hyperspectral image covering the visible, near-infrared and shortwave infrared regions of the electromagnetic spectrum has been resized to 66 bands comprising bands of broad spectral range of the multispectral LISS-IV image. The first 10 bands are from the green wavelength region (ranging from 500 to 600 nm), next 10 bands are from the red wavelength region (ranging from 601 to 702 nm) and remaining 46 bands are from near-infrared region (ranging from 703 nm to 1200 nm). This spectrally resized hyperspectral image is used for further processing. NFINDR has also been applied on the resized hyperspectral image and the spectral profile is displayed in Fig. 6.

The endmember matrix dimensions generated from NFINDR application is 66×10 which has been subjected to spectral resampling. This step resamples the 66 hyperspectral bands to 3 multispectral bands of the LISS-IV image. The spectral response function (SRF) is calculated for green, red and near-infrared bands individually and has been plotted in graphs to ensure Gaussian distribution (Figs. 7, 8 and 9).

A resampled spectral response matrix of dimension 3×10 is generated that it represents the extracted 10 endmembers in 3 resampled bands of the multispectral image. The resampled multispectral bands of endmembers have been taken as input to perform linear spectral unmixing on the LISS-IV multispectral image data. This model calculates the abundance estimates of the 10 endmembers present within each pixel of the multispectral image. The abundance distribution of endmembers present within each pixel makes the image more spatially enhanced. The fusion model takes the generated abundance data and the signature spectra of the 10

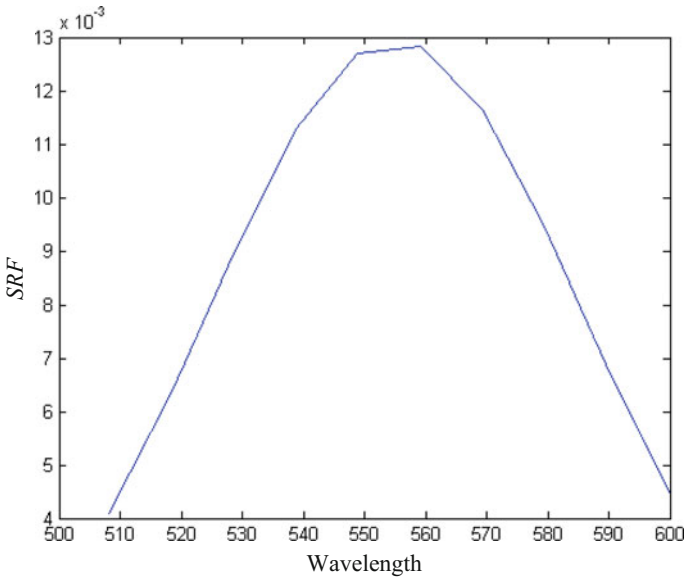


Fig. 7 Spectral response for green band [wavelength vs. $R_i(\omega)$]

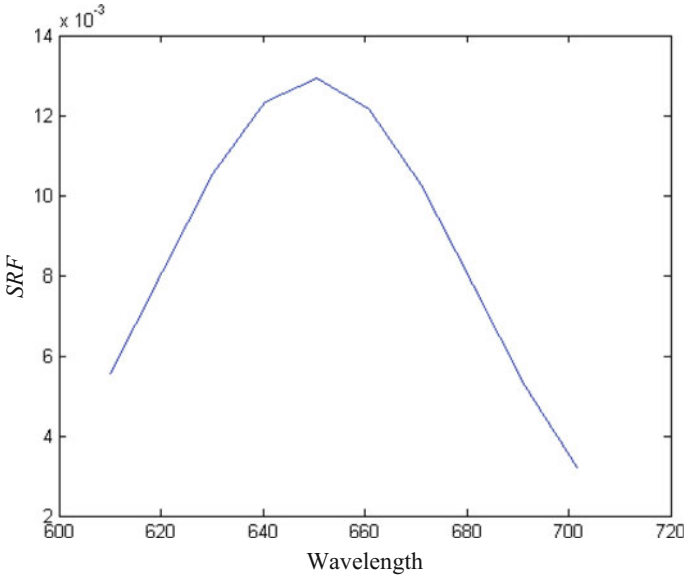


Fig. 8 Spectral response for red band [(wave length vs. $R_i(\omega)$]

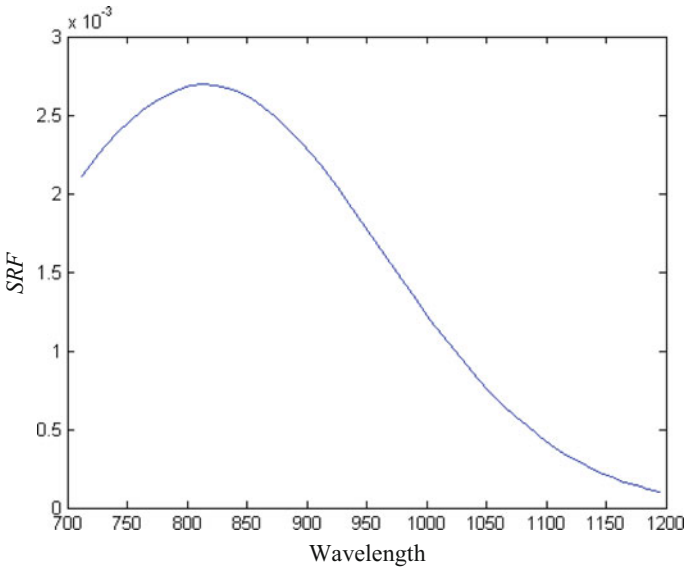


Fig. 9 Spectral response for NIR band [wave length vs. $R_i(\omega)$]

Fig. 10 Fused image at 529 nm wavelength

endmember as input and performs linear spectral unmixing on the Hyperspectral data. The result generates a new image with fused pixels, each having the spatial resolution of the multispectral data and spectral details of the hyperspectral data (Fig. 10). The spectral signatures of endmembers in the fused image are displayed

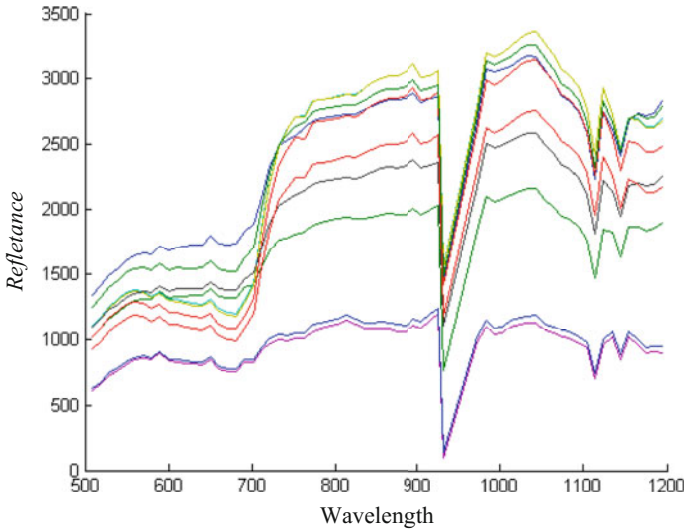


Fig. 11 Spectral profile of the 10 endmembers in the fused image

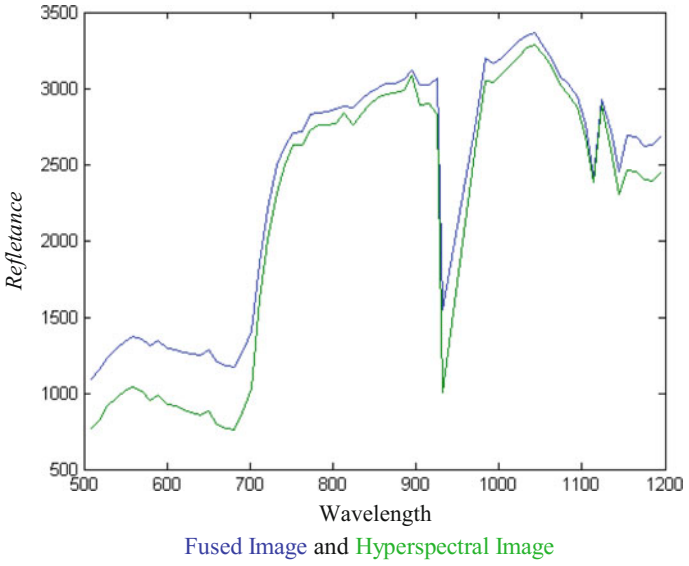


Fig. 12 Spectral profile of mangrove species pixels of fused and hyperspectral image

in Fig. 11. Figures 12, 13 and 14 display the spectral profile of mangrove species pixels, beach soil and aquaculture in the fused as well as hyperspectral images.

Figure 15 displays the neural network parameters used for classification of pixels. The training of the neural network was repeated 1000 times using gradient

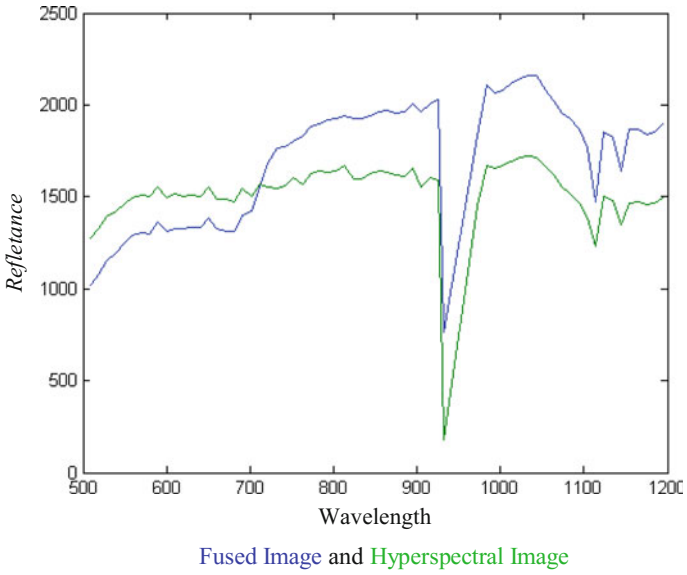


Fig. 13 Spectral profile of beach (soil) pixels of fused and hyperspectral image

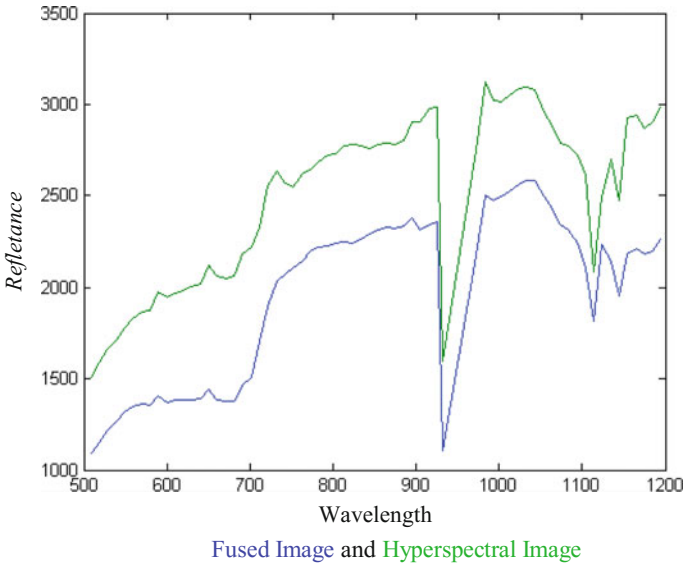


Fig. 14 Spectral profile of aquaculture pixels of fused and hyperspectral image

descent method. The number of hidden layers is considered as 1, and the minimum output activation threshold has been kept at 0.9. Figure 16 displays the plot generated during the training process of the pixels. Figure 17 displays the final classified output after executing the neural net classifier.

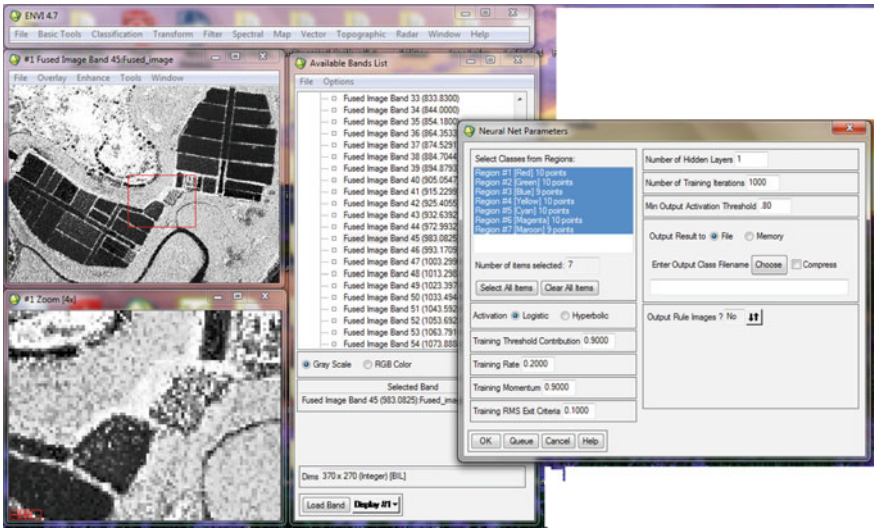


Fig. 15 Input classes for neural net classification

5.1 Spectral Quality Assessment

The indices designed for evaluation of spectral quality of fused and hyperspectral image is shown with their obtained results.

5.1.1 Spectral Discrepancy

The variation of spectral discrepancy with respect to wavelength is shown in Fig. 18, which shows satisfactory results. It is observed that except for a few bands, most of the spectral discrepancy values are below 0.02.

5.1.2 Correlation Coefficient (CC)

The spectral profile of endmembers plotted in the fused image as well as the original hyperspectral image shows high normalized crosscorrelation amongst all bands (Fig. 19). Normalized crosscorrelation values nearer to 1 signify that pixels in the bands of both images are ideally correlated and values towards 0 signify low correlation or no correlation at all. Ideally, the normalized crosscorrelation values should be more than 0.5. In our study, maximum numbers of bands have CC value above 0.9, signifying highly correlated pixels of integrated and hyperspectral images. This signifies that the spectral signature of the endmembers has been well preserved after image fusion. The near-infrared part of the bands shows a

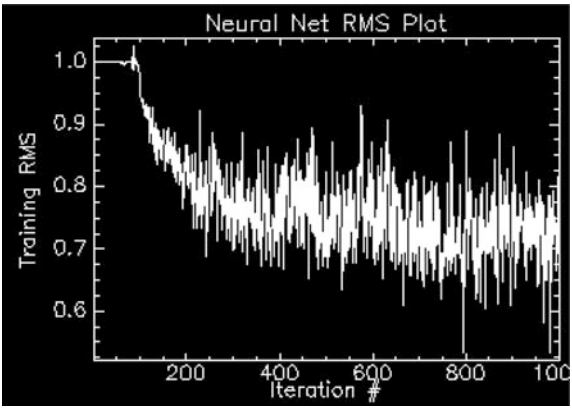
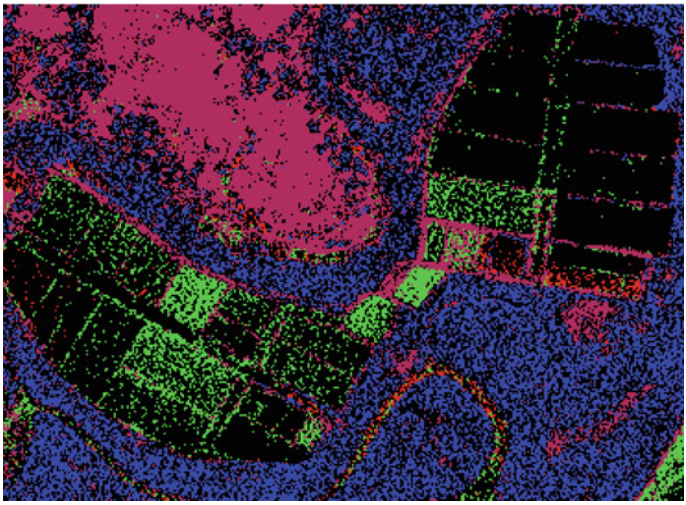


Fig. 16 Plot of training of image pixels during Neural Network Classification



Agricultural Soil 1, Beach Soil, Vegetation Type 1,
Vegetation Type 2, Waterbody, Vegetation, Agriculture Soil 2

Fig. 17 Classified Output after Neural Network Classification

crosscorrelation of more than 90% between them. A comparison of the spectral profile of mangrove species in both images shows good crosscorrelation amongst them signifying good restoration of vegetation bands (Fig. 12). However, beach (soil) and water pixels of both images show a lower correlation amongst themselves (Figs. 13 and 14). In case of low values of crosscorrelation, the primary absorption valleys are well maintained and are separable.

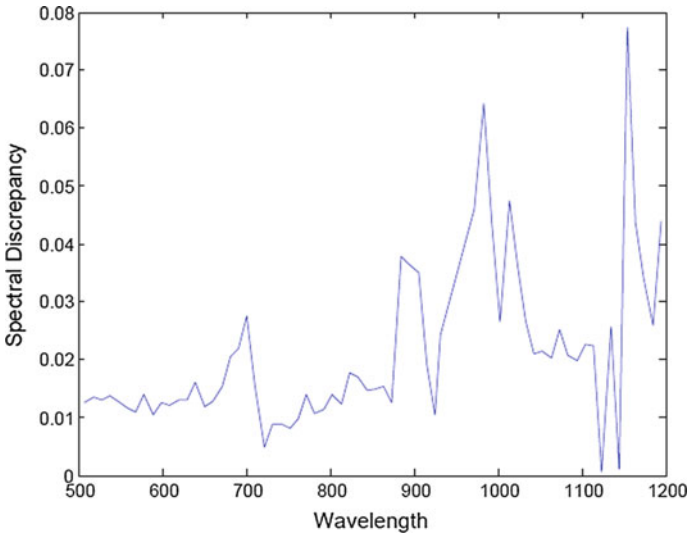


Fig. 18 Spectral discrepancy

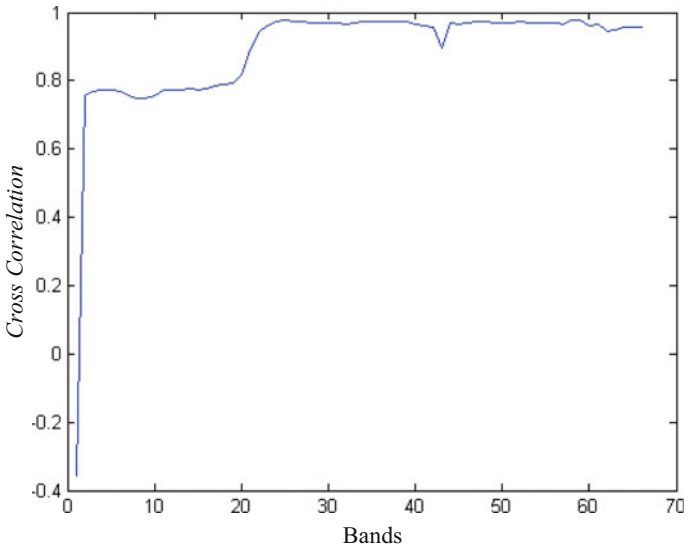


Fig. 19 Normalized crosscorrelation values between integrated and hyperspectral pixel spectra (for 66 bands)

5.1.3 Normalized Root Mean Square Error

The normalized RMSE is calculated for every band of fused and hyperspectral images and displayed in Fig. 20. It is observed that except for a few band, most of the bands show a RMSE value below 0.01 which is satisfactory.

5.1.4 Spectral Angle Mapper (SAM)

The extracted SAM values from fused and hyperspectral images for selected pixel in every band are shown in Fig. 21. It is observed that the SAM value is low in the NIR part of the electromagnetic spectrum. However, the visible region shows higher values signifying lower quality fusion in this part of the spectrum.

5.1.5 Standard Deviation

The obtained values of standard deviation between the pixels of hyperspectral and fused images are shown in Fig. 22 which show good results in NIR region. The deviation is close to zero in most of the bands signifying good-quality fused data.

5.1.6 Signal-to-Noise Ratio

The value of SNR calculated between fused and hyperspectral images is observed to vary with spectral wavelength and is shown in Fig. 23.

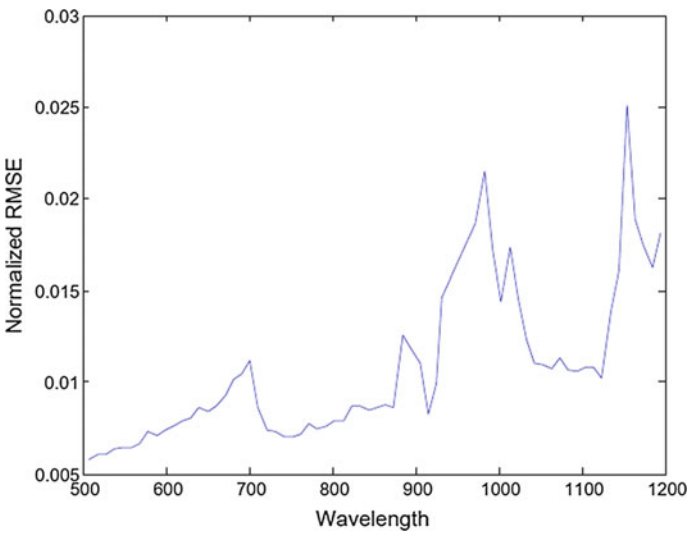


Fig. 20 Normalized RMSE

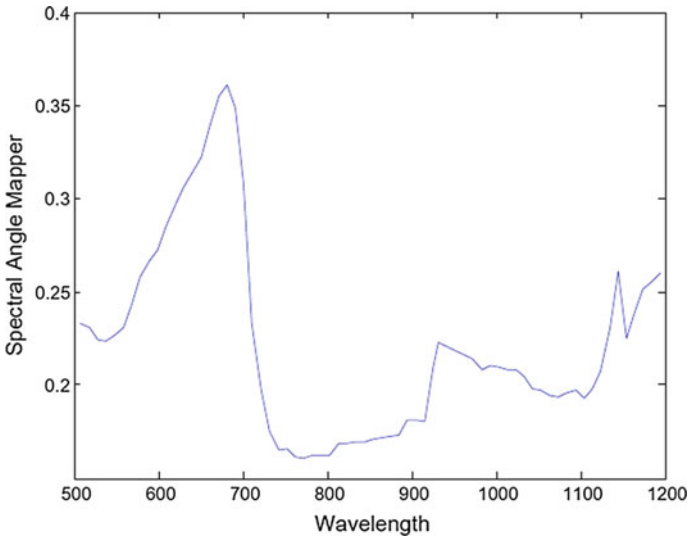


Fig. 21 Spectral angle mapper

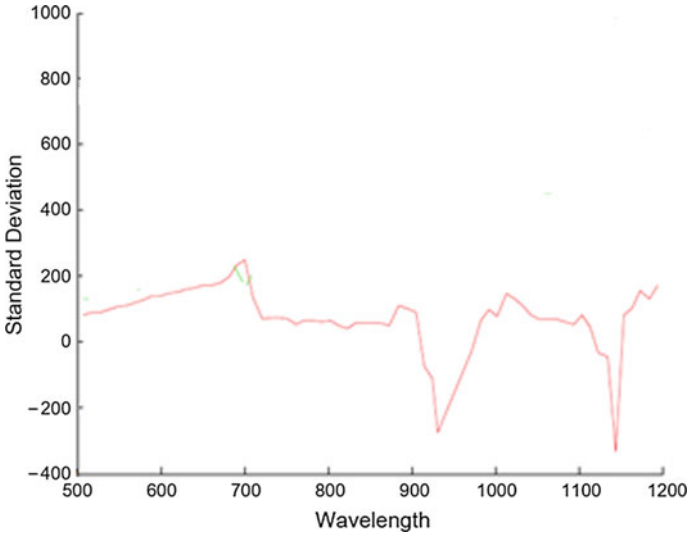


Fig. 22 Standard deviation

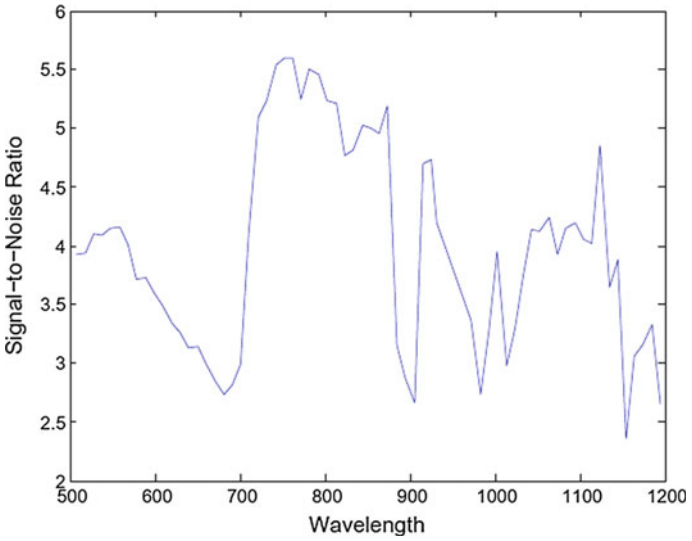


Fig. 23 Signal-to-noise ratio

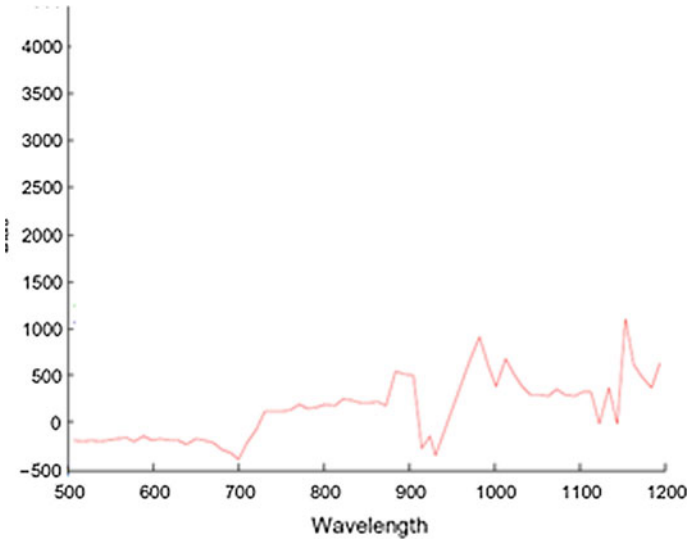


Fig. 24 Spatial quality assessment by bias

5.2 Spatial Quality Assessment

The designed indices for evaluating the spatial quality of fused image are shown with their obtained results.

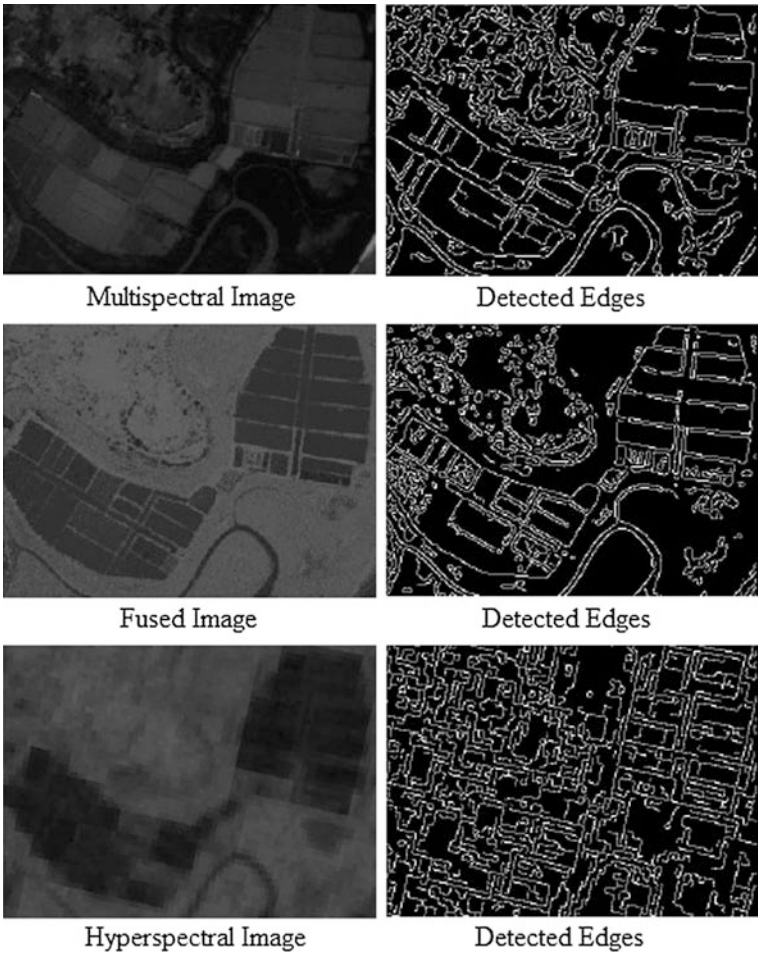


Fig. 25 Spatial quality assessment by edge detection

5.2.1 Bias

The obtained results of bias show a very good-quality fused and hyperspectral image that is displayed in Fig. 24.

5.2.2 Edge Detection

This index detects the edges in multispectral, fused and hyperspectral images, respectively, by Canny operator (Fig. 25). On visual inspection, it is observed that the obtained result of fused image is close to the result of multispectral image.

6 Conclusion

This is an attempt to obtain an image of high resolution, both spatially and spectrally by employing the available hyperspectral and multispectral images without involving any extra sensor technology. So the proposed methodology is economic and easy to use. Here, the unique signature of spectra of hyperspectral image is preserved and spatial details enhanced to that of the multispectral image. As a result, a fused image of 66 bands and 5.8 m spatial resolution has been restored by the fusion of a hyperspectral image of 30 m resolution and a multispectral image of 5.8 m resolution. This algorithm leads to a high value of correlation between the integrated and hyperspectral images which shows that the spectral and spatial signatures of the original images have been preserved successfully. This chapter has also applied indices for evaluating the spectral and spatial qualities of fused and hyperspectral image. The results signify that the objective for integrating hyperspectral and multispectral images has been successfully attained. Artificial neural network has been used to classify the pixels of the integrated image.

References

1. Krista A et al (2007) Wavelet based image fusion technique—an introduction, review and comparison. ISPRS J Photogram Remote Sens 249–263
2. Elaksher A (2008) Fusion of hyperspectral images and lidar-based dems for coastal mapping. In: Remote sensing and Spatial information sciences, Vol. XXXVII Part B3b, China
3. Chetin M, Masaoglu N (2009) Merging hyperspectral and panchromatic image data: quality and quantity analysis. Int J Remote Sens 30(7):1779–1804. doi:10.1080/01431160802639525
4. Naidu VPS, Raol JR (2008) Pixel-level image fusion using wavelets and principal component analysis. Defence Sci J 58(3):338–352
5. Klonus S, Ehlers M (2009) Performance of evaluation methods in image fusion. In: 12th international conference on information fusion seattle, WA, USA
6. Maurer T (2013) How to pan-sharpen images using the Gram-Schmidt Pan-Sharpening method. In: International archives of the photogrammetry, remote sensing and spatial information sciences, vol. XL-1/W1, ISPRS Hannover Workshop, Hannover, Germany
7. Yakhdani MF, Azizi A (2010) Quality assessment of image fusion techniques for multisensor high resolution satellite images, vol XXXVIII, ISPRS, Part 7B, Austria (Case study: IRS-P5 and IRS-P6 Satellite images)
8. Al-Wassai F, Kalyankar N, Zuky A (2011) The IHS transformations based image fusion. Comput Vision Pattern Recognit (cs.CV)
9. Al-Wassai FA, Kalyankar NV, Al-Zaky AA (2011) The statistical methods of pixel-based image fusion techniques. Int J Artifi Intell Knowl Disc 1(3)
10. Blasch E, Li X, Chen G, Li W (2008) Image quality assessment for performance evaluation of image fusion. doi:10.1109/ICIF.2008.4632263
11. Riyahi R, Kleinn C, Fuchs H (2009) Comparision of different image fusion techniques for individual tree crown identification using quichbird images. In proceeding of ISPRS Hannover Workshop 2009
12. Han Z, Tang X, Gao X, Hu F (2013) Image fusion and image assessment of fused images. Remote Sens Spat Inf Sci vol XL-7/W1, 3rd ISPRS IWIDF, China

13. Delalieux S, Zarco-Tejada PJ, Tits L,.Bello MAJ, Intrigliolo DS, Somers B (2014) Unmixing-based fusion of hyperspatial and hyperspectral airborne imagery for early detection of vegetation stress. IEEE J Appl Earth Obs Remote Sens 7(6)
14. Khan MM et al (2009) Pan sharpening of hyperspectral image using spatial distortion optimisation. In: International conference on image processing, Cario, pp 2853–2856
15. Licciardi GA, Khan MM, Chanussot J, Montanvert A, Condat L, Jutten C (2012) Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear PCA BAND reduction. EURASIP J Adv Signal Process
16. Shaw GA, Hsiao-hua, Burke K (2003) Spectral imaging for remote sensing. Lincon Lab J 14 (1)
17. Villa A, Chanussot J, Benediktsson JA, Ulfarsson M, Jutten C (2010) Super-resolution: an efficient method to improve spatial resolution of hyperspectral images. IGARS
18. Richards JA (1999) Remote sensing digital image analysis. Springer, Berlin, p 240
19. Rumelhart DE, Hinton GE, Williams RJ (1987) Learning internal representation by error propagation. In: Rumelhart D, Mc Clelland J (eds) Parallel distributed processing, vol 1. MIT Press, Cambridge
20. Chakravorty S, Chakrabarti S (2011) Pre-processing of hyperspectral data: a case study of Henry and Lothian Islands in Sunderban Region, West Bengal, India. Int J Geomatics Geosci 2(2), ISSN 0976–4380
21. Jensen (1996) Introductory digital image processing : a remote sensing perspective. 2nd edn
22. Plaza A, Martínez P, Pérez R, Plaza J (2004) A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. IEEE Trans Geosci Remote Sens 42(3):650–663
23. Winter ME (1999) N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data. In: Proceedings SPIE 3753, Imaging Spectrometry V, pp 266-277
24. KeshavaN, Mustard J (2002) Spectral unmixing. IEEE Signal Process Mag 19(1):44–57, USA
25. Khandelwal A, Rajan KS (2011) Hyperspectral image enhancement based on sensor simulation and vector decompostion. ISBN no: 978-1-4577-0267-9
26. Chakravortty S (2013) Analysis of endmember detection and subpixel classification algorithms onhyperspectral imagery for tropical mangrovespecies discrimination in theSunderbans Delta, India. J Appl Remote Sens 7(1):073523. doi:10.1117/1.JRS.7.073523
27. Dimitris M, Siracusa C, Shaw G (2001) Hyperspectral subpixel target detection using the linear mixing model. IEEE Trans Geosci Remote Sens 39(7)
28. Duran O, Petrous M (2004) Mixed pixel classification in remote sensing—literature survey. University of Surrey, Guildford
29. Eismann M, Hardie R (2005) Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions. IEEE Trans Geosci Remote Sens 43 (1):455–465
30. Bieniarz J, Cerra D, Avbelj J, Reinartz P, Muller R (2011) Hyperspectral image resolution enhancement based on spectral unmixing and information fusion. In: Proceedings of ISPRS XXXVIII-4-W19-33-2011 Hannover workshop on high-resolution earth imaging for geospatial information
31. Briechle K, Uwe Hanebeck D (2001) Template matching using fast normalised cross correlation. In: Proceeding of SPIE, aero-sense symposium, vol.4387. Orlando, Florida

Deep Learning Techniques for Breast Cancer Detection Using Medical Image Analysis

D. Selvathi and A. Aarthy Poornila

Abstract Breast cancer has the second highest mortality rate in women next to lung cancer. As per clinical statistics, 1 in every 8 women is diagnosed with breast cancer in their lifetime. However, periodic clinical checkups and self-tests help in early detection and thereby significantly increase the chances of survival. Invasive detection techniques cause rupture of the tumor, accelerating the spread of cancer to adjoining areas. Hence, there arises the need for a more robust, fast, accurate, and efficient noninvasive cancer detection system. In this work, an automated system is proposed for achieving error-free detection of breast cancer using mammogram. In this system, the deep learning techniques such as convolutional neural network, sparse autoencoder, and stacked sparse autoencoder are used. The performance of these techniques is analyzed and compared with the existing methods. From the analysis, it is observed that the stacked sparse autoencoder performs better compared to other methods.

Keywords Breast cancer · Mammogram · Deep learning techniques

1 Introduction

Cancer is one of the deadliest diseases faced by mankind. The chance of women dying from early-stage breast cancer is estimated to be about 3%. Statistics show that breast cancer constitutes about one-fourth of all the cancers affecting women, commonly affecting women above 40 years of age. Recent statistics show that there has been a steady decrease in the annual deaths from breast cancer among women, from 32.69 in 1991 to 24.00 in 2005 (per 100,000 populations). This decrease in cancer death rates reflects the improvements in early detection and treatment for

D. Selvathi (✉) · A. Aarthy Poornila
Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India
e-mail: dselvathi@mepcoeng.ac.in

A. Aarthy Poornila
e-mail: aarthypoornila@gmail.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_8

breast cancer. Research is being done for breast cancer prevention, detection, and treatment, in order to increase the survival rate of patients all over the world.

Medical imaging for breast cancer can be used as a noninvasive method for looking inside the body and assisting the doctors in diagnosis and treatment. Islam et al. [1] have explored different breast cancer imaging techniques such as mammography, magnetic resonance imaging (MRI), and ultrasound for breast cancer diagnosis and compared their effectiveness, advantages, and disadvantages for detecting early-stage breast cancer. In [2], Sachin Prasad Na and Dana Houserkovaa evaluated the role of various modalities (mammogram, molecular breast imaging, ultrasound, MRI, PET, thermography, galactography, CT scan) used in breast cancer screening and diagnosis. Though initial detection of breast cancer can be done using any one of the available imaging modalities, malignancy cannot be confirmed by such images alone. So the treatment does not start until microscopic examination of tissue from the tumor is done to confirm its malignancy. Shyamala et al. [3] say that when such procedures are done, there is a high risk of tumor cells being seeded into the interstitial tissue fluid or veins. There is also a risk of dragging cells along the surgical incision or needle track leading to the possibility of increasing the spread of cancer through biopsy.

The field of medical image processing gains its importance in the need of accurate and efficient diagnosis of diseases over a short period of time. Since manual process is tedious, time-consuming, and impractical for large data, a need for automatic processing arises which can revolutionize modern medicine. In breast cancer detection using mammogram, the mammographic breast image is usually preprocessed to remove the pectoral muscle, as it biases the detection process. Thus, by removing the background and pectoral muscle regions from the mammogram, the search for abnormalities can be limited to the breast profile region alone. Jawad Nagi et al. proposed a technique for automated segmentation of mammogram, which uses morphological preprocessing and seeded region growing (SRG) algorithm [4] in order to remove the pectoral muscle region from the breast profile region in the mammogram as the presence of such regions can mislead the abnormality detection.

Detection of cancer tissues with higher pixel intensity than the other breast region is done easily. In case of dense breasts, the normal tissues have intensities similar to the tumor regions, and in such cases, it is necessary to identify tumor regions successfully. Classification of the tumor tissues in the breast into benign and malignant is a tedious work. Feature extraction is an important step in mammogram analysis. In conventional methods, handcrafted features are used to represent the content of images. The neural network emerged as an alternative approach to extract the best features automatically. Maxine Tan et al. have predicted the risk of a patient being affected by breast cancer in the near future, based on the analysis of feature variations (density features, structural features, and texture features) in a series of mammographic images [5]. Singh and Gupta [6] proposed a two-step breast cancer detection method. First, the malignant area is found using filtering and thresholding operations. Then, morphological operations are done to find the region boundary in the tumor patch. In [7], Mellisa Pratiwi et al. used gray-level co-occurrence matrix (GLCM)-based texture features and radial basis

function neural network (RBFNN) for breast cancer classification using mammogram. For comparative analysis, backpropagation neural network (BPNN) is used in this work. It is reported that RBFNN gives 94.29% benign and malignant classification accuracy which is 2% greater than BPNN. Kim et al. [8] classified spiculated malignant lesions and normal tissues using support vector machine (SVM) which used a set of region-based stellate features, achieving 97.9% classification performance on mini-MIAS database. In [9], Hussain et al. employed successive enhancement learning-based weighted support vector machine (SELwSVM) to classify masses and normal dense tissues with 96% accurate recognition rate, by exploiting structural properties represented by the extracted texture features.

Deep learning is a newly emerging area which exploits artificial intelligence and machine learning to learn features directly from the data, using multiple nonlinear processing layers. Deep learning models can achieve very high accuracy in image classification, on par with human-level performance. Li Deng and Dong Yu discussed various deep learning techniques, their classification into supervised, unsupervised, and hybrid deep networks depending on their architectures and applications such as computer vision, language modeling, text processing, multimodal learning, and information retrieval [10]. Schmidhuber, in [11], extensively explained about the credit assignment problem and classification of shallow and deep learners by the depth of their credit assignment paths. Comparative analysis of acyclic (Feedforward) and cyclic (Recurrent) neural networks showed that RNNs are the deepest and are more powerful than FNNs. This paper also has a detailed discussion on supervised, unsupervised, and reinforced learning. Martin Langkvist et al. explored the time-series problems in deep learning and unsupervised feature learning. İn this work, a solution of using temporal coherence or temporal pooling or hidden unit activation is suggested for high-dimensional, multivariate data [12]. Furthermore, a comparative study of various time-series problems such as stock prediction, speech recognition, motion capture, and psychological data has been done.

Guo et al. [13] reviewed 216 research papers describing about deep learning algorithms and categorized them into convolutional neural network (CNN), autoencoders, restricted boltzmann machine (RBM), and sparse coding. Deeper analysis of the applications such as computer vision shows that such algorithms outperform accuracy of human raters. Arevalo et al. in [14] used CNN to learn features from mammograms, which are subsequently fed into a classifier. CNN architecture takes the raw pixels of the image as input, to learn a set of nonlinear transformations that represent better image data. This also outperforms a set of handcrafted features. Ketan Sharma and B. Preet used convolutional neural network classifier in [15] to boost the classification performance in breast cancer detection. CNN performs better than the other classifiers in terms of accuracy and misclassification rate of mammograms. Autoencoders are used in learning features in an unsupervised fashion in deep networks. Baldi in [16] gave detailed explanations about the types and the complexity of both linear and nonlinear deep networks and also elucidated about Hebbian learning, clustering, and information theory.

Mammographic risk scoring is automated by computing breast cancer risk from the extracted mammographic features in [17] by Kersten Petersen et al. In this work, the features are learned using a convolutional sparse autoencoder (CSAE) which uses sparse autoencoders in a convolutional architecture. Classification is performed by a softmax classifier based on these extracted features. Both density segmentation and texture scoring can be done in this manner. Jun Xu et al. in [18] used SSAE in breast cancer detection using histopathological images. High-level features can be learned from images by stacking sparse autoencoders together. These features are extracted patchwise and are fed into a classifier for classification. This work inspects various deep learning techniques and justifies that the performance of SSAE is superior compared to all other techniques.

Mammography, which is the present "Gold Standard" for breast imaging, is widely used for screening and diagnosing breast cancer in women. The radiation exposure in case of mammogram imaging is very minimal and also cheaper compared to all other techniques. There are many publicly available mammogram databases. The mini-MIAS database is one such public database that is used in this work, where an attempt has been made to explore the human intelligence-based deep learning techniques for achieving error-free detection of breast cancer.

2 Proposed System

The proposed system uses human intelligence-based deep learning techniques for detecting breast cancer using mammogram imaging modality. The various steps involved in the proposed system are shown in Fig. 1.

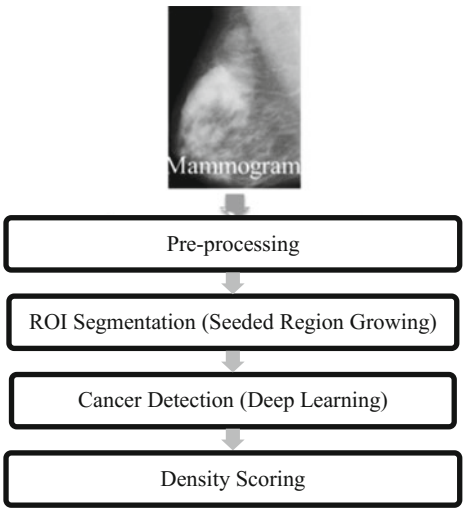


Fig. 1 Flow chart of proposed system

2.1 Input Imaging Modalities

The selection of a specific modality among all the medical imaging modalities available to study the organ present inside the human body is important. This will provide a wealth of information to understand the nature of diseases. The proposed system uses X-ray images (mammogram) which generate the information about the anatomical structure of a lesion.

2.2 Preprocessing

Image preprocessing enhances some important image features for further processing and suppresses unwanted distortions. Figure 2 shows the various stages of preprocessing method.

Mammogram usually suffers from digitization noises such as straight lines and radio-opaque artifacts in the form of labels and wedges. In the computer-aided detection process, identification of the pectoral muscle region and breast profile extraction is an essential step in the methodology of preprocessing.

Median filter and other morphological process are involved in the preprocessing step for the removal of noise and radio-opaque artifacts, respectively.

2.3 ROI Segmentation

An ROI is a certain region or part that has to be filtered out from the original image in order to perform further operations on it. This region can be extracted by multiplying the original image with a generated binary mask, which is of the same size as that of the image to be processed. The steps involved in ROI segmentation are given in Fig. 3.

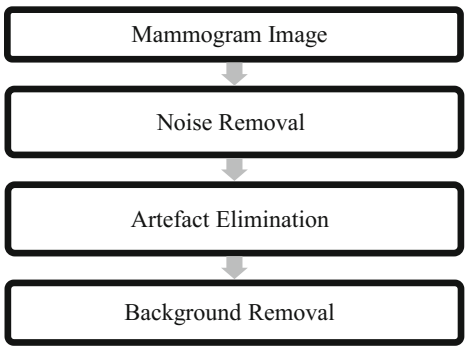


Fig. 2 Preprocessing method

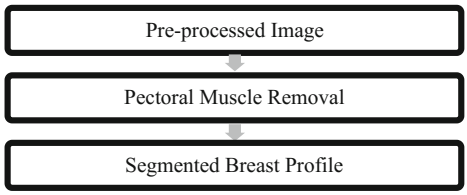


Fig. 3 ROI segmentation

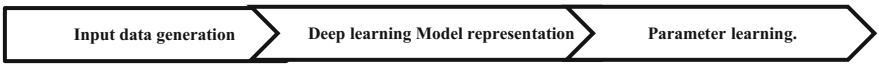


Fig. 4 Overall proposed approach

Region-based segmentation method such as seeded region growing is used for pectoral muscle removal. This approach of segmentation involves the initial seed point selection and determination of the neighboring pixels to be added to the region.

2.4 Cancer Detection by Deep Learning Technique

The architectures of "human-brain-inspired" deep nonlinear models compose complex features in the deeper layers of the network by analyzing the simple features learned in the previous layers. These features prove to be very effective descriptors in object recognition problems. During the training phase of these models, the features are encoded iteratively, and then, the learned weights are updated for improved optimization of the network. The features can be learned either using CNN in a supervised manner or using autoencoders in an unsupervised way. The features learned in a layerwise method are fed into a trained classifier, which predicts the labels. The classifier being a supervised layer has to be trained using a set of training images along with the associated label. The trained network should be able to accurately predict the label for an unseen image.

In this work, a deep convolutional neural network is used for supervised learning of features and a sparse autoencoder (SAE) and a stacked sparse autoencoder (SSAE) is employed for unsupervised learning of features from the mammogram.

2.5 Overall Approach

The overall proposed approach consists of three modules as shown in Fig. 4:

Input Data Generation. The use of entire images as input to the deep network is computationally expensive. Also, downsampling the images may lead to loss of fine details. So, a compact representation can be learned for local neighbors (or

patches) from the image. The size of each patch is limited to 100 × 100 in order to limit the weights and bias to be trained.

Deep Learning Model Representation. The patches are processed by a multilayer convolutional architecture. It consists of three layers in case of deep CNN as shown in Fig. 5. The unsupervised learning network is made of autoencoders. The sparse autoencoder (SAE) consists of four hidden layers, and the stacked sparse autoencoder (SSAE) is composed of three such autoencoders containing four hidden layers each.

Parameter Learning. The parameters of feature representations are obtained by encoding the input using a cascade of trainable transformations. These parameters are learned in a layerwise fashion with or without using the labels. While an individual layer is not deep, the stacked architecture is used. Thus, the individual training of ("shallow") layers results in a deep learning procedure. Figure 6 shows the steps required for transformation of one feature representation to another one.

Finally, a classifier maps the last feature representation into label space. A softmax classifier is used in the proposed system for this purpose. All the extracted patches within the testing image using the sliding window approach are tested by applying the trained hypothesis function. That is, every patch is sent through the trained network for prediction.

2.6 *Risk Scoring*

After the classifier maps the patches into label space, the mammograms can be scored risky or normal based on the density score of the patches. The mammograms

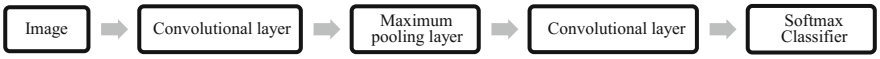


Fig. 5 Multilayer convolutional architecture

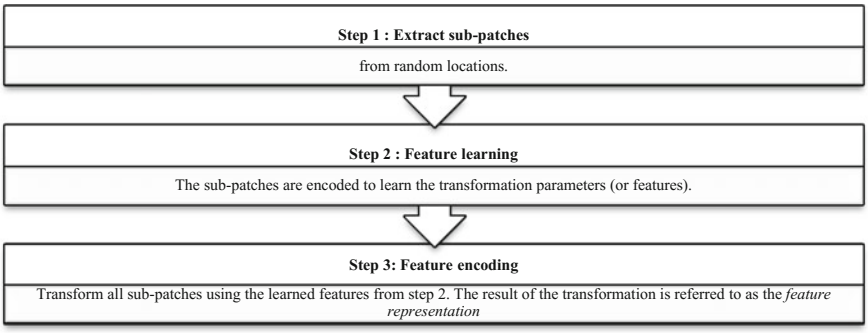


Fig. 6 Feature transformation

are classified as risky, even if at least one of the constituent patches is dense. This technique is similar to that of multiple instance learning.

3 Breast Cancer Imaging Modalities

Different modalities are available for breast cancer screening and detection. Efforts have been made to improve the accuracy of breast cancer diagnosis using different imaging modalities. Figure 7 gives the comparison of different modalities in terms of effective radiation dose to the human body.

Even though the combined PET and CT provide a better view of the suspect cells, the imaging techniques such as CT and PET cause high radiation exposure to patients. In case of using these techniques for screening, the sheer amount of radiation the body is exposed to, while imaging, may itself trigger cancer development. So it is always safer to use imaging techniques such as mammogram and MBI which have comparatively lower effective radiation dose.

In this work, mammogram images are used which is safer in terms of effective radiation dose. Traditional mammograms use X-rays to generate information about the anatomical structure of a lesion.

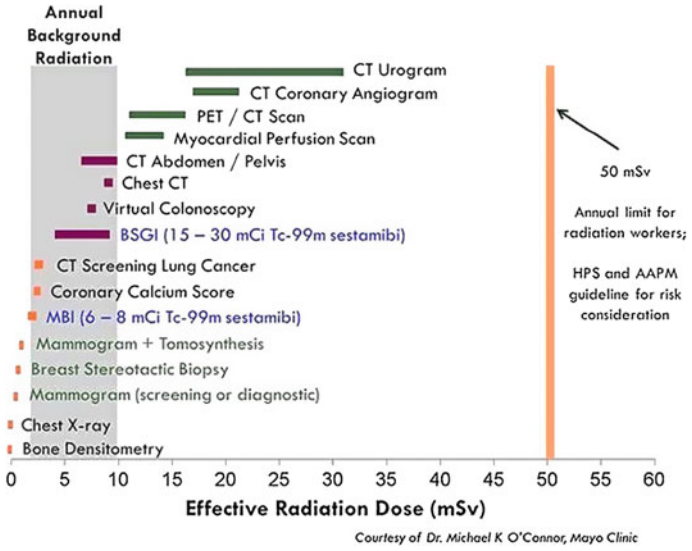


Fig. 7 Effective radiation dose for various medical imaging modalities

4 Deep Learning Techniques

Deep learning techniques can be applied to clustering, classification, and pattern recognition in medicine, biology, robotics, etc. The latest advances in the field have been caused by the invention of deep learning methods. The key component of deep learning is the multilayered hierarchical data representation typically in the form of a neural network with more than two layers. Such methods allow automatically synthesizing data descriptions (features) of a higher level based on the lower ones. A neural network (NN) consists of many interconnected neurons. The number of neurons and the interconnection depends on the application for which it is used. The deep learning techniques can be broadly classified into [10]:

1. *Deep networks for unsupervised or generative learning* (unlabeled data)
2. *Deep networks for supervised learning* (labeled data)
3. *Hybrid deep networks* (combination of 1 and 2).

4.1 Unsupervised Learning

Unsupervised learning is advantageous in cases where there is a huge amount of unlabeled data. By applying unsupervised deep learning techniques to such data, features that are better compared to handcrafted features can be learned.

Autoencoder. An autoencoder is used in unsupervised deep learning techniques. An autoencoder consists of an encoder followed by a decoder as shown in Fig. 8. The encoder part transforms the input x to z using $h^{e}(.)$. The decoder part tries to get back the original input x from z. The decoder outputs $\widehat{x}$ which is the approximate reconstruction of the original input x.

An autoencoder is trained using unlabeled data in an unsupervised fashion, to produce an output which replicates the input. During the training phase, a cost function is minimized. The usually used cost function is cross-entropy as it has lower stagnation periods compared to mean square error.

The encoder part of the autoencoder maps the input vector x, to the vector z using the $h^{e}(\cdot)$ as follows:

$$z^{e} = h^{e}(W^{e}x + b^{e}) \tag{1}$$

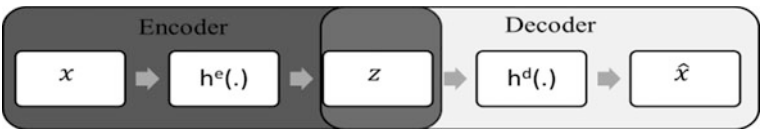


Fig. 8 Autoencoder

where W^{e} is the weight matrix, b^{e} is the bias vector, and $h^{e}(\cdot)$ is the activation function. The activation function may be sigmoid, hyperbolic tangent or rectified linear function. The sigmoid function, $h^{e}(x) = \frac{1}{1+e^{-x}}$, is used in the proposed system. The decoder part of the autoencoder maps the encoded representation z to the approximate reconstruction of the original input vector x using the function $h^{d}(.)$ as follows:

$$\widehat{x} = h^{d}\left(W^{d}x + b^{d}\right) \tag{2}$$

where h^{d} is the transfer function for the decoder, W^{d} is the weight matrix, and b^{d} is the bias vector.

Sparse Autoencoders (SAE). The use of regularizer to the cost function introduces sparsity in an autoencoder [19]. This regularizer is a function of the average output activation value of a neuron. The average output activation measure of a neuron i is defined as follows:

$$\widehat{\rho}_{i} = \frac{1}{n}\sum_{j=1}^{n} z_{i}^{e}\left(x_{j}\right) = \frac{1}{n}\sum_{j=1}^{n} h\left(w_{i}^{eT}x_{j} + b_{i}^{e}\right) \tag{3}$$

where n is the total number of training examples, x_j is the jth training example, w_{i}^{eT} is the ith row of the weight matrix W^{e}, and b_{i}^{e} is the ith entry of the bias vector, b^{e}. A neuron is active, if its output activation value is high. The neuron in the hidden layer fires in response to only a small number of the training examples for a low-output activation value. The neurons can be made to fire to a small number of training values by adding a term to the cost function that constrains the values of $\widehat{\rho}_{i}$. Thus, each neuron specializes by responding to some feature that is only present in a small subset of the training examples.

Sparsity Regularization. A sparse autoencoder can be realized by using a sparsity regularizer with a regular autoencoder. When most of the entries are zero with the exception of only a small number of nonzero entries, this feature representation is called sparse representation. These representations are robust to noise, simple, and cost-efficient. By incorporating population and lifetime sparsity, compact encoding and example-specific features can be obtained.

Sparsity regularizer enforces sparsity on the output from the hidden layer. The sparsity regularization term is made large when the average activation value, $\widehat{\rho}_{i}$, of a neuron i and its desired value, ρ, are not close in value [19]. One such sparsity regularization term is the Kullback–Leibler divergence.

$$\Omega_{\text{sparsity}} = \sum_{i=1}^{D^{e}} \mathrm{KL}(\rho||\widehat{\rho}_{i}) = \sum_{i=1}^{D^{e}} \rho \log\left(\frac{\rho}{\widehat{\rho}_{i}}\right) + (1-\rho)\log\left(\frac{1-\rho}{1-\widehat{\rho}_{i}}\right) \tag{4}$$

Kullback–Leibler divergence is a function which measures the difference between two different distributions. In this case, it takes the value zero when ρ

and $\widehat{\rho}_i$ are equal to each other and becomes larger as they diverge from each other. Minimizing the cost function forces this term to be smaller, hence ρ and $\widehat{\rho}_i$ to be close to each other.

L2 Regularization. When training a sparse autoencoder, it is possible to make the sparsity regularizer small by increasing the values of the weights w^{e} and decreasing the values of z^{e} [19]. Adding a regularization term on the weights to the cost function prevents it from happening. This term is called the L2 regularization term and is defined by:

$$\Omega_{\text{weights}} = \frac{1}{2}\sum_{l}^{L}\sum_{j}^{n}\sum_{i}^{k}\left(w_{ji}^{\mathrm{e}}\right)^2 \tag{5}$$

where L is the number of hidden layers, n is the number of observations (examples), and k is the number of variables in the training data.

Cost Function. The cost function for training a sparse autoencoder is an adjusted mean square error function as follows:

$$E = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\left(x_{tkn} - \widehat{x}_{kn}\right)^2 + \lambda * \Omega_{\text{weights}} + \beta * \Omega_{\text{sparsity}} \tag{6}$$

where λ is the coefficient for the L2 regularization term, and β is the coefficient for the sparsity regularization term.

Stacked Sparse Autoencoder (SSAE). A SSAE can be constructed by stacking together multiple layers of basic stacked sparse autoencoders (SAEs) such that the output from the first layer is fed as the input to the next layer. In this work, the SSAE is formed by stacking together three SAEs. Each of these SAEs has four hidden layers. This is followed by a softmax classifier.

4.2 Supervised Learning

Deep networks for supervised learning, which are intended to directly provide discriminating power for pattern classification purposes. It is often done by characterizing the posterior distributions of classes conditioned on the visible data. Target label data are always available in direct or indirect forms for such supervised learning. They are also called discriminative deep networks.

Convolutional Neural Network. A CNN is different from the ordinary backpropagation neural network (BPN) because a BPN works on extracted handcrafted image features, whereas a CNN works directly on an image to extract useful, necessary features for classification. A CNN is usually made up of a number of convolutional layers, pooling layers, fully connected layers, classification layers, etc. When the size of the image given as input to the CNN is $m \times m \times r$, where m is the height and width of the image and r is the number of channels, and

the convolutional layer has k filters of size $n \times n \times q$, k feature maps of size $m - n + 1$ is produced by convolving the image with these filters. Each map is then subsampled typically with mean or max-pooling layers. Subsampling rate usually varies from 2 to 5. After the convolutional layers, there may be any number of fully connected layers.

5 Implementation

The implementation is done using the computer system with MATLAB R2016a having a Windows 10 OS, Intel® Core™ i7-5500 CPU, x64-based processor, 2.40 GHz clock speed, and 8 GB RAM.

5.1 Preprocessing

The procedure for implementing the steps involved in preprocessing of mammogram such as digitization noise removal, artifact suppression, and background removal is discussed below.

Digitization Noise Removal. A 2D median filter of size 3×3 is employed in removing the digitization noises present in the mammogram [4]. In this technique, the center pixel in the 3×3 square covered by the filter mask is replaced by the median of all the 9 pixels in the neighborhood. While computing median in the edges, zero padding is done.

Artifact Suppression and Background Separation. Removal of radio-opaque artifacts present in the mammograms can be done using morphological operations. İn order to convert the mammogram image which is in grayscale to a binary image, the threshold is selected to be $T = 0.0706$. This value of the threshold is found to hold good for most mammograms through manual inspection [4]. The step-by-step procedure involved in artifact suppression and background separation is as follows:

1. First, the objects present in the binary image of the mammogram has to be labeled. The objects include radio-opaque artifacts and the breast profile.
2. The "area" of each object is calculated by counting the number of pixels in the object region.
3. The object with the largest area, which is usually the breast profile region, is selected. Using morphological opening operation, all other objects in the binary image are removed except the breast profile.
4. In order to smoothen the image, isolated pixels have to be removed. This is accomplished by setting a pixel to 1, only when it is surrounded by 5 or more 1 s in a 3×3 neighborhood.
5. The resulting image is first eroded and then dilated using a structuring element of radius 5, which is flat and disk-shaped.

6. Finally, the breast profile is segmented from the background by multiplying the resulting binary image from the previous step with the original grayscale mammogram image. The final grayscale image is devoid of noises and radio-opaque artifacts.

5.2 ROI Segmentation

The pectoral muscle has to be removed from breast profile to have unbiased tumor detection. Seeded region growing (SRG) technique is used for pectoral muscle removal.

Pectoral Muscle Segmentation. The steps involved in the removal of the pectoral muscle region through SRG technique [4] are as follows:

1. *Orientation determination*: As a prerequisite to performing SRG, the breast orientation (left or right) in the mammogram has to be known. The orientation of the breast can be found by comparing the sums of first and last 5 columns in the preprocessed image. If the sum of first five columns is greater than the last five columns, then the breast is right orientated else it is left-orientated.
2. *Seed Point Selection*: Depending on whether the breast is left-oriented or right-oriented, the seed point is selected to perform SRG. The pixel in the 5th row and 5th column is selected as the seed point, if the mammogram is right-oriented. İn case of left-oriented mammogram, the pixel in the 5th row and last 5th column is selected as the seed point.
3. *Seeded Region Growing*: In this technique, the region is grown by comparing the intensity values of the neighboring unallocated pixels. When the difference between these intensity values is less than a threshold value, it is iteratively added to the region. The threshold is selected to be $T = 32$. This value of the threshold is found to hold good for most mammograms for reliable pectoral muscle removal [4]. At the end of SRG, a binary image of the segmented pectoral muscle is obtained.
4. The resulting image is first eroded and then dilated using a structuring element of radius 3, which is flat and disk-shaped.
5. Finally, the binary image obtained from the previous step is multiplied with the preprocessed grayscale image to give the final grayscale mammogram image with the segmented pectoral muscle.

5.3 Cancer Detection Using Deep Learning

The proposed system uses various deep learning techniques to detect breast cancer in mammogram images. The cancer detection using deep learning includes the

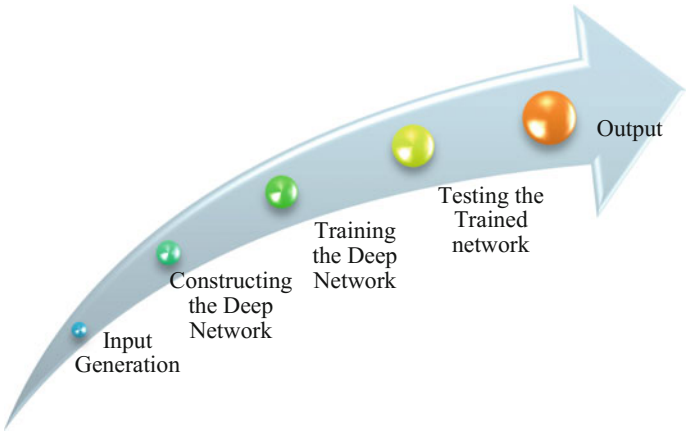


Fig. 9 Steps in cancer detection using deep learning

following implementation steps: input generation, construction of deep network, training the network, and testing the network. The steps in cancer detection are shown in Fig. 9.

Constructing the Deep Network. In this work, three types of deep networks are employed, namely CNN (supervised), SAE, and SSAE (unsupervised).

Convolutional Neural Network. Employing CNNs to classify medical images can be done in three ways. The first method is to build and train the CNN to obtain features. The second method is to use "off-the-shelf CNN features" without retraining the CNN. The third method is to use CNN in fine-tuning the results obtained using other deep learning models.

The first technique is used in building the CNN in this work. The CNN is constructed with 3 layers as shown in Fig. 10. The layers are as follows:

1. Convolutional layer,
2. Maximum pooling layer, and
3. Convolutional layer

The input to the CNN is a 4D matrix containing 100 × 100 image patches. The maximum pooling layer downsamples the convolved image to one-fourth of the original image size. The classification is performed by a softmax classifier.

Sparse Autoencoder. The deep network is constructed by stacking an unsupervised autoencoder and a supervised softmax classifier.

The autoencoder has 4 hidden layers with sparsity parameter $\rho = 0.01$ and weighting term of sparsity regularizer $\lambda = 1$. During the training phase, the autoencoder learns a hypothesis function

$$h : X \rightarrow Y \tag{7}$$

where X is the input image, and Y is the approximate reconstruction of the input image. The encoder part maps the input X to the hidden layers, and the decoder part tries to reconstruct X from the encoded features. The autoencoder applies this hypothesis function h to the input X during the testing phase. Figure 11 shows the deep network that is constructed by stacking autoencoder and softmax classifier.

In the unsupervised part, features are learned using autoencoders. A sparse autoencoder is used in this work. When most of the entries are zero with the exception of only a small number of nonzero entries, this feature representation is called sparse representation. These representations are robust to noise, simple, and cost-efficient. By incorporating population and lifetime sparsity, compact encoding and example-specific features can be obtained.

Stacked Sparse Autoencoder (SSAE). A SSAE can be constructed by stacking together multiple layers of basic stacked sparse autoencoders (SAEs) such that the output from the first layer is fed as the input to the next layer. In this work, the SSAE is formed by stacking together three SAEs. Each of these SAEs has four hidden layers. This is followed by a softmax classifier. The architecture of SSAE is shown in Fig. 12.

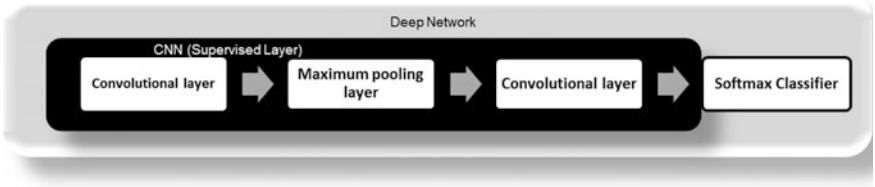


Fig. 10 Convolutional neural network

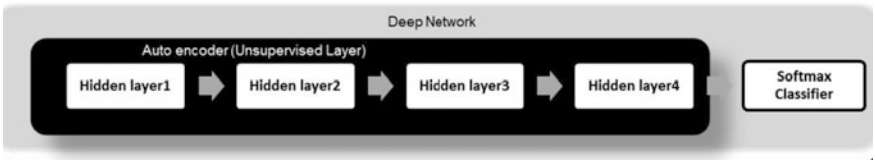


Fig. 11 Sparse autoencoder

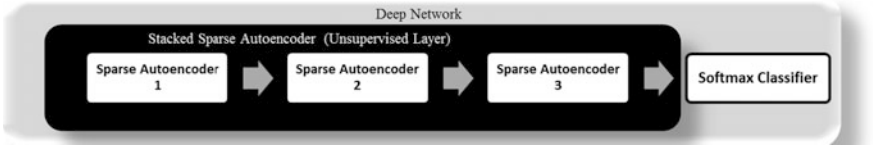


Fig. 12 Stacked sparse autoencoder

Softmax Classifier. Softmax classifier is a supervised layer. Softmax classifier outputs probabilities rather than margins. Probabilities are much easier for us as humans to interpret, so that is a particularly nice quality of softmax classifiers. The regularization term is appended to the loss function and is used to control the weight matrix *W*. By controlling *W*, classification accuracy can be increased. The three class labels for density scoring are background, fatty, and dense. Figure 13 shows the classes in density scoring.

Training the Deep network. Complex classification problems can be solved by using neural networks with multiple hidden layers, where each layer learns different features. Though such networks are attractive for medical image analysis, the training process can be expensive in terms of memory and time.

Greater accuracies of deep learning models can be achieved by using a large amount of training data. However, using such huge amounts of data can slow down the training process. Once a deep learning model is trained, it can be used in real-time applications.

The SAE deep network can be trained by training the autoencoder and softmax classifier separately and then training the deep network after stacking the softmax classifier and SAE. In case of SSAE deep network, each of the autoencoders is trained separately and is again trained after stacking them together. The autoencoder is trained in an unsupervised way, but the softmax classifier and the whole network are trained in a supervised way. Figure 14 depicts the deep network training.

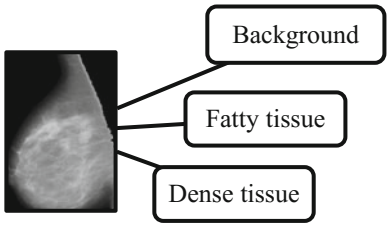


Fig. 13 Softmax classifier classes

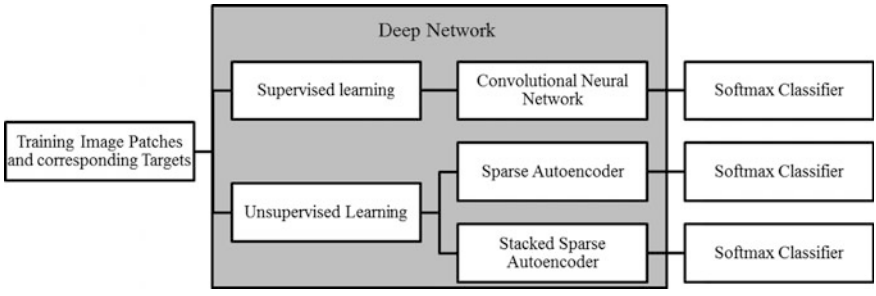


Fig. 14 Training the deep network

Testing the Trained Network. The output of the trained network exactly matches the targets only if the training was done perfectly. If the desired output is not obtained, the network can be trained again. As the network parameters change each and every time the network is trained, the output may differ in a desired way. Another approach is to increase the number of layers of the network because increasing the number of layers increases the flexibility of the network. Or a different training function can be used. Finally, the training data can be increased, as it is likely to improve the performance of the network. Figure 15 depicts the testing of data using the trained deep network.

6 Results and Discussion

There are many different publicly available mammographic databases. The mini-MIAS mammographic database is used in testing the techniques described in this work.

6.1 Database

The database contains 322 digitized films of 1024 × 1024 images. The ground truth markings for all the mammograms are provided by the radiologist, and also, the location of abnormalities is provided.

6.2 Preprocessing

The various preprocessing stages of mammogram are digitization noise removal, artifact suppression, and background removal. The results obtained from each of these stages are discussed in the following sections.

Digitization Noise Removal. A 2D median filtering with 3 × 3 neighborhood window is used to remove the digitization noises present in the input image as

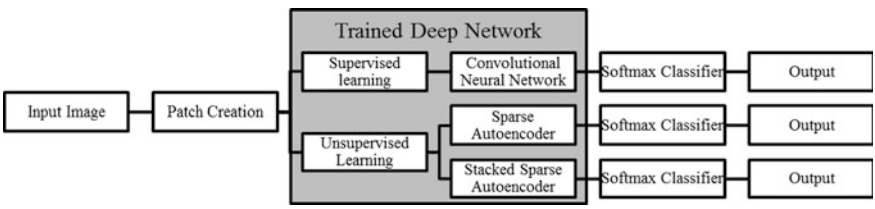


Fig. 15 Deep network testing

shown in Fig. 16a. The resultant image after removal of digitization lines is given in Fig. 16b.

Artifact Suppression and Background Separation. Mammograms with radio-opaque artifacts are shown in Fig. 17a. These artifacts are removed by morphologically opening the binary image. The binary image is shown in Fig. 17b. The resulting binary image after all the morphological operations is done on it as shown in Fig. 17c. The final image is obtained by multiplying Fig. 17a, c, and the final grayscale image is shown in Fig. 17d.

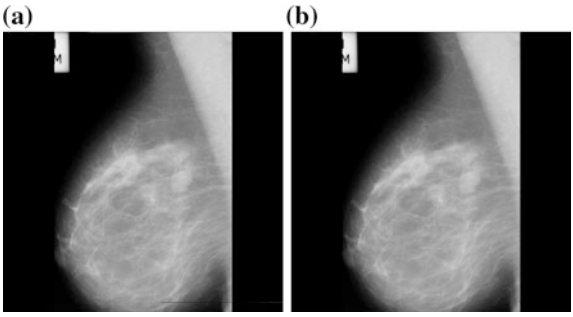


Fig. 16 **a** Original image and **b** Noise-free image

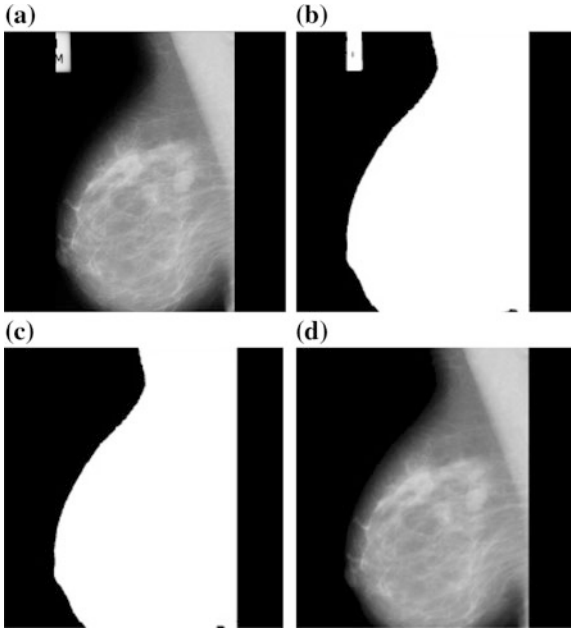


Fig. 17 **a** Noise-free image, **b** binary image, **c** morphologically opened image, and **d** preprocessed image

6.3 ROI Segmentation

The undesired effect of pectoral muscle on the detection of breast cancer warrants it be removed before the detection step. The removal of pectoral muscle is done using SRG technique. The grown region is shown in Fig. 18b. This image is multiplied with the preprocessed grayscale image (Fig. 18a) to get the final grayscale mammogram image with the segmented pectoral muscle shown in Fig. 18c.

6.4 Cancer Detection Using Deep Learning

The cancer detection using deep learning includes the following implementation steps: input generation, construction of deep network, training the network, and testing the network.

Patch Creation. The input image is divided into patches of size 100 × 100 to reduce complexity in computation. The patches extracted from mammogram are shown in Fig. 19. The training patches are selected such a way that there are 10% background patches, 45% fatty tissue patches, and 45% dense tissue patches.

The use of entire images as input to the deep network is computationally expensive. Also, downsampling the images may lead to loss of fine details. So, a compact representation can be learned for local neighbors (or patches) from the image. Thus, images are separated into patches of size 100 × 100 as shown in Fig. 20.

Construction of the Deep Network. The deep networks used in this work are CNN, SAE, and SSAE.

Convolutional Neural Network. The construction of a CNN involves the selection of optimal value for various parameters such as learning rate and number of epochs which influence the accuracy, runtime, etc. The learning rate is set to be

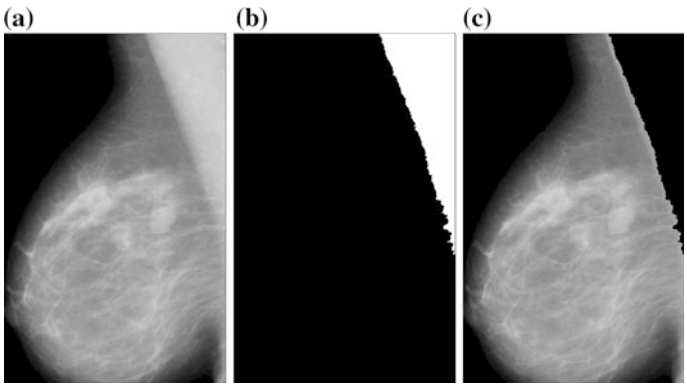


Fig. 18 **a** Preprocessed image, **b** segmented pectoral muscle, and **c** after pectoral muscle removal

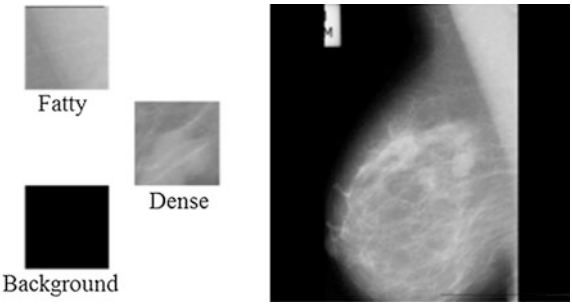


Fig. 19 Patch extraction for mammogram

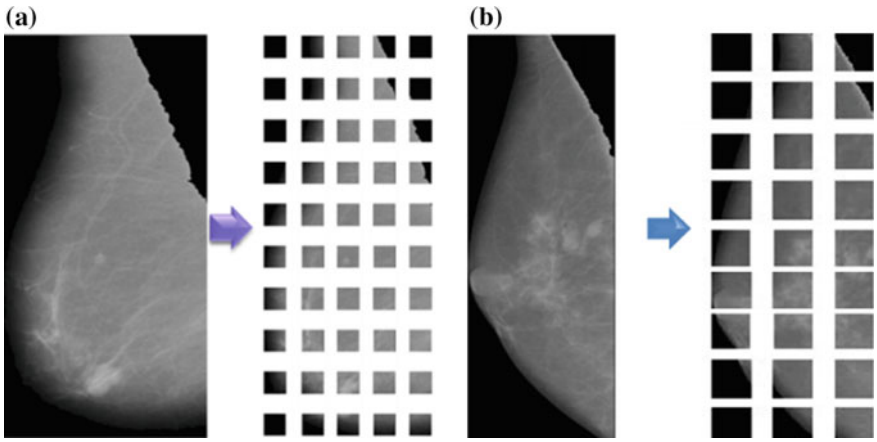


Fig. 20 Mammogram separated into 100 × 100 patches

Fig. 21 Comparison of changing the number of epochs

Number of Epochs	Accuracy	Time
200	77.41	18.65
400	81.85	36.82
600	88.15	54.17
800	93.7	70.99
1000	97	93.86

0.001. Because setting a higher value for learning rate results in suboptimal performance, while lower values increase the training time. Figure 21 shows the effect of varying the number of epochs on the accuracy and runtime. Increasing the number of epochs increases the accuracy, but has longer runtimes.

Sparse Autoencoder. The construction of a sparse autoencoder involves the process of adjusting the network parameters and other settings so that the network

gives optimal results for the current input data. The parameter to be considered while constructing the network includes number of epochs, cost function, decoder function, training algorithm, etc. Figure 22 shows the comparison of various decoder functions in terms of accuracy. By comparing, it is found that linear transfer function gives the best performance in terms of accuracy.

Figure 23 shows the comparison of various cost functions or loss functions in terms of accuracy. By comparing, it is found that minimizing cross-entropy function gives the best performance in terms of accuracy.

Figure 24 shows the comparison of various training algorithms in terms of accuracy and runtime. By comparing, it is found that the resilient backpropagation algorithm gives the best performance in terms of accuracy as well as minimum runtime.

Figure 25 shows the performance comparison in terms of various sizes of training sets in terms of accuracy. By comparing, it is found that accuracy and runtime increase with the increase in the size of the dataset used for training.

6.5 *Performance of the Deep Network*

Performance measurement is done to see whether the output obtained is in line with the target that was intended to be achieved. The performance metrics used for evaluating the performance of the deep network are listed below.

Cross-Entropy Function. The use of cross-entropy function as the minimizing function provides better network performance and short stagnation periods. Thus, the learning time is faster compared to using MSE (mean square error) as minimizing function [20].

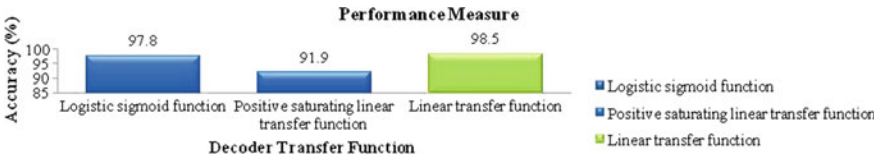


Fig. 22 Comparison of decoder transfer functions

Fig. 23 Comparison of cost functions

Fig. 24 Comparison of training algorithms

Training Algorithm	Accuracy	Time
Levenberg-Marquardt	98.5	29.84
Bayesian Regularization	88	
BFGS Quasi-Newton	98.5	32.62
Resilient Backpropagation	**98.5**	**27.69**
Scaled Conjugate Gradient	98.5	34.73
Conjugate Gradient with Powell/Beale restarts	98.5	35.32
Fletcher-Powell Conjugate Gradient	98.1	
Polak-Ribiere Conjugate Gradient	89.6	
One Step Secant	96.8	
Variable Learning Rate Gradient Descent	90.7	
Gradient Descent with Momentum	93.7	
Gradient Descent	92.2	

Fig. 25 Comparison of size of training sets

$$\text{Cross Entropy} = \frac{1}{m}\sum_{k=1}^{m}[t_k \ln(y_k) + (1 - t_k)\ln(1 - y_k)] \tag{8}$$

where t is the target, and y is the actual output.

Confusion Matrix. The confusion matrix is a two-dimensional table with the output and target classes. This shows whether the system is able to classify the classes correctly or it confuses the classes during classification. Various performance measures such as accuracy, misclassification rate, and specificity can be derived from the confusion matrix.

6.6 Performance Measures

Accuracy and misclassification rate can be calculated from the confusion matrices.

Accuracy. Accuracy is the ratio of correct classifications to the total number of inputs.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TOTAL}} \tag{9}$$

where TP—true positive and TN—true negative

Misclassification Rate. Misclassification rate is the ratio of wrong classifications to the total number of inputs.

$$\text{Misclassification Rate} = \frac{(\text{FP} + \text{FN})}{\text{TOTAL}} \tag{10}$$

where FP—false positive and FN—false negative.

Sensitivity or Recall. Sensitivity measures the proportion of correctly identified positives. It is the ratio of true positives to the total correct classifications.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{11}$$

where TP—true positive, TN—true negative, and FN—false negative.

Specificity. Specificity measures the proportion of correctly identified negatives. It is the ratio of true negatives to the total wrong classifications.

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})} \tag{12}$$

where TN—true negative, FP—false positive, and FN—false negative.

Precision. Precision is the ratio of a number of instances correctly classified as positive to all events classified as positive.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{13}$$

where TP—true positive and FP—false positive.

Negative Predictive Value (NPV). Negative predictive value is the probability of an event classified as negative that is actually negative.

$$\text{Negative Predictive Value} = \frac{\text{TN}}{(\text{FN} + \text{TN})} \tag{14}$$

where TN—true negative and FN—false negative.

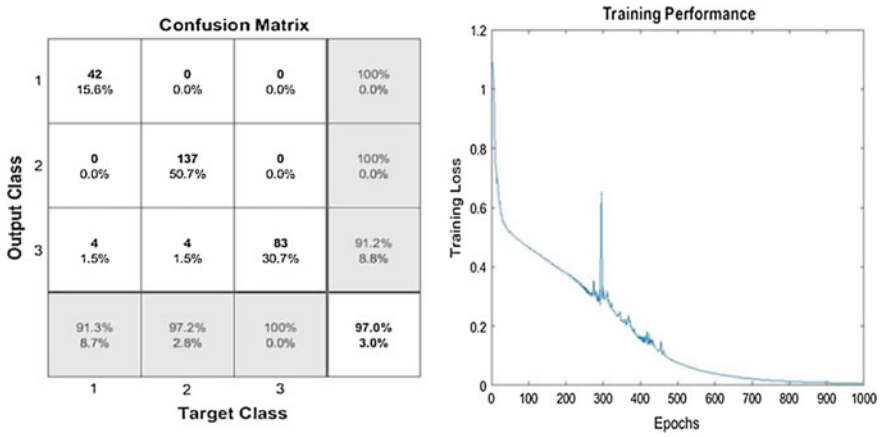


Fig. 26 Confusion plot and training performance in terms of training loss of CNN

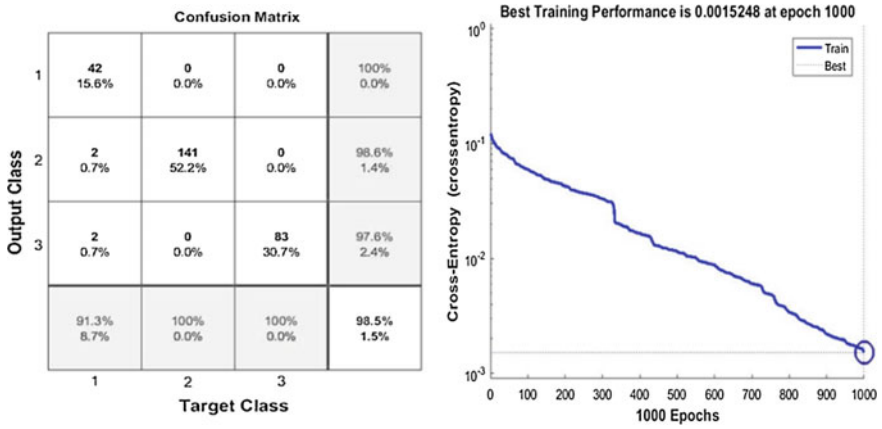


Fig. 27 Confusion plot and training performance of SAE in terms of minimizing the cross-entropy function

Figure 26 shows the confusion plot and training performance in terms of training loss for CNN, and Figs. 27 and 28 show the confusion plots and training performance in terms of cross-entropy for SAE and SSAE, respectively.

The performance of the deep learning techniques such as convolutional neural network, sparse autoencoder, and stacked sparse autoencoder is compared in terms of various performance measures such as accuracy, sensitivity, specificity, precision, and negative predictive value. Figure 29 shows the bar chart comparing the results obtained for various performance measures of these deep networks. From the comparison chart, it can be inferred that the stacked sparse autoencoder performs better than both CNN and SAE in all the cases.

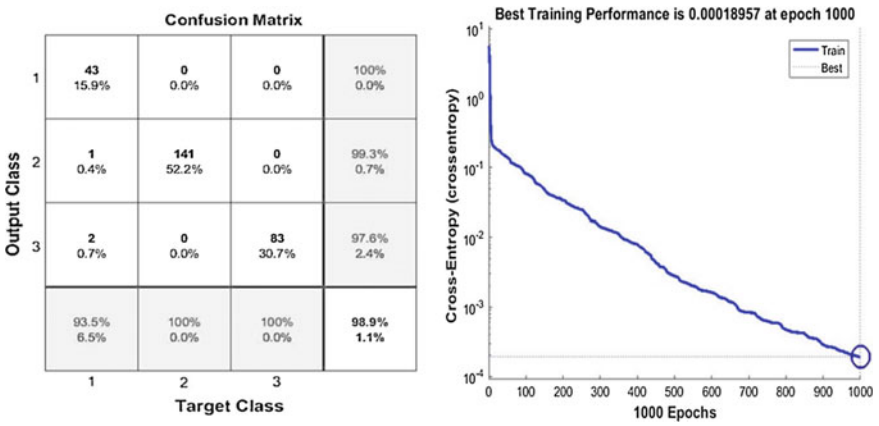
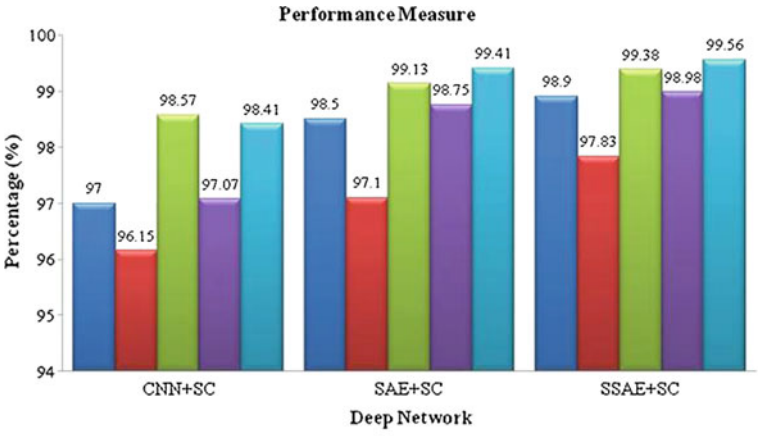
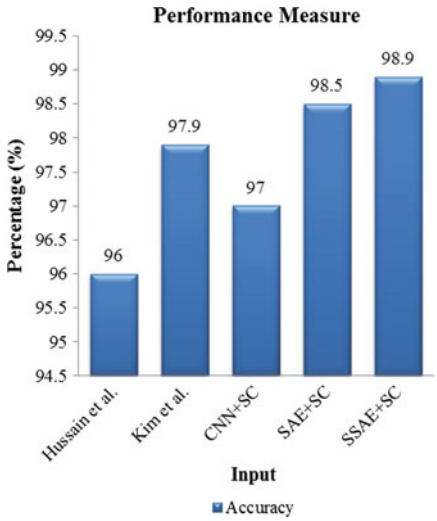


Fig. 28 Confusion plot and training performance of SSAE in terms of minimizing the cross-entropy function



	Accuracy	Sensitivity or Recall	Specificity	Precision	Negative Predictive Value
CNN+SC	97	96.15	98.57	97.07	98.41
SAE+SC	98.5	97.1	99.13	98.75	99.41
SSAE+SC	98.9	97.83	99.38	98.98	99.56

Fig. 29 Comparison of performance of the three deep learning techniques in terms of various performance measures such as accuracy, sensitivity, specificity, precision, and negative predictive value using MIAS database



Abbreviation	Full Form
CNN	Convolutional Neural Network
SAE	Sparse Autoencoder
SSAE	Stacked Sparse Autoencoder
SC	Softmax Classifier

Fig. 30 Comparison of performance of the three deep learning techniques with previous techniques using MIAS database

The performance of the deep learning techniques used in this work (convolutional neural network, sparse autoencoder, and stacked sparse autoencoder) was compared with the performance of the previously proposed techniques which detected breast cancer in mini-MIAS database. The performance was compared in terms of accuracy. Figure 30 shows the bar chart comparing the various techniques. By comparing all these techniques, it can be found that the stacked sparse autoencoder performs better compared to all other techniques.

6.7 Density Scoring

After training the deep network, this trained network is used in predicting the density score of mammogram. The input image is first divided into patches, and the density of each patch is predicted. Figure 31 shows the density scoring results for a mammogram. The mammogram can be classified as abnormal even if any one of the patches extracted from the mammogram is classified as dense. The mammogram is said to be normal only if all the patches constituting the mammogram image are classified either as fatty or background. This technique is similar to that of multiple instance learning (MIL). Figure 31a shows two dense patches which correspond to the dense mass present in the mammogram, and Fig. 31b shows the density scoring of patches obtained from a normal mammogram which shows no dense patches.

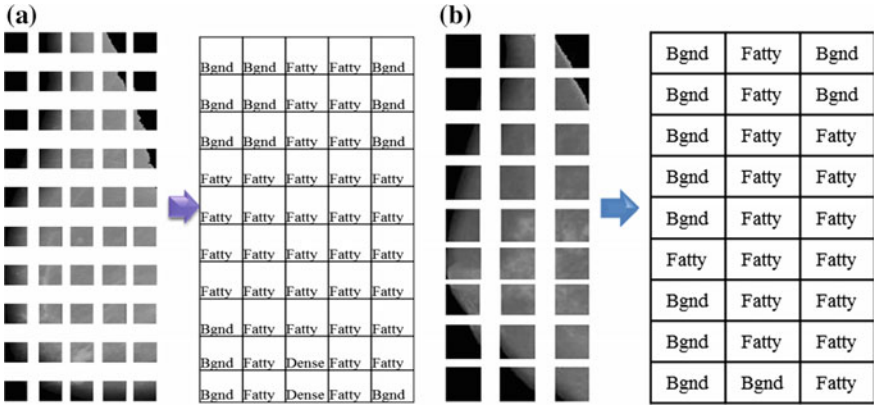


Fig. 31 Density scoring for mammogram

Thus, by comparing both these results, it is clear that only mammograms with no dense patches can be classified as normal. Even if there is an instance of at least one dense patch, the mammogram is classified as dense.

7 Conclusion

In the proposed work, deep learning techniques such as convolutional neural network (CNN), sparse autoencoder (SAE), and stacked sparse autoencoder are used for breast cancer detection using mammographic images. The mammograms from the mini-MIAS database, which is publicly available, are used in this work. The mammograms are first preprocessed to remove digitization noise, radio-opaque artifacts, background, and pectoral muscle, which reduce the effectiveness of the deep network in detecting cancer. The proposed model has achieved an accuracy of up to 97% for CNN (supervised learning), 98.5% for SAE, and 98.9% for SSAE (unsupervised learning) in classifying dense mammogram images.

The use of deep learning techniques in breast cancer detection from mammogram helps to identify smaller masses more accurately to pinpoint their locations, distinguish between cancerous and harmless lesions, and provide quantitative information about these lesions if they are cancerous. This helps in providing earlier treatment for women with breast cancer and spares other women the pain and anxiety of undergoing a biopsy.

References

1. Islam MS, Kaabouch N, Hu WC (2013) A survey of medical imaging techniques used for breast cancer detection. In: IEEE conference on Electro-Information Technology (EIT). IEEE Press, Rapid City, South Dakota, pp 1–5
2. Prasad SN, Houserkova D (2007) The role of various modalities in breast imaging. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 151(2):209–218
3. Shyamala K, Girish HC, Murgod S (2014) Risk of tumor cell seeding through biopsy and aspiration cytology. J Int Soc Prev Commun Dent 4(1):5–11
4. Nagi J, Abdul Kareem S, Nagi F, Khaleel Ahmed S (2010) Automated breast profile segmentation for roi detection using digital mammograms. In: IEEE EMBS conference on biomedical engineering & sciences. IEEE Press, Kuala Lumpur, Malaysia, pp 87–92
5. Tan M, Zheng B, Leader JK, Gur D (2016) Association between changes in mammographic image features and risk for near-term breast cancer development. IEEE Trans Med Imaging 35(7):1719–1728
6. Singh AK, Gupta B (2015) A novel approach for breast cancer detection and segmentation in a mammogram. Procedia Comput Sci 54:676–682
7. Pratiwia M, Alexandera, Harefaa J, Nandaa S (2015) Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network. Procedia Comput Sci 59:83–91
8. Kim DH, Choi JY, Ro YM (2012) Region based stellate features for classification of mammographic spiculated lesions in computer-aided detection. In: 19th IEEE international conference on image processing. IEEE Press, Orlando,Florida, pp 2821–2824
9. Hussain M, Khan S, Muhammad G, Bebis G (2012) A comparison of different Gabor features for mass classification in mammography. In: 8th international conference on signal image technology and internet based systems. IEEE Press, Naples, pp 142–148
10. Deng L, Yu D (2014) Deep learning: methods and applications. Now publishers, Boston
11. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Networks 61:85–117 Elsevier
12. Langkvist M, Karlsson L, Loutfi A (2014) A review of unsupervised feature learning and deep learning for time-series modelling. Pattern Recogn Lett 42:11–24 Elsevier
13. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: a review. Neurocomputing 187:27–48 Elsevier
14. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Guevara Lopez MA (2015) Convolutional neural networks for mammography mass lesion classification. In: 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE Press, Milan, pp 797–800
15. Sharma K, Preet B (2016) Classification of mammogram images by using CNN classifier. In: International conference on advances in computing, communications and informatics (ICACCI). IEEE Press, Jaipur, pp 2743–2749
16. Baldi P (2012) Autoencoders, unsupervised learning, and deep architectures. In: JMLR: workshop and conference proceedings. pp 37–50
17. Kallenberg M, Petersen K, Nielsen M, Ng AY, Diao P, Igel C, Vachon CM, Holland K, Winkel RR, Karssemeijer N, Lillholm M (2016) Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging 35(5): 1322–1331
18. Xu J, Xiang L, Liu Q, Gilmore H, Wu J, Tang J, Madabhushi A (2016) Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Trans Med Imaging 35(1):119–130
19. Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1. Vision Res 37(23):3311–3325 Elsevier
20. Nasr GE, Badr EA, Joun C (2002) Cross entropy error function in neural networks: forecasting gasoline demand. In: FLAIRS-02 Proceedings. pp. 381–384

A Tour Toward the Development of Various Techniques for Paralysis Detection Using Image Processing

Banita Banita and Poonam Tanwar

Abstract The facial paralysis is a disease that occurs due to facial nerve damage. It is the 7th cranial nerve which resides just behind the ear of human body. Facial nerve travels through a narrow canal, bonny canal called "fallopian canal" in skull, and exits from the skull via the stylomastoid foramen and then passes through parotid gland and subdivided to supply the facial muscles. Facial paralysis is a neuromuscular disorder which involves muscle of expression of the face and taste buds. The common form of facial paralysis is Bell's palsy that is somehow having same symptoms, but it affects exact half portion of face, i.e., either left side of face which includes eyes, cheeks, lips, and chin or right side of face. But in facial paralysis, one quadrant of face gets affected, i.e., either eyes or lips. In examination of facial paralysis, the face is divided into four quadrants and one or more quadrant may be affected as per the degree of facial paralysis. That is why maximum research defined Bell's palsy as facial palsy. But there is difference between the Bell's palsy and facial palsy. It is believed that the various grading system helps to evaluate the degree of facial paralysis in the patients. Incidence of facial paralysis of idiopathic disorder is 25 per 100,000 annually or about 1 in 60 persons in a lifetime. The rate of growth of patients increased with the age of human being. In the USA, it affects about 40,000 people every year. As rate of growth of facial paralysis patients increasing per year so, Computer Science Engineers are participating in this for betterment in recovery and accuracy using grading system and 3D image processing which is not yet used in facial paralysis area. Paper is used to present all common techniques for detecting the facial paralysis and calculating the degree of facial paralysis by using grading system. Paper represents the comparison of all the common grading systems and different problem under one umbrella. Main objective of this paper is to present a review on the progress of research efforts made in

B. Banita (✉)
School of Computer Science, Lingaya's University, Faridabad, Haryana, India
e-mail: banita861023@yahoo.com

P. Tanwar
School of Computer Science, Manav Rachna International University, Faridabad, Haryana, India
e-mail: Poonam.tanwar@rediffmail.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_9

various grading systems and then analyzing the scaling methods. The commonly used House–Brackmann grading system (HBGS) is discussed which can help to find the degree of facial paralysis. Higher the grade, more is the paralysis. The development of new method for detection and grading facial nerve function will be very challenging, but it will help the facial paralyzed patients up to some extent.

Keywords Grading system · Facial paralysis · Bell's palsy · Conduction velocity · Image processing

1 Introduction

The seventh cranial nerve is a facial nerve which is composed of motor fibers, parasympathetic nerve fibers, and sensory nerve fibers. It originates from brain and supplies to the muscles of the face, submandibular and sublingual salivary glands, lacrimal glands, and stapedius muscles. It also supplies to the taste buds in the anterior 2/3rd of the tongue, and it also supplies to the skin of the external ear and tympanic membrane. After leaving from the brain (pons), this nerve enters the internal auditory meatus with acoustic nerve. It continues its course in its own bonny canal which is about 30 mm in length and exits from the skull via stylo-mastoid foramen of skull. Then, it passes through the parotid glands and subdivides to supply the facial muscles. The facial nerve and intermediate nerve separate from eighth nerve at the lateral fundus of internal acoustic canal and run through the skull bone via bonny canal. The complete interruption of the facial nerve at the stylo-mastoid foramen paralyzes all the muscles of the face. Facial paralysis can be detected by noticing various facial expressions.

There are generally two categories of facial paralysis. The first category of facial paralysis is Bell's palsy which is also termed as peripheral facial paralysis. Bell's palsy occurs when there is damage to the VII cranial nerve. Bell's palsy leads to asymmetry in eyes, forehead, lips, and all the remaining facial regions. Bell's palsy affects the half portion of face due to nerve damage. The other category of facial paralysis is facial palsy. It deals with particular facial region which is affected due to damage to the corresponding branches of facial nerves. The facial palsy is due to damage to any branch of the facial nerve, and these branches spread over the face. Facial paralysis occurs only when there is injury to the nerve or swollen nerve which results in asymmetry in the facial expressions due to involvement of corresponding branches.

The affected portion of the face in case of Bell's palsy as well as in facial palsy can be detected by using facial grading systems. Grading system can be defined by traditional approaches and computer-based approaches.

1. **Traditional approaches**: It includes the marking points on the prominent landmarks on the face. Dysfunctioning of the facial region is assessed by measuring the distance between points on facial landmarks, and the level of paralysis is measured. Generally, traditional approaches are time-consuming. Traditional approaches are further classified in three scales:

(1.1) **Gross Scale**: These scales are used to access the overall facial motor functioning which is applicable only in Bell's palsy, and it deals with facial motor functioning. May scale, House–Brackmann, etc., are the common gross scale of facial grading system.

(1.2) **Regional Scale**: This assessment is made in terms of scoring different facial regions individually and is applicable in facial palsy which deals with facial regions evaluation. Smith scale, Sunnybrook scale, etc., are common regional scale of facial grading system.

(1.3) **Specific Scale**: It generally deals with questionnaire. Standard questions are being asked to the patient, and respective responses are recorded and evaluated by the scoring system. Specific scales are used in Bell's palsy as well as facial palsy. Nottingham system, Burres–Fisch scale, etc., are its commonly used facial grading system.

2. **Computer-Based Approaches**: It will assess the facial paralysis by using specialized equipment and various image processing techniques. "Maximum Static Response Assay (MSRA)" evaluates bilateral facial paralysis using regional and specific facial scaling system. Likewise, MSRA works by comparing amplitude of facial movement during facial expressions and at rest. Similarly, "Automated Face Image Analysis" (AFIA) system automatically detects and evaluates the current status of facial paralysis. In case of lip evaluation, AFIA will evaluate the facial paralysis degree which is based on lip movements. Different grading system involves different sections of facial expression just to detect the facial paralysis.

2 Grading System

2.1 Facial Nerve Function Index (FNFI)

Facial nerve dysfunctioning is categorized into two types. The first category refers to loss of motor function only in muscles of facial expression that is shown in Fig. 1. Loss of motor control may decrease the interobserver variability. Motor defects may involve some region of face more than others due to branching of facial nerve.

The second category refers to the collection of possible complications that occur due to nerve injury.

Total Facial Nerve Index (TFNI): It is defined as ratio of sum of each FNI to the number of indices [1].

$$\text{TFNI} = (\text{Sum of each FNI}/\text{Number of indices})$$

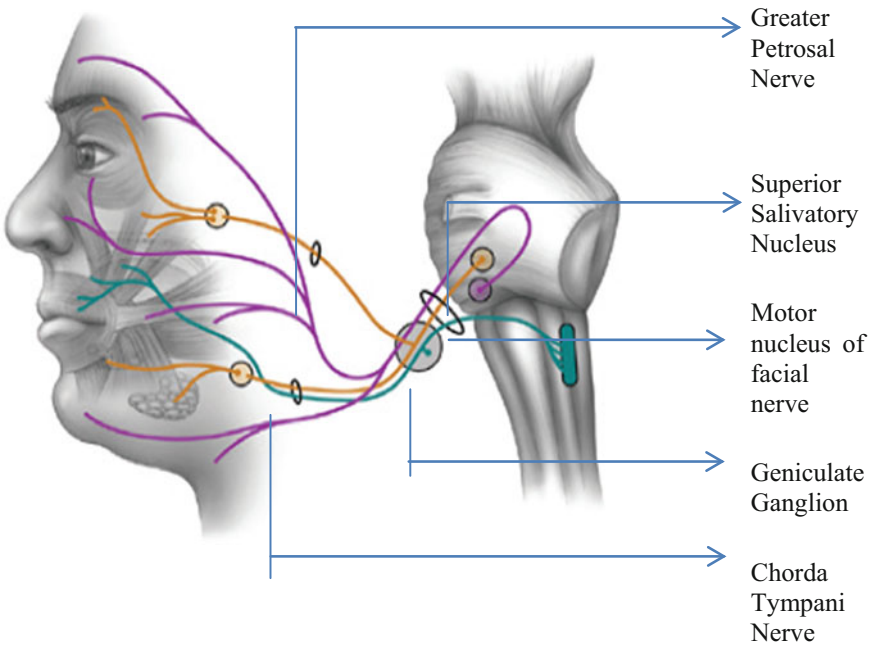


Fig. 1 Different area supplied by facial nerve

Facial grading system also includes three major sections:

1. *Evaluation of testing symmetry*: It is the initial phase of grading system which detects the current symmetry of face and is further compared by the asymmetry on the face.
2. *Degree of execution of facial muscles*: It can be evaluated with the help of electrodes. The current is passed to each nerve by using stimulator and various impulses so output can be recorded as a response to the current which will help to evaluate the working of the muscles supplied by that nerve.
3. *Degree of synkinesis associated with each movement of face*: It will help to detect the actual percentage of affected area of face. The affected area on face can be due to nerve damage, swollen nerve, or pinched nerves etc.

Here, the five cases of facial nerve indices are compared which are closing eyes, raising nose, raising eyebrow, plumping cheeks, and opening mouth. In Fig. 2, each bar shows the different facial expressions. If the probability of total facial nerve index (TFNI) is more than 0.80, then the patient may recover after 30 days, and if probability of TFNI value is less than 0.40, then the patient has to be diagnosed as early as possible as it indicates that the facial nerve of the patient is in bad condition. This system gives 95% of accuracy results [1].

Junyu Dong et al. [1] implemented the technique of salient point detection and used the K-means clustering for evaluating the degree of facial paralysis of patient

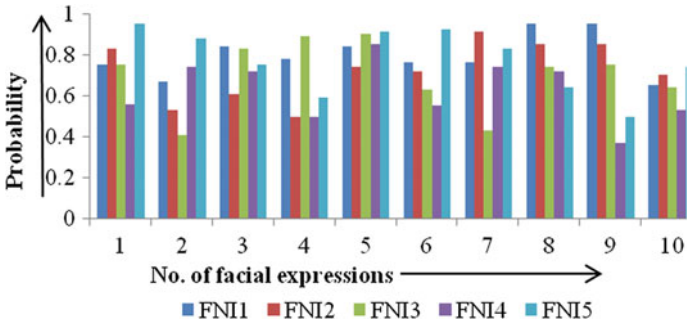


Fig. 2 Probability of facial expression [1]

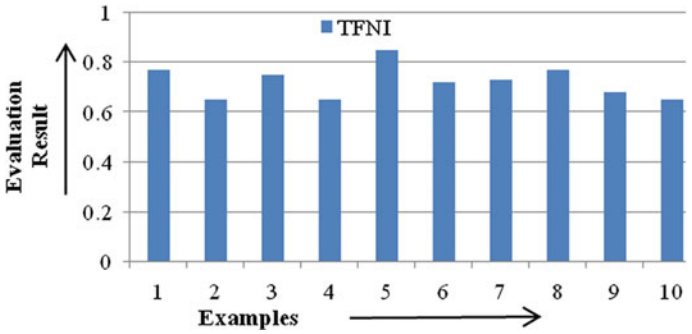


Fig. 3 Total facial nerve index [1]

to prove that the performance of system is better as compared to system considered by Cootes in 1998 and elected to use the active appearance model (AAM) for face recognition due facial paralysis.

Dong et al. [1] represent the five facial nerve indices, samples of 100 patients with different facial expressions were used in frontal view, and the result of point displacement is shown in Fig. 3. The result was well consistent with the diagnosis based on a group of doctor's decision and gives the accuracy of about 95%. This evaluation is strictly based on the facial expressions of patient and by comparing those facial expressions with the normal face of patient to calculate the accuracy, i.e., TFNI = sum of each FNI/number of indices. Similarly, the percentage of motor function of facial nerve can be calculated as follows:

$$\text{TPr} = \text{unhealthy TFNI}/\text{healthy TFNI [1]}$$

The greater the difference of two side of face, the severer is the disease and results in smaller ratio. Corner and salient point detection techniques are simple as well as efficient for facial paralysis grading.

2.2 House–Brackmann Grading System (HBGS)

It is a commonly used scale which is designed to systematically quantify the functioning of facial nerve recovery after trauma to the facial nerve or the Bell's palsy and treatments given to the patient. There is no recommendation made regarding the effect of physical therapy effect in Bell's palsy patients [2], and it is not intended as a sole source of guidance in managing Bell's palsy [2] also. Currently, no cause for facial nerve paralysis has been identified.

Causes for facial paralysis are as follows:

(a) Injury/stroke,
(b) Brain tumors, and
(c) Tumors of the parotid gland [2]

House–Brackmann grading system is to detect the severity of disease, i.e., normal to slight weakness, noticeable weakness to obvious weakness, asymmetry at rest which is the severe dysfunction, and no motion that is the total paralysis which is also indicated in Table 1. Their grading sections are based on the symptoms of facial expression which is further compared with rest of the symmetry on face.

House–Brackmann grading system (HBGS) is divided into six grades which are normal (HB-1 to HB-VI). HBGS provides easy grading of patient's condition [3]. HBGS can work on forehead, eye, and mouth in case of facial paralysis patients. But it can be more useful if it can be implemented in 3D images. The standard method of House Brackmann facial grading system was introduced in 1983 for measuring the facial nerve function. This method was strictly based on 2D and is considering the case of Bell's palsy which means half portion of face is affected. HBGS in terms of face includes only 3 section, i.e., forehead, eye, and mouth [3]. Even in grading system, a new technique is introduced called self-organizing map which consists of number of map neurons using nonlinear projection which are arranged in map using 2D [3].

Table 1 House Brackmann grading system [2]

House Brackmann facial grading system		
Grade		Defined
1	Normal	Normal in all areas
2	Mild dysfunction	Slight weakness
3	Moderate dysfunction	Noticeable weakness
4	Moderately dysfunction	Obvious weakness
5	Severe dysfunction	Asymmetry at rest
6	Total paralysis	No motion

2.3 Linear Measurement Index (LMI)

It includes Adour and Swanson Scale facial paralysis recovery profile and the Yanagihara grading system for facial palsy. LMI is the most widely used method in the facial paralysis detection for the patients who have gone through the treatment or surgery. It includes the rating scale that at what percentage the patients can be cured or uncured or uncertain. For calculating linear measurement index, facial anatomical landmarks are mandatory and are calculated by using displacement of number of facial anatomical landmarks during facial movement which is compared with response.

Barbosa et al. [4] introduced a new approach in terms of landmarks by combining optimized Daugman's algorithm on iris recognition by detecting the iris boundary and localized active contour model (LACM). As such, Daugman's algorithm utilizes the detected 4 key points that include supraorbital (upper eyelid), infraorbital, inner canthus, and outer canthus. This model is used to extract iris and facial landmarks or key points. Using Haar classification on 10 key points on the edge of face which includes eyes and mouth which helps to detect asymmetry.

Figure 4 shows the comparison of performance for facial palsy classification in terms of sensitivity, specificity, and harmonic mean. This technique of combination of Daugman's algorithm with localized active contour works on iris movement, and some portion is on edges of mouth. For detecting iris boundary, Duagmann's algorithm is used. Basic purpose is to detect the uneven portion of iris exposure when performing different eye movements like rising of eyebrow.

Figure 5 represents the comparison of previous research to new techniques. It considers only in the case of Bell's palsy where exact half portion of face is affected. In the eye region, it is clearly noticeable that based on key point approach, eye region is having lowest accuracy as compared with mouth and forehead. This better performance was achieved by the use of combined iris and LAC-based key point detection which yielded in better performance as compared to that of solely key point-based approach [4]. It reveals the difference between healthy side and the side which is affected by raising eyebrows where both eyes are directed upward.

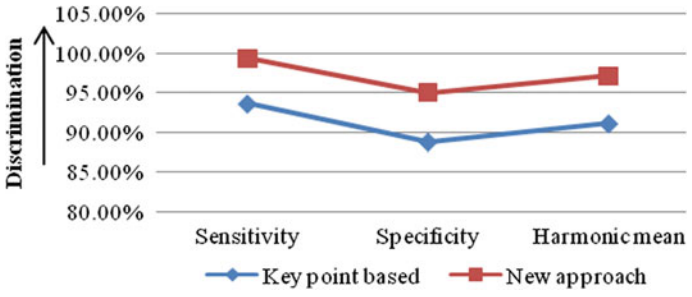


Fig. 4 Comparison of performance for Bell's palsy classification [4]

The limitation of linear measurement index is that it is a time-consuming process and it takes approximately 20 min for calculation.

Different levels of grading are represented by different colors in the graph, and in Fig. 6, the evaluation is based on Emergent Self-Organizing Map (ESOM) in case of Bell's palsy [3]. 46 different cases were studied out of which 21 images are found healthy having image resolution of 150 × 180, whereas the facial palsy having image resolution 691 × 1024 which will further proceed with a flowchart having the following procedure:

1. Image upload,
2. Noise removal,
3. Region of interest (ROI) rectangle,
4. Feature extraction,
5. Bell's palsy analysis, and
6. Diagnosis for grade level.

JohnVong et al. [2014] concluded that the use of Automated Grading System (AGS) significantly reduces the duration of medical treatment and will increase the consistency. It also concluded that Geometric Mean Measure (GMM) and F-measure maximize the accuracy, and hence, they are good indicators and achieve a classification rate of up to 95% in 2D region [3]. HBGS replaces other scales, but it does not allow one to accurate grade. In the year 2003, it is focused on four main areas on face: forehead, eye, midface, and mouth. The major drawback of HBGS is that it is not much reliable as compared to Sunnybrook grading scale. But still HBGS is widely used just because of its simple calculation of degree of facial palsy.

2.4 *Sunnybrook Facial Grading Scale (SFGS)*

This grading system works on the various parts of the face and on the bases of abnormality recorded, the score points are given, and then, final score was evaluated by the formula which is given below

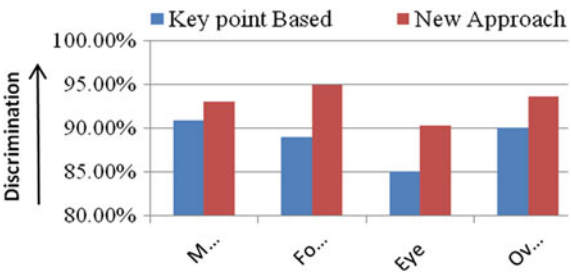


Fig. 5 Percentage of discrimination to different regions of face [4]

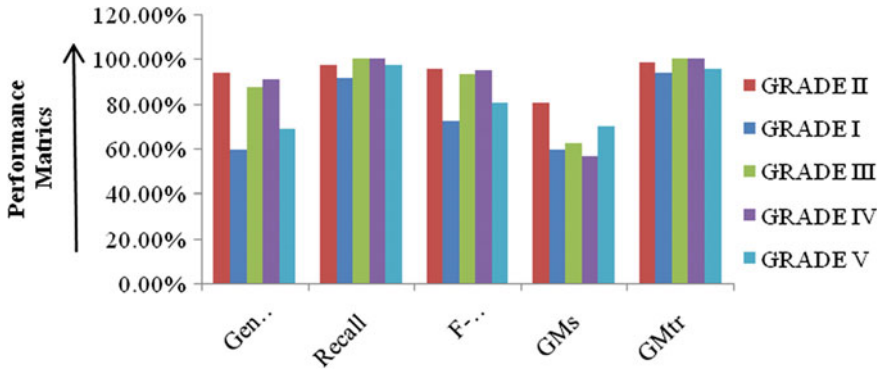


Fig. 6 Performance using emergent self-organizing map [3]

$$\textbf{Sum of resting symmetry points} * 5 - \textbf{Sum of voluntary movement points} * 4 - \textbf{Sum of skyninesis points} * 1 = \textbf{Final score which has maximum of 100}.$$

As the multiple areas of the face are scored from 0 to 5, therefore, it takes lot of time as compared to the HBGS.

3 Facial Paralysis

The facial paralysis has already been defined but as per the review of this facial paralysis, much work has been done in this field which is described below.

Saishanmuga et al. [5] proposed a method for personal identification which is strictly based on iris recognition by the use of neural networks and genetic algorithm. It concludes in better error finding by genetic algorithm and execution time.

Khanra et al. [6] invented new design of upper limb below elbow prosthetic patients. Electromechanical prosthetic arm was invented for paralyzed patients whose grasping power may vary, but it does not deal with facial paralyzed patients. Similarly, automated wheelchair is designed by Nashte et al. [7] which helps to make movement of handicapped person without using hands. Movement of wheelchair in left, right, forward, and stop is performed w.r.t. iris movements, but it will not help paralyzed patients to recover.

Paralysis may occur on any part of human body where there is weakness in muscles or dysfunctioning due to injury.

Spandana et al. [8] proposed a new algorithm for early detection of breast cancer with the use of **image processing**. It can be considered in paralysis cases also as it leads to high fluctuating asymmetry.

Birbaumer et al. [9] introduced brain–computer interface named thought translation device as an alternative communication channel for completely paralyzed patients with amyotrophic lateral.

4 Bell's Palsy

Bell's palsy is a type of facial paralysis which has already been explained in introduction. It includes simple steps for evaluation of the Bell's palsy. Issahar Ben Dov in 2012 introduced various symptoms and evaluation of diaphragmatic paralysis [10]. Bell's palsy patients and their respective treatment have been introduced. It is further concluded that in many patients no specific therapy is needed. The conclusion is strictly based on unilateral and bilateral paralyzed patients.

Reginald Baugh et al. in 2013 introduced the evidence-based framework for decision-making strategies so as to assist the clinicians and provide the only appropriate approach to diagnosing and managing this program of care.

Kleiss et al. in 2015 [11] proposed the method to validate existing health-related quality of life by questioning the patients.

Tiemstra et al. in 2007 [12] described a complete survey of Bell's palsy which further concluded that complete nerve palsy has lower rate of spontaneous recovery. It also results in facial asymmetry and muscular contractures which may require cosmetic surgical procedure or Botulinum toxin injections for which consultation with an ophthalmologist or cosmetic surgeon is needed.

Alberton et al. in 2006 [13] presented a review of treatment using antiviral agents to reduce the mobility associated with Bell's palsy. It concluded that randomized controlled trials are required before a definite recommendation can be made on use of antiviral agents.

FM. Sullivan et al. in 2009 presented the outcome for Bell's palsy in which 63% recovered patients with no treatment in 3 months which further increased to 85% after 9 months. This is indicated that acyclovir antiviral drug alone does not provide any benefit to Bell's palsy patients.

Aishwarya Balakrishnan in 2015 [14] presented the various symptoms of Bell's palsy by the use of electroneurography and also discussed that by adding antiviral to steroids does not provide any added benefit in achieving the recovery of facial muscles. Brenner et al. in 2004 [15] discussed computer-based approach to describe the muscle action units to evaluate the asymmetry in facial nerve branches. Gilden [16] proposed a study on Bell's palsy to determine the various factors causing facial weaknesses.

5 Conduction Velocity for Paralysis Patients

Nerve conduction velocity is common technique to evaluate the sensory nerves of human body. For conduction velocity, electrodes will be taped on to cleaned skin. Small electric current or stimulus is applied over there. Then, these electrodes will measure the current traveling down the nerve pathway. If nerve is damaged, then the current will be slower or weaker. For this procedure, anesthesia is not used. This nerve conduction velocity test is sometimes combined with electromyography, i.e., EMG. This can be calculated as per the formula given below

$$\text{Conduction Velocity} = \text{Distance}/\text{Time}. \quad \text{Or}$$
$$\text{Conduction Velocity} = \text{Distance}/\text{Onset latency}.$$

Or

$$\text{Conduction Velocity} = \text{Distance (cm)}/\text{Latency 2}-\text{latency1(msec.)} \times 10$$

[17] Conduction velocity is used to record the responses and further processed to identify the % of damage of the nerve.

Figure 7 shows the basic procedure for conduction velocity which is performed by various electrodes based on the facial regions. The procedure for conduction velocity includes three main steps:

- **Active electrodes**: These are placed on the center of muscle belly on human body. These electrodes receive the current which is passed via stimulator.
- **Reference electrodes**: These electrodes are placed about 3–4 cm away from the active electrodes.
- **Stimulator:** This is placed over nerve that supplies the muscle. With the help of stimulator, electric current is passed via nerve to muscle which is recorded on a graph in the form of waves to determine nerve damage. Current is different in case of motor nerve conduction and sensory nerve conduction. 20–50 mA current is required by stimulator in case of motor nerve conduction, whereas 5–30 mA current is required by stimulator in case of sensory nerve conduction.

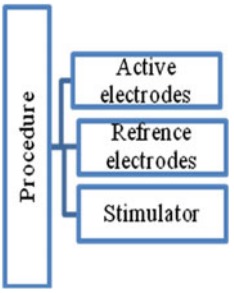


Fig. 7 Classification of electrodes during conduction velocity

With the use of electric current, each node shows its response in the form of waves which has been recorded as shown in Fig. 8, and during conduction velocity, different outcomes have been recorded with different values. There is a particular range for this conduction velocity; e.g., range of normal conduction velocity was 48.8 ± 3.68 (mean ± SD) m/s.

The waveform of one electrode from which current is passed with the help of stimulus is shown in Fig. 9.

By noticing this waveform, we can easily calculate amplitude and latency. Amplitude reflects the number of muscle fibers that are depolarized due to passage of stimulus. Amplitude is measured from the baseline to the negative peak.

Conduction velocity can be calculated by dividing distance with the onset latency. In case of motor nerve conduction value in facial region, there is constant value of amplitude and onset latency, i.e., for amplitude, value should be 1, and for maximum onset latency, value should be 4.0.

Latency is defined as the time from stimulus to the initial negative deflection from the baseline of waveform. Latency is measured in milliseconds. There are two types of latency as shown in Fig. 10:

Latency varies directly with the distance of stimulating electrodes from muscles.

a. ***Onset latency***: It is defined as the time required for electrical stimulator to initiate potential.
b. ***Peak latency***: It represents the latency along with the majority of axons. It is measured as the peak of waveform amplitude that means the first negative peak which is being recorded by electrodes in terms of waveforms. Peak latency cannot be used to calculate the conduction velocity.

Conduction is obtained by stimulating nerve at two different points along the nerve at least 10 cm apart [17].

In sensory nerve action potential, two types of latency can be calculated:

1. Onset latency and
2. Peak latency

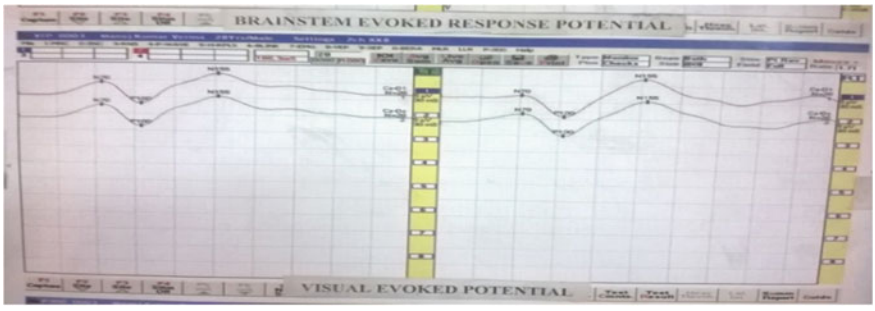


Fig. 8 Recorded waveforms using electrodes in case of brain stem

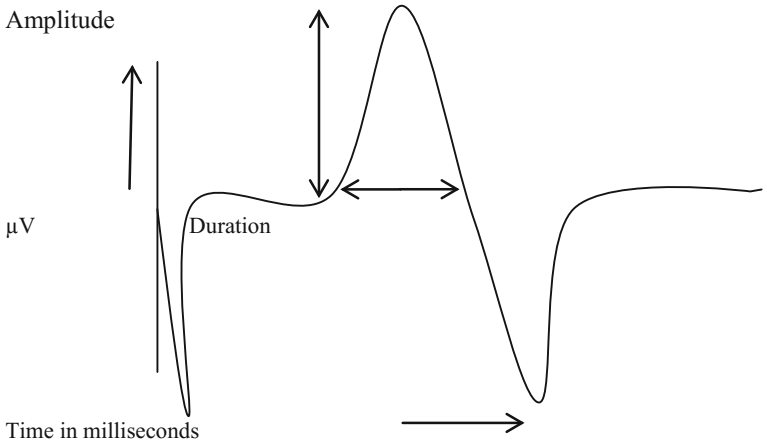
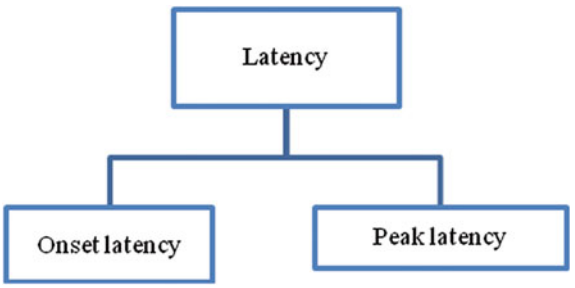


Fig. 9 Compound motor action potential

Fig. 10 Types of latency in terms of waveform



But peak latency is not useful in case of conduction velocity of facial paralyzed patients. Low sensory nerve action potential indicates a definite disorder of nerve. It is measured as the first negative peak in the waveform responses which is recorded with the help of electrodes.

The waveform is recorded with the help of electrodes which are placed on the nerve. The onset latency is defined as the time required for electrical stimulation. Drop of 50% is considered normal when recording a proximal sensory nerve action potential (SNAP). The existence of initial positive deflection is due to the inappropriate placement of active electrodes from the motor point. Abnormal results are recorded because of the nerve damage due to which the waveform will not be appropriate. The nerve damage includes the following:

1. Conduction block: blockage of nerve pathways.
2. Damage to nerve axons.

Main part of this section depends on the value of conduction velocity. If it is lower than the range, i.e., 48.8, then current is somehow not passing properly, or muscles are not working properly. The waveforms are used to calculate the

amplitude of each electrode. These electrodes represent all muscle where the electrodes have been placed on patient's body. The recording electrodes are placed over the muscles with the active electrode over the belly of muscle and the reference electrode is at least 3 cm distal to active electrode over distal muscle. Generally, electrode of stimulator is placed over the wrist. The active electrode is placed directly below the pupil with the patients looking straight ahead in case of facial paralysis.

Robinson et al. in 1993 concluded that difference in gender in case of nerve conduction velocity varies with height, temperature, as well as age of the person. It is observed that nerve conduction velocity will decrease with increase in age. Facial nerve splits into number of branches and supplies the muscles of the face. One of these branches is just below the eye that is why electrode is directly placed below the pupil.

Albers et al. from Michigan [17] concluded that the conduction velocity of facial nerve can be calculated by putting electrodes on nose as well as below the eye, and these can be grounded on forehead or on wrist as well. As in Fig. 11, it is very much clear that stimulator has been applied on the 7th cranial nerve which is just behind the ear of human body. Electric current is passed through the stimulator, and the waveform is received w.r.t the electrodes that is to show how much current is received on nose as well as below the eyelid. Using same concept and technique, it is observed that lesser the current results to the asymmetry of the face which clearly shows that there is damage to the nerve due to which the current is not passing through the nerve properly.

Two pairs of electrodes are used in the case where there is comparison required between two halves of facial regions. But it is implemented only in case of Bell's palsy and not in facial palsy patients.

Albers et al. [17] also introduced the blink reflex. Two pairs of electrodes are used in such a way that reflex responses are recorded simultaneously as shown in Fig. 12. The active recording electrodes are placed on oculi muscles bilaterally.

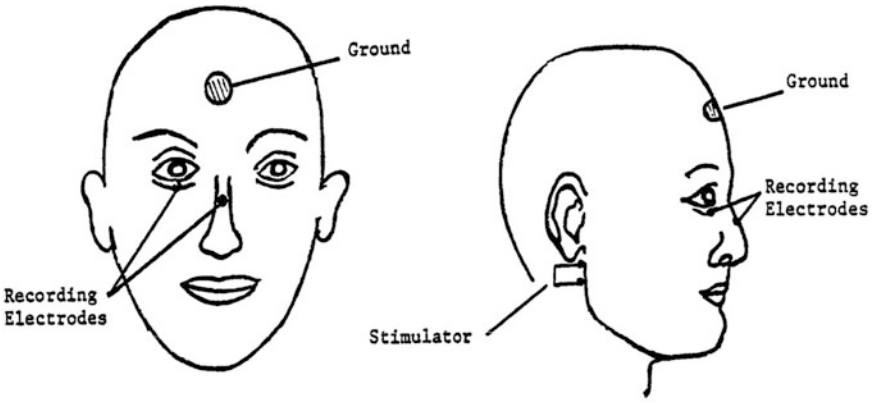


Fig. 11 Use of single pair of electrodes [17]

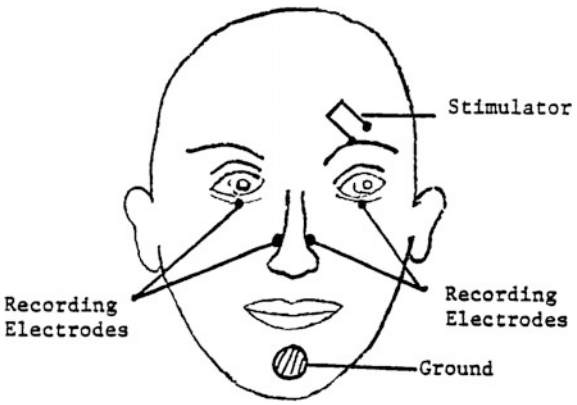


Fig. 12 Use of two pairs of electrodes [17]

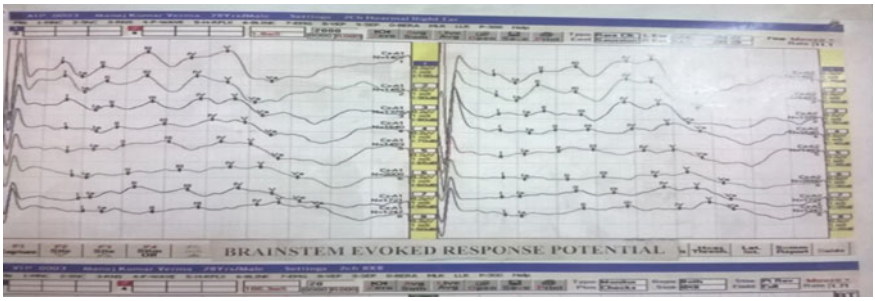


Fig. 13 Recorded waveforms when number of electrodes used

The inactive recording electrodes are placed on the lateral side of nose. The ground is placed between active electrode and reference electrode, and rest of the process is same; i.e., with the help of stimulus, electric current is passed via electrodes and amplitude can be calculated.

Various wave impulses are shown in Fig. 13; these waves represent the electrode outputs which are placed on paralyzed patients just to calculate the conduction velocity.

5.1 Compound Motor Action Potential (CMAP)

It also consists of waveform, but peak latency is not required in Compound Motor Action Potential (CMAP).It represents three basic processes:

- **Nerve conduction time**: It can be calculated with the help of electrodes. It is defined as the average time taken by the nerve to respond when some electric

current is passed via stimulator. It is the ratio of the time taken to pass the electric current from nerve to the muscle.

- **Delay in time across motor nerve**: It is based on asymmetry due to nerve dysfunctioning. Delay in time across the motor nerve is only due to facial paralysis which occurs due to nerve damage or dysfunctioning.
- **Depolarization of nerve**: Muscles play a great role in body movement. The permeability of ions Na^+, k^+, and Cl^- plays a great role in the conduction of the current in the nerve fibers by the method depolarization and repolarization.

In the stimulator, the current required is 20–50 mA for the nerve conduction.

CMAP is measured between the baseline and negative peak from the waveform. Conduction velocity can be calculated by dividing the change in distance by change in time. Drop of 15% is considered normal when recording a proximal compound motor action potential.

5.2 Sensory Nerve Action Potential (SNAP)

SNAP includes the same feature as in peak latency. The major difference is in waveform as it indicate greater drop as compared to the drop in CMAP.

Nerve conduction shows responses in impulse form which differ from nerve to nerve. There is no single normal value existing for all the individual regions on face as it depends upon various factors such as height and temperature. Nerve conduction velocity is directly proportional to the nerve diameter and myelination degree. It is also observed that sometimes responses may be noticed as normal even if there is nerve damage, and thus, to overcome this clause, **image processing** is used for same. By using image processing and comparing the two halves of face, asymmetry can be easily justified. At some extent, nerve conduction velocity is helpful because it is used to detect nerve damage or nerve disorder. It also helps to detect the disorder and condition where there is injury in nerve which results in dysfunctioning of muscles.

During nerve conduction velocity, the body temperature should not be too low or too high. If the temperature of body is low, then nerve conduction will slow down the nerve conduction. Similarly if body temperature is high, then it results in faster nerve conduction, due to which exact responses cannot be recorded. In case of nerve fiber damage, responses in terms of impulse show a decreased impulse but in normal speed of electric current from stimulator. It is strictly based on HBGS grading system because HBGS mainly includes the forehead, nose, mouth, and chin. These areas are covered with electrodes as shown above in Fig. 14. Here, in this case, electrodes are placed on forehead, and also hence, these can be grounded on the wrist of the patient. But the stimulator or electric current is given on the facial nerve, i.e., seventh cranial nerve so as to record the various waveform w.r.t current. By calculating amplitude as well as latency, we can only identify the weakened nerve on face which can be the branch of facial nerve. This depends only

on the current value which is actually passed w.r.t. the current received on the surface where electrodes are placed. Dilip in 2010 concluded that gender has a definite effect on nerve conduction variables. It is also observed that in case of CMAP, there is longer latency in terms of amplitude, and in case of males similarly in SNAP, there is longer latency in males, whereas higher amplitude is recorded in case of females.

The impulse waveform of right ear of human body is clearly shown in Fig. 15. Highest as well as lowest waveform is identified by number as I, II, III, IV, and so on. There is one technique which only works for four facial expressions. There is use of mouth edge position and lateral canthus, and disease level was predicted [18].

IECM algorithm is a new approach to find the disease level at some extent. It includes eyebrow rise, eyebrow closed tightly, screwing nose, and mouth opened. It will not work on chin, forehead, and cheeks which is the major part in paralysis. Below graph is the upgradation in detecting the disease level as compared to existing one as shown in Fig. 16. But the accuracy can be more if more expressions are used in same technique.

The average value of proposed method increased due to upgradation in the value when eyebrow is closed tightly. Rest of the regions which have been taken during IECM is approximately same which is clearly shown in graph.

Many techniques are used for paralyzed patients to help in for paralysis detetction at some extent, but it will not help in recovery from paralysis. As in HBGS, it is clearly shown that higher the grade of paralysis, less chances of recovery are there.

There are different techniques used for evaluating the degree of paralysis, but the important part for this work is carried out only for Bell's palsy patient but not for facial palsy patients. We can calculate the level of disease in patients which are

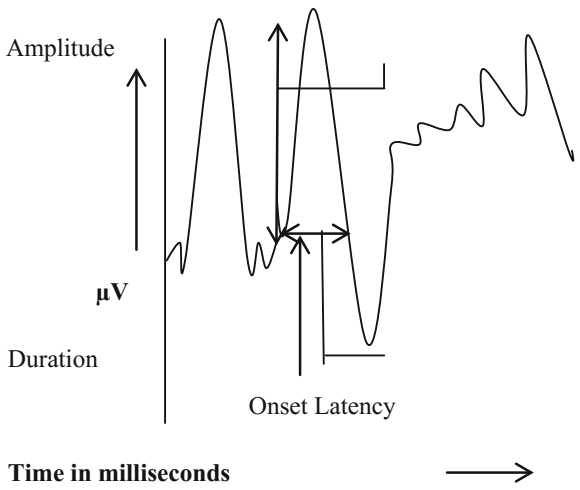


Fig. 14 Sensory nerve action potential

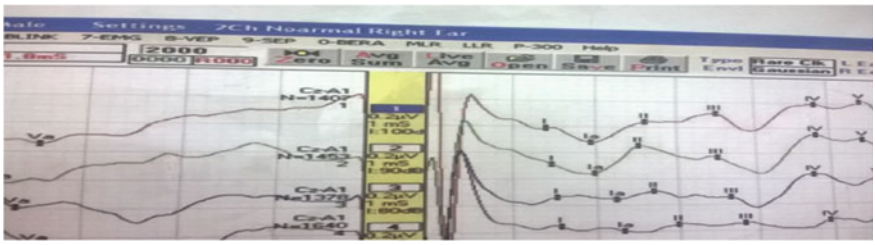


Fig. 15 Waveform of right ear of human body

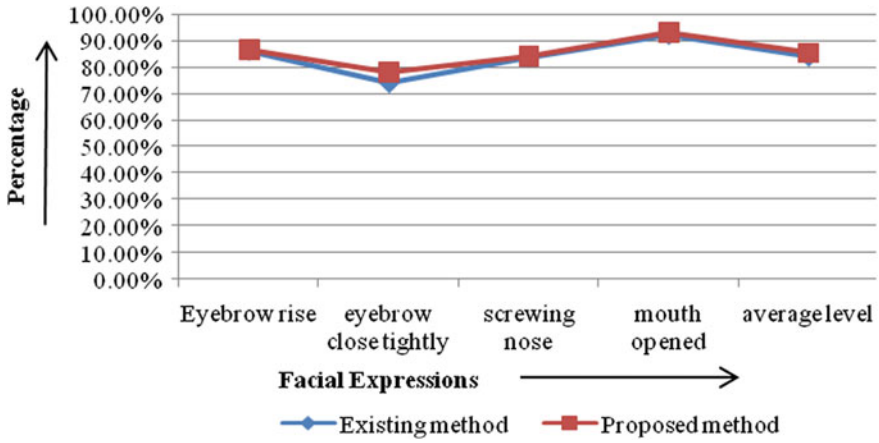


Fig. 16 IECM algorithm results in upgradation [18]

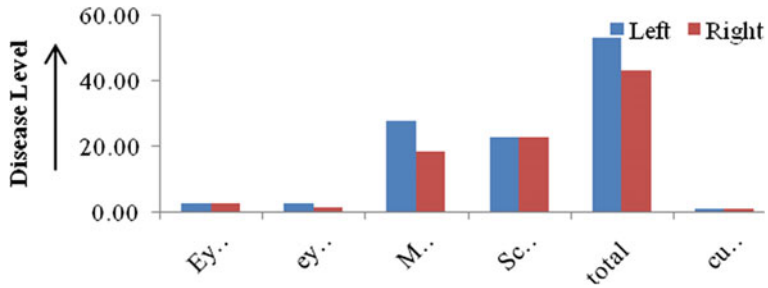


Fig. 17 Level of disease in patient suffering from Bell's palsy [18]

suffering from Bell's palsy. Figure 17 shows the comparison, whereas Fig. 18 shows the level of disease using IECM algorithm.

Above graph shown in Fig. 17 is the disease level of patient having facial paralysis. It is based on the Bell's palsy where exact half portion of face is affected

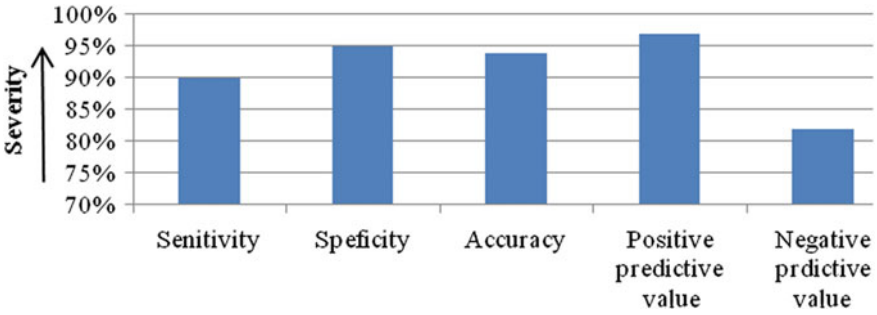


Fig. 18 Degree of facial paralysis [19]

and the basic concept to calculate the degree of facial paralysis is to compare the face of affected side to the normal side of face. Greater the difference, more is the paralyzed area, and smaller the values during comparison of both side of face indicates lesser the paralysis, which can recover easily. As in graphical representation, it is clear that left side of face of patient is affected or paralyzed as compared with right side of face.

Anurag et al. [19] introduced salient point selection algorithm to evaluate the degree of facial paralysis. Different facial expressions are taken during implementing salient point selection algorithm to compare the expression of both right and left sides of the face. Here, Wiener filtering is used for noise removal under preprocessing phase. Salient point selection algorithm also deals with same concept to calculate the degree of evaluation by taking ratio of affected side of face by unaffected side of face. Anurag et al. [19] proposed the degree of accuracy in evaluating the disease as well as categorization.

Nerve passes electric current with the help of stimulator as nerve is like electric wire which passes electric current from one end to another end of human body. This electric current can be recorded in terms of impulses or waveforms. Impulses may show some variation, and these variations can be due to nerve damage. It is not mandatory that whole impulse can be affected, but a small passage of impulse gets affected in case of nerve damage. There are various conditions on which nerve conduction velocity depends. Some common conditions which will affect nerve conduction velocity are age, upper or lower limb, temperature, etc. [20]. There cannot be any universal reference value for particular nerve which can be taken into consideration.

The conduction velocity is evaluated only in case of elbow and axilla. Similar techniques can be used in terms of facial paralyzed patients. Procedure for elbow nerve conduction velocity is similar as in conduction velocity of facial palsy, but the difference may vary in terms of electrode pairs used. Common fields which are taken in case of elbow are latency and amplitude. Standard deviation of elbow conduction velocity with respect to mean value of elbow conduction velocity is evaluated and shows a parallel variation of mean as well as standard deviation as shown in Fig. 19.

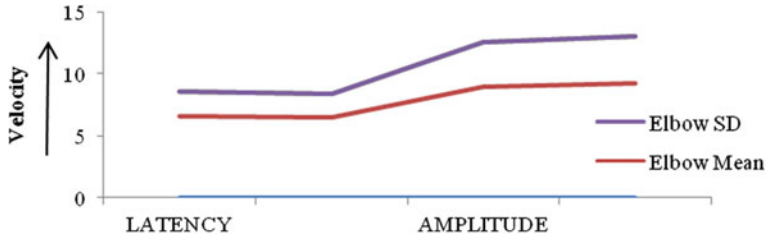


Fig. 19 Radial motor nerve of elbow [20]

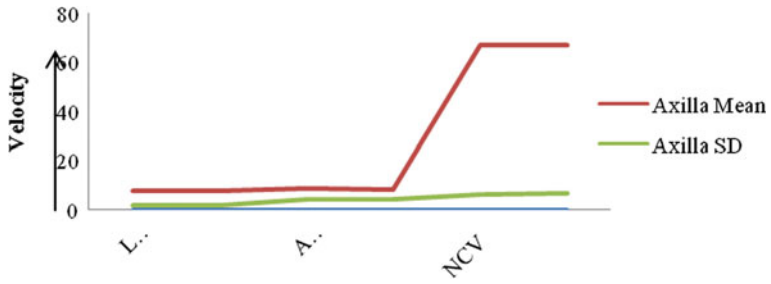


Fig. 20 Radial nerve motor study of axilla [20]

Radial nerve includes both sections of normal motor and sensory nerve conduction fibers. There must be nerve conduction velocity values of different regions on human body. In case of elbow, there is no recording of data value in case of nerve conduction velocity which is measured in meter per seconds. Latency is measured in milliseconds, and amplitude is measured in millivolt.

Here, in Fig. 20, the procedure to calculate the latency and amplitude and NCV of axilla is same as in elbow. But in this, mean value of axilla is higher as the standard deviation value of axilla is more. The values are not varying parallel. There is complete implementation for evaluating elbow, axilla, and wrist. There is no constant value or universal value for evaluating every region of human body in terms of nerve conduction velocity [20]. The implementation is only on adult age-group of 17–20 years. Here, radial nerves of both the right and left hand were compared for latency, amplitude, and NCV. There is not much difference between left and right sides of wrist, elbow, and axilla as well. Mean value of nerve conduction velocity of radial motor nerve is not constant throughout as in Fig. 21.

There is no significant difference between latency and amplitude of right and left arms, respectively. Also in case of conduction velocity of axilla, there is no significant difference ($p > 0.05$) [20].

The limitation of the conduction velocity of elbow, axilla, and wrist is that it considered only the age-group of 17–20 years. Hence, it is not sufficient for comparison between other age-groups [20].

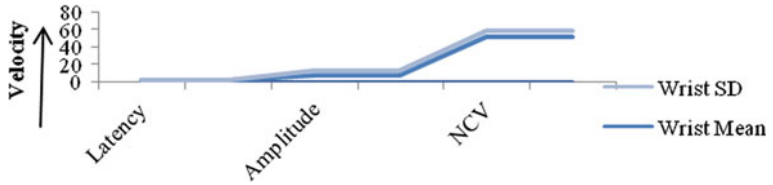


Fig. 21 Sensory conduction velocity of wrist [20]

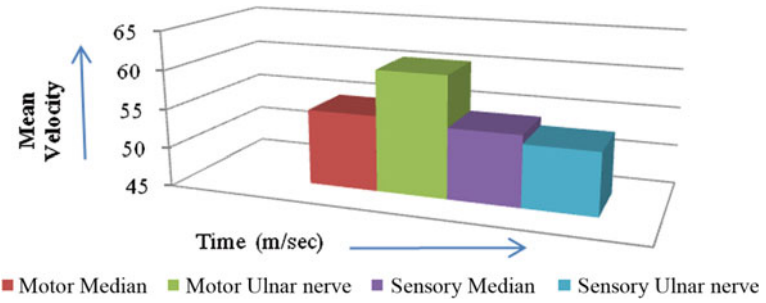


Fig. 22 Mean velocities [21]

Aging plays an important role in terms of detecting the level of paralysis. In case of facial paralysis patients, maximum cases are present in old age. Aging results in physiological changes in human body. This physiological change occurs due to some dysfunctioning of neuromuscular junction in neuromuscular and reduction in nerve conduction velocity, or it may be due to slow muscle response. It directly depends upon aging factor. It is not mandatory that facial paralysis depends on the age of the patient except the flaccidity of muscles which decreases with age. In some cases, adults suffer from facial palsy, but old-age human beings are found normal. Thus, it is easy to say that just because of nerve damage or nerve dysfunctioning, it leads to facial palsy. Every human body has its own strength and own muscular power. It depends on the muscles and nerve that up to what extent it will respond positively. But 80% cases depend upon the aging factor. 80% cases are found that increase in age results in muscles weakness. Due to muscle weakness, there may be some asymmetry between two halves of face.

Nerve conduction study consists of 3 types:

(a) Motor nerve: It includes CMAP.
(b) Sensory nerve: It includes SNAP.
(c) Mixed nerve.

It is observed that conduction velocity is 6 m/sec faster in females. Mean velocities are evaluated in terms of median and ulnar nerves of arms [21] as shown in Fig. 22.

Mohamed in 2007 shows the evaluation of both motor and sensory nerve in terms of their median and ulnar nerves. There is not much difference between median value of motor and median value of sensory nerve, whereas in case of ulnar nerve, there is noticeable variation between ulnar nerve motor and ulnar nerve sensory potential. Nerve conduction velocities vary with temperature. There is a relation between body temperature and nerve conduction velocity, i.e., nerve conduction velocity response changes with change in the body temperature of patients. Different value of temperature shows different results in nerve conduction velocity.

Here, median value for both motor and sensory nerves is 54.71 ± 5.69 and 54.04 ± 7.02 m/s, whereas ulnar value in case of motor and sensory nerves is 60.57 ± 5.00 and 52.92 ± 5.89 m/s.

Different implementations result in different outcomes. Common observations include one major factor that is declined in muscular performance with advancing age at some extent. Likewise, Henriksen in 1956 concluded that there must be some variations in nerve conduction velocity with temperature. Temperature effects are from 12 to 40 °C on motor conduction of human nerve [21]. Between 29 and 38 °C, the motor conduction velocity decreased 2.4 m/sec. for each 1 °C decrease in temperature [21].

6 Image Processing for Paralysis Detection

Image processing techniques have been widely used for evaluating the degree of paralysis. In case of image processing, different researchers have their different opinions in terms of noisy images, textures, and evaluation in degree of paralysis.

Gutte et al. in 2014 [22] sclassify textures as static and dynamic with use of local binary pattern features. It works with invariant textures. Image processing techniques play a great role in detection of degree of paralysis for the treatment of paralysis patient.

Kashyap et al. in 2011 [23] introduced Wiener filtering technique to remove the noise and mutation-based bacteria.

Facial paralysis makes a very critical position to human, as shown in Fig. 23 Left side of face is affected. As it is the case of Bell's palsy, hence eyes, nose, cheeks, and lips are equally affected, but it works on collected images, and only 4 facial expressions are used for conclusion. More facial expressions can be added to improve accuracy.

Anurag et al. [19] worked on evaluating the degree of paralysis. Salient point selection algorithm (SPSA) is presented on different facial expressions. Severity of the paralysis with corresponding images is used to calculate the degree of paralysis. Prasad et al. in 2003 [24] presented the facial expression analysis from image sequences, and their synthesis was presented. Facial animation for video analysis and synthesis has been implemented.

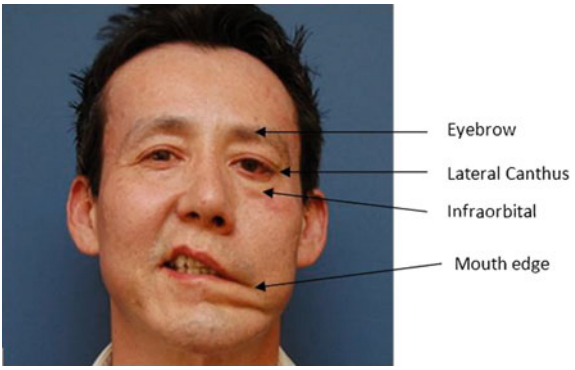


Fig. 23 Facial paralyzed image [18]

Scokin et al. [25] proposed a system that discriminates between patient with facial nerve palsy and normal subjects using only 3 expressions (resting, smiling, raising eyebrow). Active appearance models (AAMs) for localized facial landmarks were good for healthy subjects, and Incremental Parallel Cascade of Linear Regression algorithm is used for paralyzed patients. AAMs and Incremental Parallel Cascade of Linear Regression algorithm are two image processing techniques using 1080×1920 resolution which was well implemented by Hyun Scokim et al. Here, atomic grading system is divided into image or video based. Total 49 facial landmark points have been chosen for evaluation.

Samsudin et al. [26] concluded that image processing has been used widely in many areas but is limited in facial rehabilitation area. Furthermore, the development of accurate method for detection and grading facial nerve functioning will be very challenging, but it will give a lot of benefit to facial paralyzed patients.

Ater et al. in (2014) [27] conclude an outcome of a clinical observational study together with algorithmic development. Here, nine facial expressions were considered for developing mobile application for diagnosis and grading facial nerve palsy. Algorithm demonstrates an excellent binary classification and high correlation to facial nerve palsy grading methods.

Tang et al. [28] introduced a toolbox which includes fast operation with extensive applicability and concluded that time used in visualization is greatly reduced as interactive data language (IDL) supports for open hardware accelerated graphics.

Rao et al. [29] introduced a novel image enhancement technique which is based on M band wavelet. Basic work is to generate medical image enhancement algorithm which will generate high-resolution denoised good contrast images. It has been tested on various medical image processing techniques such as X-rays, MRI, CT, and OCT (optical coherence tomography). Interface is designed to browse the raw images and to plot the enhanced images along with other utilities.

Sadek et al. [30] proposed a new perceptual forensic singular value decomposition (SVD)-based approach. A logarithmic transformation has been used to develop a new prospective vision in utilizing the SVD properties. Daniel Goldberg

et al. [2006] proposed image processing technique to better distinguish white matter disruption which is helpful in medical images [30].

Image processing is also helpful in medical science for detecting the paralysis area and comparison between various techniques used for paralysis detection using image processing have been presented in Table 2.

Table 2 Comparison between various techniques used for initial paralysis detection and their outcomes

S No	Methodology	Outcomes
1.	1. Use of Harris corner detector, SUSAN edge detector 2. Use of K-means clustering algorithm	1. Sample size is taken as 100 patients 2. Frontal viewed human faces are considered 3. Standard size = 200 * 250 4. Accuracy = 95% 5. It will calculate certain distances to estimate the state of facial paralysis
2.	1. Electromechanical prosthetic arm 2. AT89C51 minus controller and L23D motor driver are used	1. It also brings back muscle movement to paralyzed patients' hand and reduces muscle atrophy
3.	1. Eye-Asseto glasses worn by a user to detect the eyeblinks [31]	1. User can operate the computer through eyeblinking.
4.	1. Image preprocessing 2. Recognition techniques 3. Testing and classification of emotions 4.Mutation Bacteria Foraging optimization	1. Image is free from speckle noise by use of mutation-based bacteria foraging optimization technique 2. Mean and variance = 0.05 3. Result shows noises are not removed properly so for removing the remaining noise, mutation-based bacteria foraging
5.	1. PC MATLAB environment 2. PSoC microcontroller 3. DC motor	1. Movement of wheelchair in left, right, forward, and stop by moving iris position in left, right, straight, and closing of eye, respectively
6.	1. Use of IECM algorithm 2. Preprocessing of images and estimation of level of disease	1. Sample size is taken as 3 person 2. Average level calculated in existing method = 83.6% Proposed method = 85.38% 3. The disease level is predicted
7.	1. Image enhancement 2. Segmentation of masses 3. Extraction of border of mass	1. Sample size is taken as 44 patients out of which 4 patients have high fluctuating asymmetry which can be easily indication for breast cancer 2. Patients who have speculated masses have high fluctuating asymmetry 3. AHE gives desirable level of enhancement, i.e., EBCM value of AHE is lower than that of HE and higher than that of original for given all types of mammogram images

(continued)

Table 2 (continued)

S No	Methodology	Outcomes
8.	1. House–Brackmann facial nerve grading system 2. Numerous diagnostic test has been used to evaluate patients with acute facial paralysis	1. It results in survey as in medical treatment and medicines
9.	1. Training and testing phase are used 2. Salient point selection algorithm	1. Sample size is taken as 9 images 2. sensitivity = 90% 3. Accuracy = 94% 4. Positive prediction value = 97%
10.	1. Corticosteroids treatment used to reduce inflammation 2. Prednisolone treatment [32]	1. Result shows complete recovery and synkinesis at 12 months in relation to severity of palsy at baseline
11.	1. Oral corticosteroids 2. Acyclovir 400 mg can be given	1. Facial asymmetry and muscular contractures may require cosmetic surgical procedure or botulinum toxin (Botox) injections
12.	1. Acyclovir monotherapy was shown to be inferior to prednisone monotherapy	1. The evidence reviewed would favor the combination of acyclovir and prednisone if commenced within the first 72 h of symptoms onset 2. Result is shown in terms of medical treatment
13.	1. 2 * 2 factorial randomized double-blind trial [33] 2. Patients divided in Four groups: ● Prednisolone + Acyclovir ● Prednisolone + Placebo ● Acyclovir + Placebo ● Placebo + Placebo	1. Sample size is taken as 752 patients 2. 132 patients is ineligible 3. 551 out of 620 patients are eligible were randomized 4. 55 patients dropped out of study before final determination of House–Brackmann status. 5. Thus, final outcomes were for 496 patients.
14.	1. Electroneurography 2. Meta-analysis to determine whether steroid treatment plus antiviral provides better degree of facial muscle recovery than does steroids alone	1. It only deals with survey 2. No simulation is implemented
15.	1. Model-based codec is presented that is able to encode head and shoulder video sequences at bit rate of 1 Kbit/s	1. Gabapentin is used to treat TN and is effective in approximately 40% of cases 2. Operating microscope is used to move the offending vascular loop and secure the TN
16.	1. Linear discriminant analysis (LDA) 2. Support vector machine (SVM) 3. Leave one out cross-validation (LOOCV) 4. Use of incremental training of discriminative models	1. Accuracy rate was 88.9% 2. Images were resized to 540 * 960 to reduce processing time 3. The precision and recall rate were 92.3 and 90.0%, respectively

(continued)

Table 2 (continued)

S No	Methodology	Outcomes
17.	1. 2D and 3D measurements on images or videos	1. If value of P degree is larger than zero, the subject is declared to have facial paralysis 2. The correct recognition rate can reach 100%
18.	1. An algorithm introduced based on analyzing a set of 9 facial expressions	1. Sample size is taken as 13 predefined facial landmarks 2. accuracy = 95.5% 3. Result also includes addition of separate grading of each facial area
19.	1. Images from Mater Misericordiae health service in Brisbane 2. Use of House–Brackmann grading system	1. Sample size is taken as 46 facial palsy images and 21 normal face images 2. Resolution = 691 * 1024 and 150 * 180, respectively 3. Images are graded manually into five scales using House–Brackmann grades 4. ROC curve of Hamming distance with SVM which gives better results of about 0.81
20.	1. Additional data needed to determine the combination of antiviral and corticosteroid alone	1. The density of facial nerve in Bell's palsy is probably decided within the first two to three weeks after onset of symptoms
21.	1. Based on IDL 6.4 with common configured computers	1. Time used in visualization is greatly reduced as IDL supports for open GL-based hardware accelerated graphics
22.	1. MSE and PSNR are taken for comparison of enhanced images generated	1. Sample = 2 cases are considered 2. MSE decreases drastically when we use proposed method using two-stage decomposition
23.	1. New perceptual forensic SVD-based approach is proposed which is based on global SVD 2. Logarithmic transformation	1. Objective quality measure by using NMSE values for developed and proposed techniques is 0.0223 and 8.8058e−009, respectively
24.	1. 84-feature points (FPs) defined on a neutral face 2. Focus is on the recognition of spontaneous expression 3. Hidden camera to record subject expression [34]	1. It is the theoretical portion which is survey

7 Conclusion

This paper is used to present the detailed survey of various paralysis detection techniques and methods in use, and the outcome of this survey is an active appearance model which can be useful to achieve better performance in terms of measuring facial expression. IECM technique is helpful to evaluate the degree of recovery of facial expression, but it can be more accurate if more expressions are taken in IECM technique. In terms of medical science, tablet acyclovir can be helpful as an antiviral drug, but it alone has not much benefit but it has more effect when tablet prednisolone is added with it. Analysis in terms of facial paralysis images has been in 2D but not in 3D. For upgradation and to evaluate the accuracy as well as efficiency, 3D can be used in facial paralysis as much work has been done in 2D. House–Brackmann grading system is very helpful as it includes simple method for calculation and evaluates the degree of facial paralysis. The number of technologies has been in the category of Bell's palsy but not in the facial palsy which does not include bilateral face but includes a part of facial region.

References

1. Dong J, Wang Q, Wang S, Liu LA (2011) Evaluation of facial paralysis degree. College of information science and engineering, Ocean university of China
2. Baugh R, Ishii L, Schwartz SR (2013) Clinical practice guideline Summary Bell's palsy. AAP-HNS Bulletin
3. Song I, Vong J, Yen NY, Diederich J, Yellowlees P (2013) Profiling bell's palsy based on House-Brackmann score. JAISCR 3(1)
4. Barbosa J, Lee K, Lee S, Lodhi B, Cho JG, Seo WK, Kang J (2016) Efficient quantative assessment of facial paralysis using iris segmentation and active contour based key point detection with hybrid classifier. BMC Med Imaging
5. Saishanmuga Raja V, Rajagopalan SP (2013) IRIS recognition system using neural networks and genetic algorithm. Int J Comput Appl 0975-8887 68(20)
6. khanra D, Sudesh S (2011) Below elbow upper limb prosthetic for amputees and paralyzed patients. Int J Comput Appl 0975-8887 16(5),Vellore, India
7. Nashte AA, Savkare SS, Kole SG (2016) Automation of wheelchair using Iris movement. IJCER 6(03)
8. Spandana P et.al (2013) Novel image processing techniques for early detection of breast cancer MATLAB and lab view implementation. IEEE Point of care health care
9. Birbaumer N, Kubler A, Ghanayim N, Hinterberger T, Perelmounter J, Kaiser J, Iversen I, Kotchoubey B (2000) The thought translation device (TTD) for completely paralyzed patients. IEEE Trans Rehabil Eng 8(2)
10. Issahar B-D (2012) Diaphragmatic paralysis-symptoms,evaluation,therapy and outcome, Israel
11. Kleiss IJ, Beurskens CHG, Stalmeier PFM (2015) Synkinesis assessment in facial palsy validation of Dutch synkinesis assessment questionnaire. Acta Neurol Belg. Department of Otolaryngology and Head and Neck Surgery, Radboud University Medical Center, PO Box 9101. Ingrid J.Kleiss ingrid.kleiss@radboudumc.nl Published in springer 171–178
12. Tiemstra JD, Khatkhate N (2007) Bell's palsy: diagnosis and management. Am Fam Physician 76(7)

13. Alberton DL, Zed PJ (2006) Bell's Palsy: a review of treatment using antiviral agents. Ann Pharmacother 40:1838–1842
14. Balakrishnan A (2015) Bell's palsy: causes, symptoms, diagnosis and treatment. J Pharm Sci Res 7(11)
15. Michael J, Brenner J (2004) Gail neely.:approaches to grading facial nerve function. vol 18, No 1
16. Gilden DH (2004) Bell's palsy. New Engl J Med
17. Albers James W (1995) Emeritus professor of neurology: nerve conduction manual. University of Michigan, Michigan
18. Anguraj K, Padma S (2012) Analysis of facial paralysis disease using image processing techniques. Int J Comput Appl 54(11)
19. Anguraj K,Padma S (2015) Evaluation and severity classification of facial paralysis using salient point selection algorithm. Int J Comput Appl 0975-8887 123(7)
20. Chouhan S (2016).Normal motor and sensory nerve conduction velocity of radial nerve in young adult medical students. J Clin Diagn Res 10(1)
21. Jagga M, Lehri A, Verma SK (2011) Effect of aging and anthropometric measurements on nerve conduction properties-a Review. J Exerc Physiotherapy 7(1)
22. Gutte PH, Kharat PK (2014) Description of rotation-Invariant textures using local binary pattern features. Int J Comput Appl 0975-8887 99(9)
23. kashyap KL, Shantaiya S (2011) Noise removal of facial expression image using wiener filter. In: IJCA-national conference on emerging trends in CSE and IT
24. Prasad S, Galetta SL (2009) Trigeminal neuralgia. In: Encyclopedia of lifesciences. Wiley, New York
25. Scokim H, Kim SY, Kim YH, Park KS (2015) A smartphone based automatic diagnosis system for facial nerve palsy.
26. Samsudin WSW, Sundaraj K (2012) Image processing on facial paralysis for facial rehabilitation system: a review. In: IEEE international conference on control system, Penang, Malaysia
27. Ater Y, Gers L, Bryt O, Halperin D (2014) Mobile application for diagnosis of facial palsy
28. Tang M, Chen F (2013) Three dimensional visualization toolbox for medical image based on IDL. Int J Sig Process 6(5)
29. Rao Y, Sarwade N, Makkar R (2015) Denoising and enhancement of medical images using wavelets in LAB View. Int J Image Graph Signal Process
30. Sadek RA (2012) SVD based image processing application: state of art, contributions and research challenges. IJACSA 3(7)
31. TribhuwanKumarTewari,AnshulArya, Sameer Rastogi.: Message reading through eye blinking, IJCA 0975-887, Vol.2,No.6,JIIT, Noida Sec 62,2010
32. Axelsson S (2013) Bell's Palsy-medical treatment and influence of prognostic factors. vol 18. Sweden
33. Sullivan FM, Swan IRC, Donnan PT, Morrison JM, Smith BH, McKinstry B, Davenport RJ, Vale LD, Clarkson JE, Stewart K (2009) A randomized controlled trial of the use of acyclovir or prednisolone for the early treatment of bell's palsy: bell's study. Health Technol Assess NIHR HTA programme 13(47)
34. Bettadapura V (2012) Face expression recognition and and analysis: the state of the art, college of computing Georgia institute of technology. Cornell University Library

Chlorella Algae Image Analysis Using Artificial Neural Network and Deep Learning

S. Lakshmi and R. Sivakumar

Abstract Generally, solutions and results to a problem in image processing involve a lot of trail and testing with huge set of sample images. Chlorella is a single-cell, freshwater green algae, and it consists of green plant pigments, chlorophyll, vitamins, minerals, and protein, fiber, and omega fatty acids. The size of the chlorella cells are 10–30 μm. Due to its photosynthetic process, it converts carbon dioxide into fresh oxygen. Automatic identification and classification of algal community are very difficult due to various factors such as change in size and shape with climatic changes, various growth periods, and the presence of other microbes. In this chapter, an elaborate analysis of artificial neural network concepts and convolutional neural network (CNN) of deep learning technique that automatically measure the algae growth through the image classification techniques from algae digital images using MATLAB is presented.

Keywords Digital image processing · Convolution neural networks (CNN) · MATLAB · Back propagation · Radial basis · Probability neural network · Chlorella · Algae

1 Introduction

India is an agriculture-based country. Getting the expert's advice is an essential one not only to improve the production but also helpful to face the challenges such as unexpected diseases, natural calamity, and contamination due to other species such as bacteria, rotifer, or other microalgae efficiently and take actions effectively at right time. Farmers in rural India have minimal access to agricultural experts,

S. Lakshmi (✉)
Jeppiaar SRR Engineering College, Chennai, Tamil Nadu, India
e-mail: lakshmi1503@gmail.com

R. Sivakumar
Tejas Biotech P Ltd, Chennai, Tamil Nadu, India
e-mail: sivaa21@yahoo.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_10

who can inspect the plant and render advice. Delayed expert responses to queries often reach farmers too late which not only leads the shortage of production but also reduce the quality of the product. There is a growing interest from medical and biology researchers to develop automated tools and algorithms that can efficiently extract and utilize useful information from large scale biomedical image databases. This section will give an introduction about algae and its cultivation procedure and measurement techniques, MATLAB, Image processing, neural network, and deep learning.

1.1 Algae

Nowadays, algae are considered as a source for the production of renewable energy. The concentration of carbon dioxide is considerably raised in the atmosphere which leads to the climatic changes, globally. Fuel costs are raised due to the shortage of fossil fuels. Hence, we are in need of finding alternate solutions to this problem. Algae are rich in fats, carbohydrates, and proteins. Algae can grow in freshwater as well as in marine water. It may grow even in wastewater, and it will give fresh oxygen to the environment. Algae oil is used to make biodiesel, and the algae protein can be used as feedstock and food supplement. Biodiesel is produced from algae, and it requires many developments. Research is going on to extract the specific features from algae.

Chlorella is a single-cell, freshwater green algae, probably one of the first organisms seen on the earth. The name chlorella is obtained from the Greek word–chloros (meaning green) and the Latin word–ella (meaning small). It is an edible form of algae, and it consists of green plant pigments, chlorophyll, vitamins, minerals, protein, fiber, and omega fatty acids. Due to its high nutritional profile, it is also called as super food. The sample chlorella pond image is depicted in Fig. 1. Some advantages of chlorella are listed below:

- It allows diversification from traditional crops in case land and water resources are short.
- It has also been used as a complementary ingredient of feed for poultry and as a protein and vitamin supplement to aqua feeds.
- It is used for wastewater treatment.
- It can be produced from small scale industries to mass cultivation level.
- It is consumed as a health supplement in the USA and as food supplement in Japan.
- Chlorella is used for preventing cancer, reducing radiation treatment side effects and stimulating the immune system.
- The consumption of natural ‘whole foods’ rich in macronutrients has many health benefits for those in nonvegetarian diet [17].

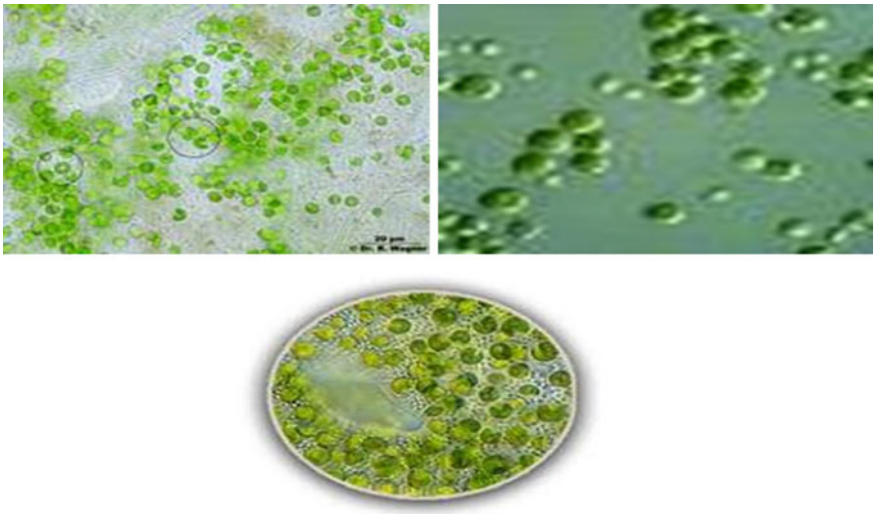


Fig. 1 Chlorella algae images

The extensive study about the ecological response of phytoplankton has been done [19], and the mathematical models have been designed [28]. The [15] has stated that the prediction of algae bloom dynamics remains a difficult problem. A wide investigation has been done [16] to use ANN as the ecological modeling tool. A system has been designed [5, 24, 30] to measure the temporary changes of a particular algae species in freshwater.

Normally, the microalgae are cultured for eight days. Joseph et al. [10] developed an ANN model by using the eight environmental variables with 7–13 days time. They are Skeletonema, water temperature, total inorganic nitrogen, phosphorus, salinity, secchi disc depth, wind speed, and tidal range.

Algae Cultivation and Measurement: Algae cultivation takes place in two ways. They are

1. Open pond system and
2. Photobioreactor.

The sample pond is shown in Fig. 2. The proposed automated tool is used to overcome the challenges when we are producing algae at mass level. The main objective of this chapter is to measure the algae growth level in the ponds to improve the production. Normally, the algae growth is measured by using three methods. They are as follows:

1. Numerical,
2. Gravimetric, and
3. Proximate.

Fig. 2 Sample algae pond

These methods can be used to measure the following:

1. Numerical is the very old technique and it requires microscope. But algae are counted manually. There may be a chance for man-made mistakes.
2. The amount of algae biomass is measured. It requires fluorometer or a spectrophotometer.
3. The preweighted filter is used.

1.2 MATLAB

MATLAB is a software for doing mathematical calculations, analysis, and algorithm development. It is termed as Matrix Laboratory which has high-performance programming language developed by MathWorks in 1984 [6, 7]. We can develop an interface with other programming languages such as C, C++, JAVA, and Python from MATLAB. It may be used in engineering, science, economics, and so on. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects.

Novel ideas and technologies used in engineering purely depend on mathematics concepts because math gives a way to express the ideas efficiently. Some of the mathematical concepts are listed here. They are as follows:

a. Linear algebra: It is a part of the workplace in MATLAB. It is used to do the following:

   - Taking decisions in life and business
   - Creation of plan
   - Predict the amount of turnover in a company

b. Numerical Analysis: It relies on the approximation. It may be used in the following fields:

   - Construction
   - Astronomy

c. Simulation: It may not provide a complete solution, but it is required to test any engineering and scientific approach for testing for the following reasons:

   - Reduces costs
   - Saves time
   - Improves the Success rate
   - Security and
   - to perform some specific tasks as follows:

     – Creating the snapshot based on the model
     – Display the workflow of the new technology
     – Define the mathematical model used to explain the technology

d. **Tool boxes:** Tool boxes are the collection of MATLAB functions, and m-files are used to solve the problems. Various tool boxes are available in MATLAB. Some of them are listed below:

   - Image acquisition toolbox
   - Image processing toolbox
   - Fuzzy logic
   - Wavelet
   - Neural network
   - Signal processing
   - Control systems and so on.

Image Processing Toolbox: It is used to manage the pixels in an image using math techniques. The basic data element is a matrix. The image is considered as a matrix. This toolbox supports four types of images. They are gray scale, binary, indexed, and RGB color images. The toolbox provides functions for converting the images from one class to another. The features like edges can be extracted easily by using the detector functions in MATLAB. Hence, it integrates computation and programming in an easy way.

1.3 Digital Image Processing

In image processing, the input is an image such as photograph or video frames, and the output will be an image or a set of parameters related to the image which will be extracted while processing the image. The main objectives of image processing are as follows:

(i) improve the information for human understanding
(ii) Processing the image for storage, transmission, and representation.

Image processing is to analyze the digital images for various applications in society [9]. It is used in several places such as medical images, spatial images, underwater images, and other biological images. Image processing is considered as a standard scientific tool, and its techniques are applied to all natural images [10]. Some developed tools are used for online monitoring, some tools are used for measurements of microorganism in water, and others are used to assist in recognition process such as enhancing images, noise elimination, and edges-extracted segmentation [11, 25].

Generally, images contain objects which may convey the information about the objects in the image. Image processing and analysis provides a way to extract and quantify the objects for answering the meaningful queries. Since the microscopic images are very complex and noisy, multiple processing steps are required for the extraction and quantification of the objects. A common approach for image analysis is explained as follows:

1. The initial step in image analysis is the acquisition of images through digital cameras or microscope-enabled mobile phones. Most of the raw images are affected by noisy pixels or defective pixels due to the problems in image acquisition system. It is necessary to correct the defected pixels and noisy pixels by identifying and removing the unwanted pixels, and then by using the image filtering techniques, the contrast is enhanced for doing the consequent process.
2. Various techniques such as image segmentation and morphological image processing can be applied to extract the features of the images.
3. After extracting the important features from the images, quantitative information about the objects will be derived.

The image matching is considered as a difficult task, whereas the images of the same picture or environment may be different due to the motion and intensity variation. Neural networks may be implemented for taking decisions about image matching results [3, 18].

1.4 Classification

Mining is a wide area that combines the techniques from various fields such as pattern recognition, artificial intelligence, statistics, and machine learning. Classification is done based on the similarity of the objects into groups. Automatic classification is necessary for designing intelligent systems. Classifiers are designed manually based on expert's knowledge. The most common classification methods are as follows:

- Decision trees: It represents a set of classification rules in a tree form. It is very simple and we can handle the mixed variables easily.
- Bayesian networks: It is represented visually as a graph structure.
- k-nearest neighbor classifier: KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common among its k-nearest neighbors measured by a distance function. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor.
- Support Vector Machine: The main aim is to find the best classification function to distinguish between members of the classes in the training data. It gives more attention to the class boundaries. The simplest way to separate two groups of data is with a straight line for one dimension, flat plane for two dimensions, or hyper plane for N dimensions. It performs the classification by finding the hyper plane that maximizes the margin between the two classes. The vectors that define the hyper plane are the support vector. It provides the robust and accurate results. The advantage of SVMs is to find the global optimum since there is no local optima in maximizing the margin.
- Artificial Neural networks
  Artificial neural networks are the processing information that are inspired on the architecture and functioning of the biological nervous system. This is composed of a large number of interconnected processing elements which are called as neurons used to solve some specific applications like pattern recognition through a learning process. Learning is the adjustments of the connections that exist between the neurons in neural network. The most complex problems are solved using combination of simple processing elements (neurons). Since all the neurons operate in parallel, the processing of information will be very fast.
  A trained neural network can act as an expert to analyze the information efficiently to answer especially for 'what if' questions. Artificial neural networks (ANN) have proved very successful in various machine learning and artificial intelligence areas. Recent advances in neural network-based deep learning architectures have demonstrated promising results in solving general image classification problems.

Architecture of Neural Network: There are two broad classification of neural network. They are as follows:

a. Feed-forward network and
b. Feedback network.

Feed-forward networks: It allows the signals to travel one way only, from input to output. Feed-forward ANNs tend to be straight-forward networks that associate inputs with outputs. They are extensively used in pattern recognition.

Feedback networks: It allows the signals to travel in both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. Feedback networks are dynamic; their 'state' is changing continuously until they reach a stable point. They remain at the stable point until the input changes, and a new stable point needs to be found.

Neural network layers: The common type of neural network consists of three layers. They are as follows:

- Input layer: the input of the network is presented to the neurons of this layer. The raw information is fed into the network.
- Hidden layer: these contain neurons that only connect to other neurons. The action of each hidden layer is determined by the actions of the input units and the weights between the hidden and output units.
- Output layer: The output of the neurons in this layer forms the output of the network. It purely depends on the activity of the hidden layers and the weights between the hidden and output units.

Example for neural network layers:

3-2-3-2 Feed-Forward Network

It contains three nodes in input layer, two nodes in the first hidden layer, three nodes in the second hidden layer, and two nodes in the output layer.

Learning process in neural network: A neural network learns from the surroundings through the interactive process. Generally, information is stored in the weight matrix of the neural network. Learning is used to determine the weights. Hence, learning is performed in two ways.

1. Fixed networks—weights cannot be changed.
2. Adaptive networks—weights can be changed.

Learning methods are classified into two major categories. They are as follows:

1. Supervised learning and
2. Unsupervised learning (self organization).

In adaptive learning, the neural network learns how to do the tasks based on the data given for training. In self-organizing, the network creates its own organization by using the received data during the learning process.

Applications of Neural Network: Neural networks are broadly used in various fields due to its prediction capabilities. Some of them are listed here:

- In business, ANN can be used to do marketing, sales, and so on.
- Customer research
- Industrial process control
- In medical, ANN is considered as a hot research area. It can be used to recognize diseases from scans such as ultrasonic scan, MRI scan, and cardiograms.
- **Deep Learning:** Deep learning is a branch of machine learning based on a set of algorithms. Many layers are present in between the input and output which allows the algorithms to use multiple layers made of linear and nonlinear transformations. Each layer uses the output of the previous layers as input, and the algorithms may be supervised or unsupervised. Numerous deep learning architectures are available such as deep neural networks, convolutional deep neural networks, deep belief networks, and recurrent neural networks. A separate tool box is available in MATLAB 2016 for deep learning for modeling the hierarchical data.

2 Problem Formulation

Automatic identification and classification of algal community are very difficult due to various factors such as change in size and shape with climatic changes, various growth periods, and the presence of other microbes. Hence, our aim is to design a tool to assist the experts in the following ways:

- to measure the growth of chlorella,
- to calculate the yield of each and every pond to forecast the production.

3 Problem Solution

Designing and building up of a system for classification of algae images for measuring the growth is a complicated task because of the significant similarities among different classes. Classical conventional techniques for preprocessing, segmentation, and post-processing were used for differentiating the size of the algae growth measurement. The entire process for algae growth measurement is depicted in the Fig. 3.

The manual way of monitoring the algae systems may provide wrong results due to the employee's fatigue. In order to raise the production of algae growth, the following steps are taken to design an automatic system. They are as follows:

1. Capturing the input images from the chlorella ponds
2. The unwanted information i.e., noisy pixels are removed by using the various filtering techniques.
3. The interested features are extracted from the noise-free images
4. Features are stored in a database and proper training is given
5. Testing is taken to check the classified results.

The detailed explanation of each and every step is given below.

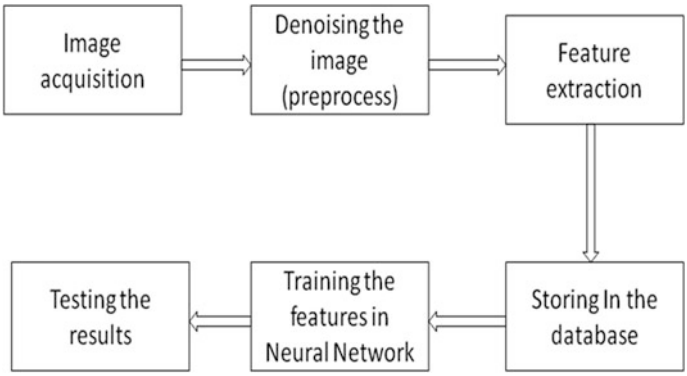


Fig. 3 Entire process of algae growth measurement

3.1 Image Acquisition

Nearly 20 digital photographs were taken daily at same time in the morning as well as in the evening for 10 days, and totally, it comes around 400 images. It is ensured that the same lighting condition is maintained to define the processing parameters. To speed up the entire process especially training and testing phase, we need to minimize the size of the image by using the resize function in MATLAB. The images are read into the MATLAB environment using imread() function in MATLAB. The syntax is

imread('filename')

The file name is the string containing the name of the file including the path where it is located in the system. By default, it searches in the mydocuments/MATLAB folder. Otherwise, we need to specify the exact path. It will bring that file to the MATLAB environment. Then, the input images are converted to gray scale images for doing further process using the function rgb2gray().

3.2 Noise Removal

Noise is unwanted information which is added with the image during image acquisition process. Moreover, any real world image is definitely affected by noise. The unwanted information is removed by applying various filtering techniques before processing the image. For dealing with this alga images, we have applied various filters in MATLAB and analyzed the results. The median filter suits to remove the noise from the algae images. The syntax for median filter is as follows:

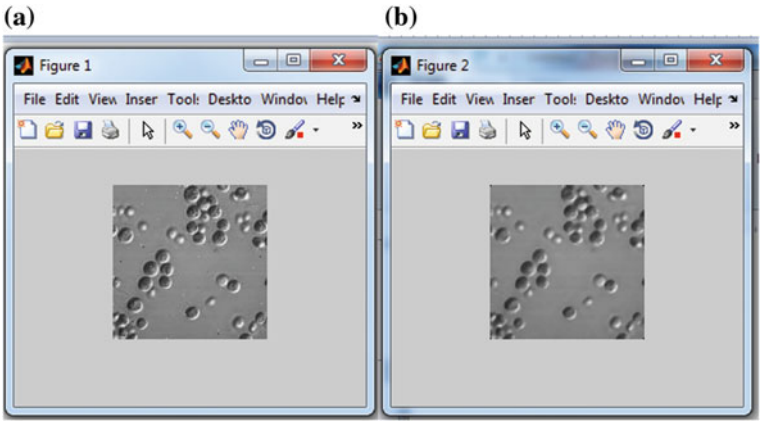


Fig. 4 **a** Noisy image. **b** Denoised

$$K = \text{medfilt2}(J);$$

where J is an noisy input image. Figure 4 shows the noisy image and the denoised image after applying the median filter.

3.3 *Feature Extraction*

When we are taking the whole image for processing, it requires a large amount of memory and power to process the whole image. Feature extraction is the process of defining a set of features or image characteristics which represents the important information for analysis and image classification. Feature extraction is a method which reduces the huge volume of data considerably, and the same time it maintains the accuracy of the data. Since the texture played a major role for image analysis, first-order and second-order statistics are used to extract the features from the images. In first-order statistics, the original image texture measures are considered such as standard deviation, mean, median, and variance, and pixel neighborhood values should not be considered. The following features are extracted in first-order statistics. They are as follows:

(i) ***Mean***: It is a measure of brightness of the given image. It is calculated by using the formula

$$\mu_p = \frac{i}{n^2} \sum_{r=0}^{n-1} \sum_{s=0}^{n-1} P_{r,s} \tag{1}$$

(ii) ***Standard deviation***: It is a measure of contrast. It is calculated as follows

$$\sigma_P = \left[\frac{1}{n^2}\sum_{r=0}^{n-1}\sum_{s=0}^{n-1}\left[p_{r,s} - \mu_p\right]^2\right]^{1/2} \tag{2}$$

(iii) ***Variance***: It is a measure to tell the gray level variation from the mean. The formula is

$$\sum_i \sum_j (i + j - 2\mu) P_{i,j} \tag{3}$$

(iv) ***Mode***: It is a measure of the gray level which occurs most frequently.

(v) ***Median***: It is a middle value of the set of gray values of an image.

(vi) ***Skewness:*** It is a measure of the asymmetry of the data around the sample mean. If it is negative, the data are spread out more to the left of the mean than to the right, and if it is positive, the data are spread out more on the right side. The normal distribution skewness is zero.

$$\mu = \frac{1}{s^3}\sum_{i=1}^{L}(k_i - \text{mean})^3 p(k_i) \tag{4}$$

(vii) ***Kurtosis:*** The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3. The kurtosis is given by:

$$\mu = \frac{1}{s^4}\sum_{i=1}^{L}(k_i - \text{mean})^4 p(k_i) - 3 \tag{5}$$

In second-order statistics, Gray Level Co-occurrence Matrix (GLCM) is used to attain statistical features. Shanmugam et al. [26] defined the techniques for extracting fourteen features from the images using co-occurrence matrix; here, we are taking only four features are considered at second-order level.

Gray Level Co-occurrence Matrix (GLCM) method is used to extract second-order statistical features. The sample chlorella image is shown in the Fig. 5 and the GLCM matrix of the taken chlorella sample image in Fig. 5 is depicted in Fig. 6. In GLCM matrix, the number of rows and columns are equal to the number of gray levels in the input image. The matrix element $P(i, j|d, \theta)$ is the relative frequency with which two pixels i and j at a displacement of d and an angle θ. The GLCM calculation procedure table is displayed in Table 1. Here, four different gray levels are displayed. If we make the window size larger, we will get more gray levels. The command graycomatrix() is used in MATLAB to construct the gray level co-occurrence matrix. We have to pass an image variable as a parameter for

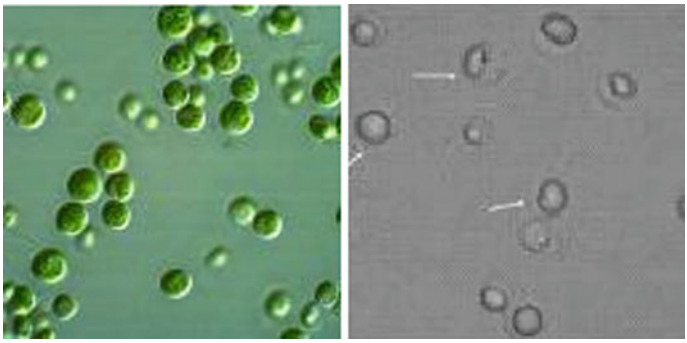


Fig. 5 Sample chlorella images (*color and gray scale*) for GLCM calculation

0	0	0	0	0	0	0	0
0	116	177	13	0	0	0	0
0	149	676	341	45	3	1	0
0	32	280	2062	855	48	17	2
0	10	77	831	9522	174	60	5
0	2	9	50	148	91	115	6
0	0	2	11	84	94	116	10
0	0	0	2	6	6	7	1

(a) Color image

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	16	50	10	0	0	0
0	0	51	476	369	2	0	0
0	0	9	367	12880	31	1	0
0	0	0	2	27	40	6	1
0	0	0	0	6	2	227	9
0	0	0	0	1	0	9	1664

(b) gray scale image

Fig. 6 GLCM matrix

Table 1 GLCM calculation procedure

Reference pixel value	Neighborhood pixel value			
	0	1	2	3
0	0, 0	0, 1	0, 2	0, 3
1	0, 0	0, 1	0, 2	0, 3
2	0, 0	0, 1	0, 2	0, 3
3	0, 0	0, 1	0, 2	0, 3

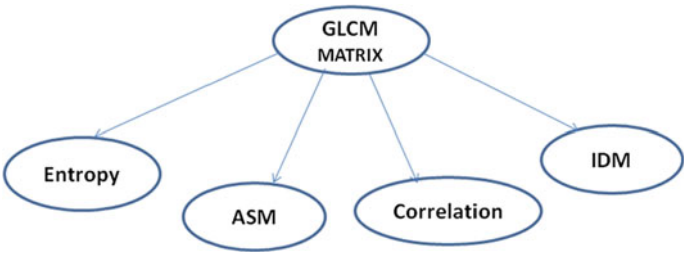


Fig. 7 Extraction of GLCM features

this function to get the values of the second-order image features. This will return the 8 × 8 matrix which is the GLCM of input image.

The syntax for creating the gray level co-occurrence matrix is

$$\text{com} = \text{graycomatrix(ii)};$$

It will generate the 8 × 8 matrix as follows.

The next command graycoprops() is used to extract the second-order statistical features such as contrast, entropy, IDM, and ASM of an image from the gray level co-occurrence matrix.

st = graycoprops(com,'all');

The first argument com is the gray level co-occurrence matrix, and the second argument 'all' is used to mention all features from the GLCM matrix which is depicted in Fig. 7.

The GLCM matrix is shown in the Fig. 6. Extraction of the four important features using GLCM matrix as follows:

1. ***Angular Second Moment:*** It is also known as energy. It is a sum of squares of entries in the GLCM matrix. It is high when image has very good homogeneity or when the values of the pixels are same. The formula for calculating this is

$$\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} Pi,j^2 \tag{6}$$

   where i, j are the spatial coordinates of the function $p(I, j)$, and Ng is gray value.
2. ***Correlation:*** It measures the linear dependency of gray levels of neighboring pixels. This is used to measure deformation and displacement, but it is also used in many areas of science. The formula for calculating the correlation is

$$\frac{\sum\sum (i,j)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{7}$$

3. ***Entropy:*** Generally, entropy shows the amount of information of the image, and it also measures the loss of information. The formula for calculating this is

$$\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} -Pi,j * \log Pi,j \tag{8}$$

4. ***Inverse Difference moment:*** It is the local homogeneity, and it is high when the local gray level is uniform. The formula for calculating this IDM is

$$\frac{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} Pi,j}{1+(i-j)^2} \tag{9}$$

3.4 Neural Network Implementation

We are extracting the features from the denoised images and then feeding them into the neural network for classification. The general steps are followed to implement the neural network for classification. They are as follows:

1. Network Creation
2. Train the network
3. Testing the network
4. Classify.

Network Creation: Nevertheless, we have many neural network types; we are using the following networks for our work.

Feed-Forward Back Propagation Network: A feed-forward back propagation network model is shown in the Fig. 8 and it is created by using **newff()** function. It has three arguments. The fist argument is the input vector, second one is the output vector and the third argument is used to specify the number of hidden layers. The default transfer function for the hidden layer is tansig and the output layer is purelin.

Ex:
newff(P,T, 6)
where P is the input vector,
T is the Output/target vector and the third argument
'6' is the number of hidden layers in the feed forward neural network.

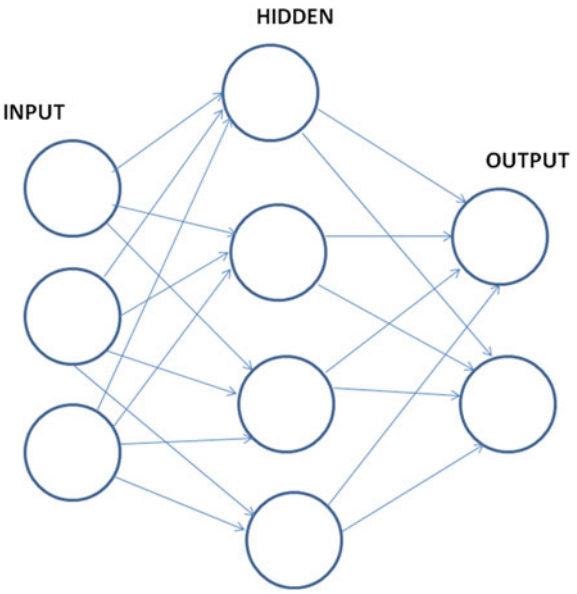


Fig. 8 Feed-forward back propagation neural network

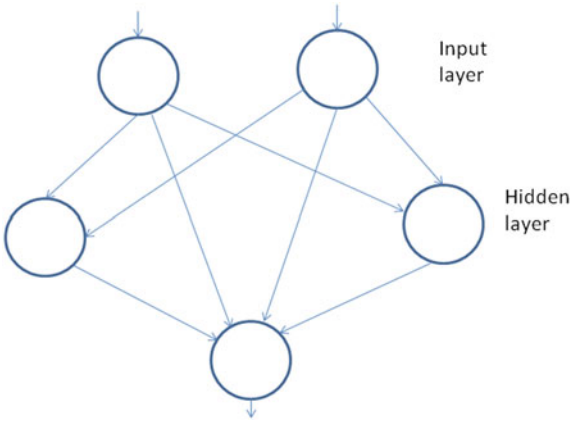


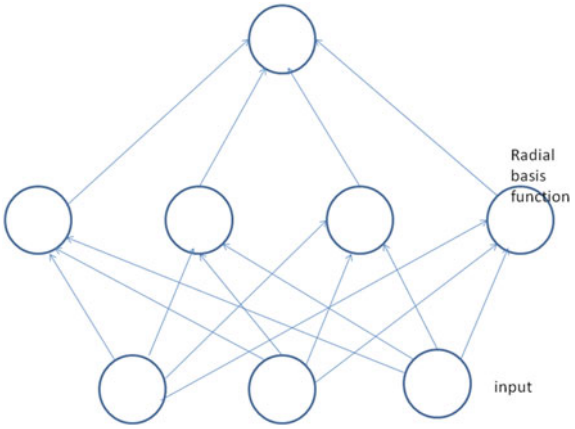
Fig. 9 Cascade-forward back propagation network

During training the calculation were carried out from the input layer to output layer and the error values were transmitted to prior layers.

Cascade-forward back propagation network: This model is similar to feed-forward networks and it is shown in the Fig. 9. It contains a weight connection from the input to hidden layers one by one. The very important point in this type is that each layer of neurons are purely related to all previous layers of neurons. The syntax for creating this type of network is done by using the function **newcf()**.

Ex: newcf(P,F,5);

Fig. 10 Radial basis network



The first parameter is the Input vector, the second parameter is the output vector and the third one is the number of neurons of the hidden layer.

Radial basis network: This network drives from the theory of function approximation. It s a two layer feed-forward networks. The hidden nodes implement a set of radial basis functions. First of all the weight from the input layers are determined then he weights from the hidden layer to output layer is determined through the linear summation. When compare with other network the radial basis training is very fast and gives very good results. The structure of the radial basis network is depicted in the Fig. 10.

The syntax for creating the radial basis network in MATLAB is **newrb()**.

Ex: newrb(P,T)

where P is the input vector and T is the output vector. Two layer network is created by the function newrb(). The input layer has radbas neurons and the weighted inputs are calculated with distance and its net input with netprod. The purelin neuron is in second layer and the weighted input is calculated with dotprod and the total inputs with netsum.

New probabilistic neural network: The syntax for creating the probabilistic neural network is

$$\text{newpnn}(\text{P}, \text{T}, \text{spread})$$

where P is the input vector, T is the target vector and spread—Spread of radial basis functions (default = 0.1) and returns a new probabilistic neural network. If spread is near zero, the network acts as a nearest neighbor classifier. As spread becomes larger, the designed network takes into account several nearby design vectors.

***SOM* (*Self-Organizing Map*);** The syntax for creating SOM is

$$\text{newsom (simpleclassInputs,[8 8])}$$

where simpleclassInputs are input to the network and the next parameter is the dimension of the next layer.

CNN: Convolutional neural networks (CNN) are used in various applications such as image classification, speech recognition, and pattern recognition. A CNN is a special case of the neural network and it consists of one or more convolutional layers, often with a subsampling layer, which are followed by one or more fully connected layers as in a standard neural network. Each and every layer has many neurons. The CNN is a multistage processing of an input image to extract high level features for doing further processing. The convolutional neural network is a type of deep learning concept which is used to solve the image classification problems effectively [14].

In visual recognition tasks the deep convolutional neural network produces better results [12, 23, 31]. Prior feature extraction from images are not required in the CNN-based algorithms. It provides outstanding results in object detection [22], object recognition [13, 27] and scene classification [1, 2, 4, 8, 20]. Four types of layers are most common in CNN. They are as follows:

- convolution layers,
- pooling/subsampling layers,
- nonlinear layers and
- fully connected layers.

Convolution layers: The convolution operation extracts different features of the input. It removes noise from the input and extracts low-level features such as edges, lines, and corners. Higher-level layers extract higher-level features. The convolution filter kernel weights are decided on as part of the training process. Since it is a unique method, it combines segmentation, feature extraction and classification in one process.

Pooling/subsampling layers: The noise-free i.e., denoised features are made by these layers. Initially it reduces the damages in the images. There are two ways to do pooling. They are

- max pooling and
- average pooling.

In pooling, the input is divided into nonoverlapping two-dimensional spaces. In average pooling, the average of the four values in the region is calculated. In max pooling, the maximum value of the four values is selected. The input is of size 4×4. For 2×2 subsampling, a 4×4 image is divided into four nonoverlapping matrices of size 2×2. In the case of max pooling, the maximum value of the four values in the 2×2 matrix is the output. In case of average pooling, the average of the four values is the output. Stochastic Diagonal Levenberg Marquadt learning algorithm is applied in training the network. Minimum redesigning is required when we are doing alteration in the database [21, 29].

Fully connected layers: It is used as the final layers of a CNN. These layers mathematically sum a weighting of the previous layer of features, indicating the

precise mix of 'ingredients' to determine a specific target output result. In case of a fully connected layer, all the elements of all the features of the previous layer get used.

4 Training and Testing

The general methodology of the algae classification is explained in a step by step manner. The detailed training and testing stages are depicted in the Fig. 11:

A. **Training stage**

1. Load an image from the database
2. Resize the image to speed up the process
3. Convert the color image into gray scale image
4. Preprocess the image for extracting the features
5. Create a neural network by using the extracted features

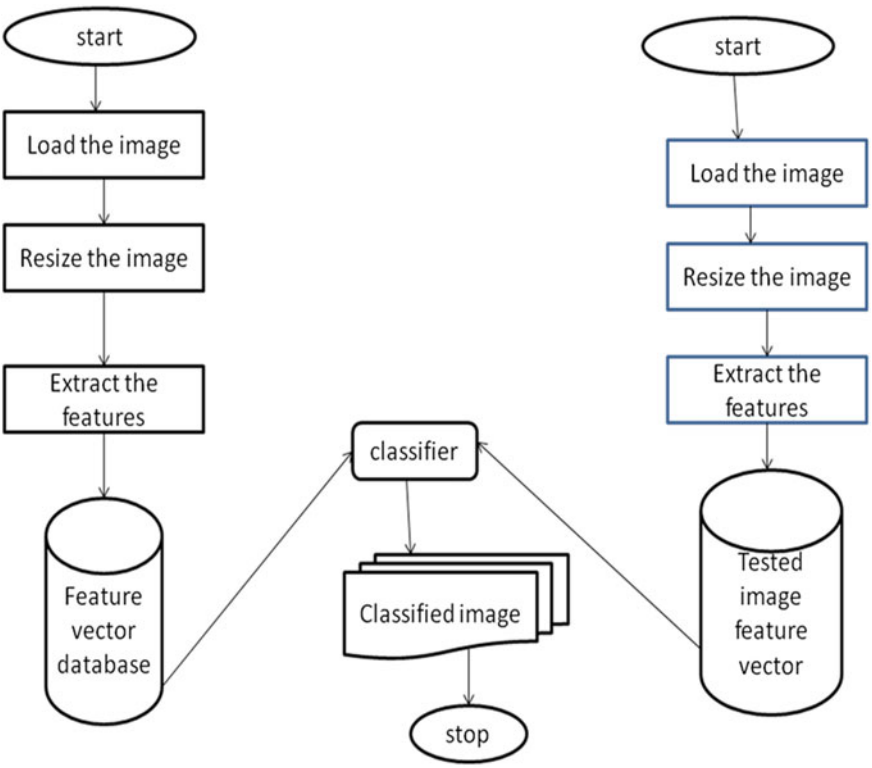


Fig. 11 Training and testing phases

6. Train the network
7. Save the results.

B. **Testing stage**

1. Load an image to be tested from the testing folder
2. Resize the image to speed up the process
3. Convert the color image into gray scale image
4. Preprocess the image for extracting the features
5. Extracted features are fed to a trained neural network
6. Classification algorithm makes the decision
7. Check the results.

5 Source Code in MATLAB for Classification

The sample source code in MATLAB is given for algae image classification using first-order statistics, second-order, and CNN for measuring the growth of chlorella algae. In the sample code we have given only one input image for making the training and testing process easier. We need to give some set of images for training as follows:

```
Input_folder = '.\'; % folder with big images
D = dir([Input_folder 'img *.jpg'])
Inputs = {D.name}'
for k = 1:length(Inputs)
   X = imread([Input_folder Inputs{k}]);
end
```

We have trained the network using the following methods:

1. the feed-forward back propagation network,
2. cascade-forward back propagation network,
3. radial basis network,
4. probabilistic neural network,
5. self-organizing map (SOM), and
6. convolutional neural network (CNN).

We have given the source code for newff in first-order, second-order, and CNN methods only in this section. We have already discussed the functions to create a network using cascade, radial basis, and so on.

(i) First-order statistics

```
clc;
close all;
clear all;
%input
i = imread('d:/book chapter/dataset-chlorella/timg1.jpg');
ii = rgb2gray(i);
kk = double(ii);
i1= mean(kk);
as = std(kk);
av = var(kk);
am = mode(kk);
ame =median(kk);
s= skewness(kk);
k = kurtosis(kk);
%output
i21 = imread('d:/book chapter/dataset-chlorella/timg2.jpg');
ii1 = rgb2gray(i21);
kk1 = double(ii1);
i2 = mean(kk1);
as1= std(kk1);
av1 = var(kk1);
am1 = mode(kk1);
ame1 =median(kk1);
s1= skewness(kk1);
k1 = kurtosis(kk1);
 p =[i1 as av am ame s k];
t = [i2 as1 av1 am1 ame1 s1 k1];
% feed-forward backpropagation network
  rand('seed',0.5000);
 net = newff(p,t,5);
 net= train(net,p,t);
 Y = round(sim(net,p));
 plot(p,t,p,Y,'o')
```

(ii) Second-order statistics

```
clc;
close all;
clear all;
 %input
i = imread('d:/book chapter/dataset-chlorella/spi-rot/onet.jpg');
ii = rgb2gray(i);
kk = double(ii);
com = graycomatrix(ii);
st = graycoprops(com,'all');
%output
i21 = imread('d:/book chapter/dataset-chlorella/spi-rot/twort.jpg');
ii1 = rgb2gray(i21);
kk1 = double(ii1);
com1 = graycomatrix(ii1);
st1 = graycoprops(com1,'all');
p = [st.Contrast st.Correlation st.Energy st.Homogeneity];
t = [st1.Contrast st1.Correlation st1.Energy st1.Homogeneity];
 % feed-forward backpropagation network
 rand('seed',0.5000);
 net = newff(p,t,6);
 net= train(net,p,t);
Y = round(sim(net,p));
 plot(p,t,p,Y,'o')
```

(iii) Source code for chlorella image classification using CNN

```
%CNN for Chlorella algae image recognition
clear all;
close all;
clc;
 [I,labels, I_test, labels_test] = readchlorella(1000); %Load the image
numLayers = 8;  % No. of layers
numSLayers = 3; % No. of Subsampling layers
numCLayers = 3; % no. of Convolutional layers
numFLayers = 2; % No. of fully connected layers
numInputs = 1;
InputWidth = 32;
InputHeight = 32;
numOutputs = 10;
```

```
%Create an empty convolutional neural network
sinet =
cnn(numLayers,numFLayers,numInputs,InputWidth,InputHeight,numOu
tputs);
sinet.SLayer{1}.SRate = 1;
sinet.SLayer{1}.WS{1} = ones(size(sinet.SLayer{1}.WS{1}));
sinet.SLayer{1}.BS{1} = zeros(size(sinet.SLayer{1}.BS{1}));
sinet.SLayer{1}.TransfFunc = 'purelin';
sinet.CLayer{2}.numKernels = 6;
sinet.CLayer{2}.KernWidth = 5;
sinet.CLayer{2}.KernHeight = 5;
sinet.SLayer{3}.SRate = 2; %subsampling rate
sinet.CLayer{4}.numKernels = 16;
sinet.CLayer{4}.KernWidth = 5;
sinet.CLayer{4}.KernHeight = 5;
sinet.SLayer{5}.SRate = 2;
sinet.CLayer{6}.numKernels = 120;
sinet.CLayer{6}.KernWidth = 5;
sinet.CLayer{6}.KernHeight = 5;
sinet.FLayer{7}.numNeurons = 84;
sinet.FLayer{8}.numNeurons = 10;
sinet = init(sinet);
 sinet.CLayer{4}.ConMap = ...
[1 0 0 0 1 1 1 0 0 1 1 1 1 0 1 1;
 1 1 0 0 0 1 1 1 0 0 1 1 1 1 0 1;
 1 1 1 0 0 0 1 1 1 0 0 1 0 1 1 1;
 0 1 1 1 0 0 1 1 1 1 0 0 1 0 1 1;
 0 0 1 1 1 0 0 1 1 1 1 0 1 1 0 1;
 0 0 0 1 1 1 0 0 1 1 1 1 0 1 1 1; ]';
sinet.SLayer{1}.WS{1} = ones(size(sinet.SLayer{1}.WS{1}));
sinet.SLayer{1}.BS{1} = zeros(size(sinet.SLayer{1}.BS{1}));
sinet.FLayer{8}.TransfFunc = 'radbas';
sinet.epochs = 3;
sinet.mu = 0.001;
sinet.teta =  0.0005;
sinet.HcalcMode = 0;
sinet.Hrecalc = 300; %Number of iterations to passs for Hessian recalculation
sinet.HrecalcSamplesNum = 50; %Number of samples for Hessian recalculation
 sinet.teta_dec = 0.4;
 [Ip, labtrn] = preproc_data(I,1000,labels,0);
[I_testp, labtst] = preproc_data(I_test,100,labels_test,0);
sinet = train(sinet,Ip,labtrn,I_testp,labtst);
```

6 Results and Discussion

All the experiments and results are discussed here. We can check the network performance and if necessary the changes to be made in the training process. We can use the function plotperf(tr) to plot the performance progress of the network, and it is shown in the Fig. 12. The property best_epoch indicates the iteration at which the validation performance reached a minimum.

The regression plot in the Fig. 13 shows the relationship between the outputs of the network and the targets. If the training were perfect, the network output and the target would be equal. The training data, validation, and testing data are plotted. The perfect result is represented by dashed line and the best-fit linear regression represented the relationship between the output and target.

The *R* value is an indication of the relationship between the outputs and targets. If $R = 1$, this indicates that there is an exact linear relationship between outputs and targets. If *R* is close to zero, then there is no linear relationship between outputs and targets. One common image is taken for analyzing the performance of the training and testing images, and the results are shown in the image Fig. 14. The following results in Fig. 15 are taken from the second-order statistics of the image, and the network layer is shown in Fig. 16.

The above network shows the network architecture based on the first-order statistics of the given image. The regression analysis of the training and testing images based on the second-order statistics of the image is shown in Fig. 17.

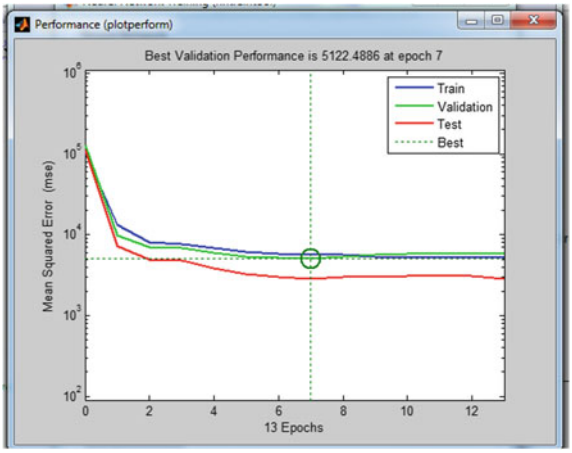


Fig. 12 Performance results of first-order statistics

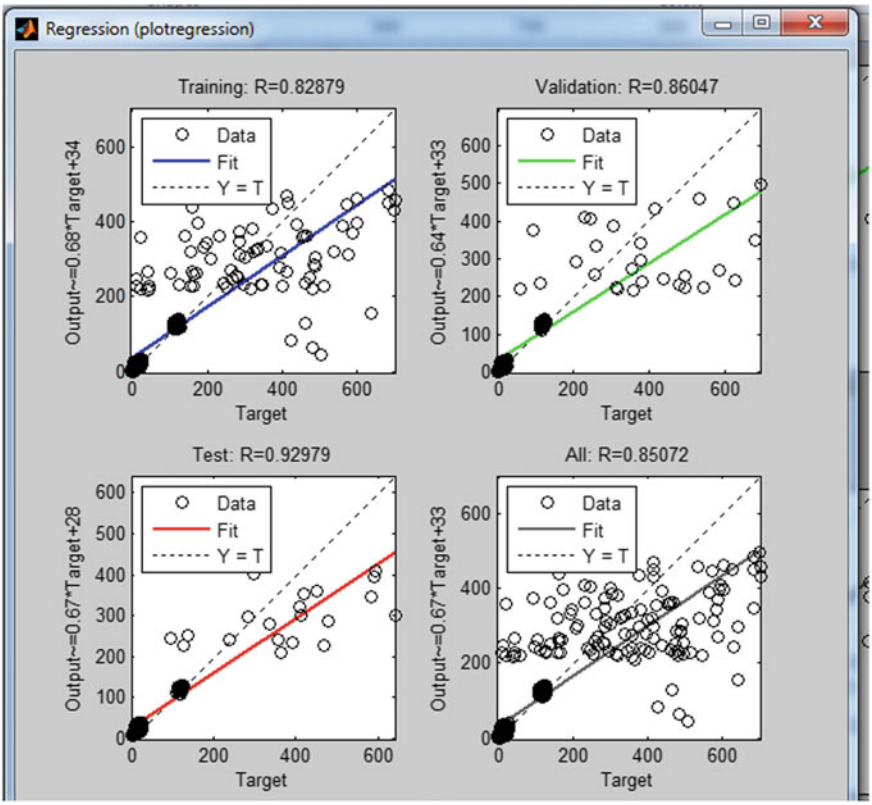


Fig. 13 Regression analysis of first-order statistics

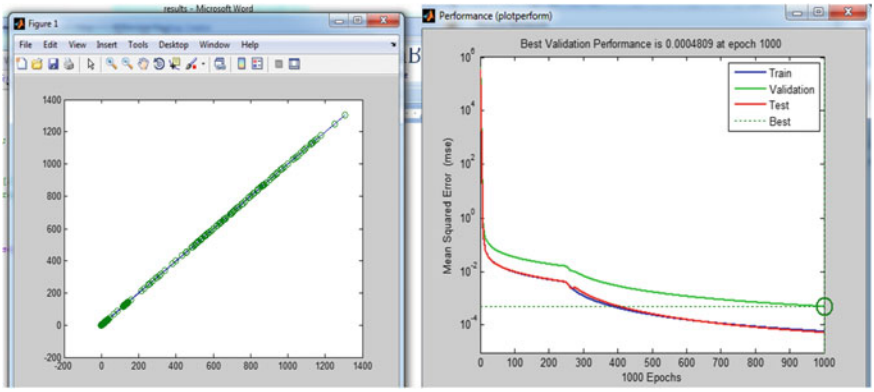


Fig. 14 Analyzing the training and testing image

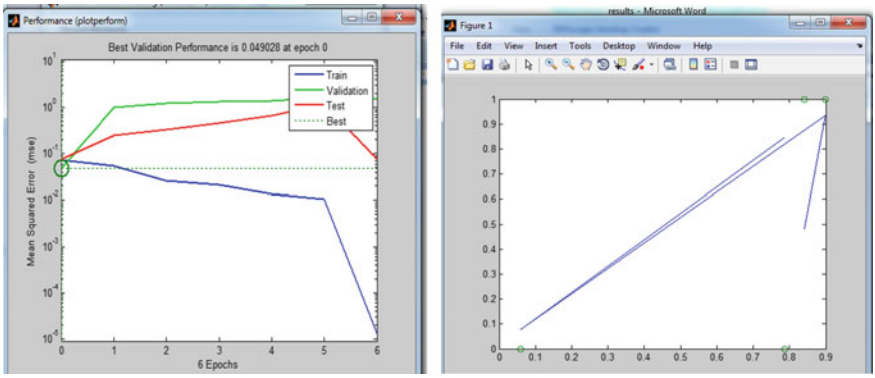


Fig. 15 Second-order statistics–results

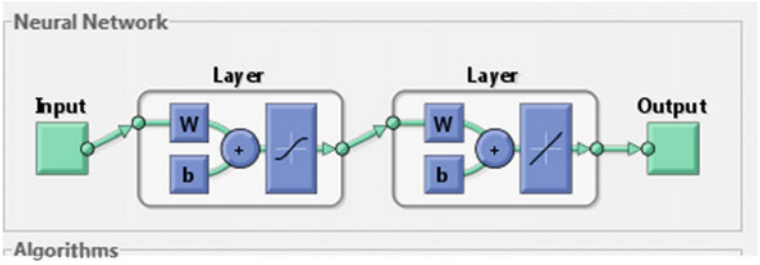


Fig. 16 Network layer

Figure 18 shows the performance analysis results of the input image which match with the training image, and the Fig. 19 shows the performance analysis of the gray color chlorella image. Generally, chlorella is in green color, and we used to give importance to color images.

We can calculate the first-order and second-order statistic values for extracting the texture features of the given image, then we can plot these statistical values for measuring the growth of the algae. Generally, algae can grow in seven days. Hence, the training set is generated for all seven days to measure the growth of the algae. Based on these training set, algae can be counted. The Fig. 20 depicts the training data generated by using the first-order statistics of the given image, and Table 2 shows the sample data for the first-order statistics features of the texture of the image. Table 3 shows the texture features calculated based on the values of the second-order statistics, and the graph shows the performance based on the second-order statistics in Fig. 21.

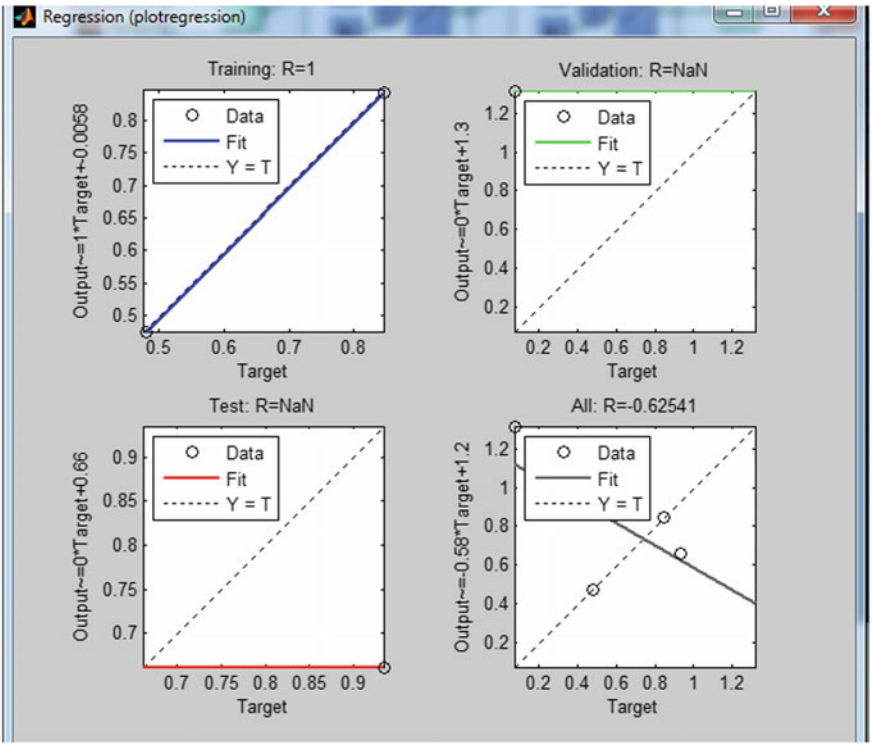


Fig. 17 Regression analysis of second-order statistics

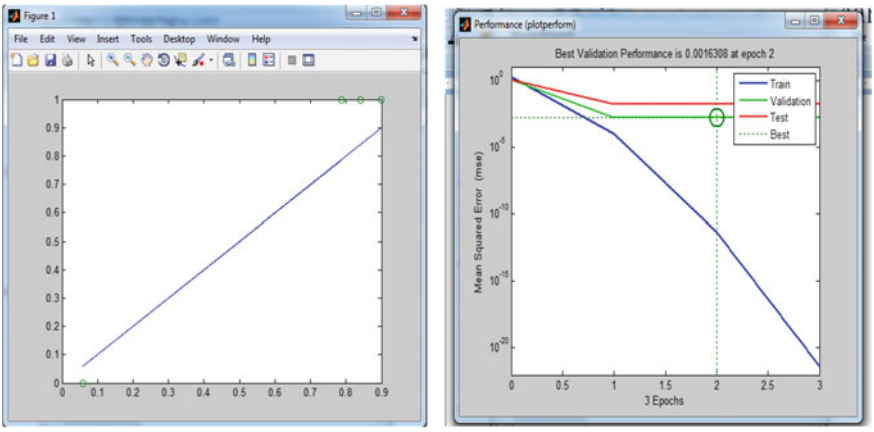


Fig. 18 Performance analysis

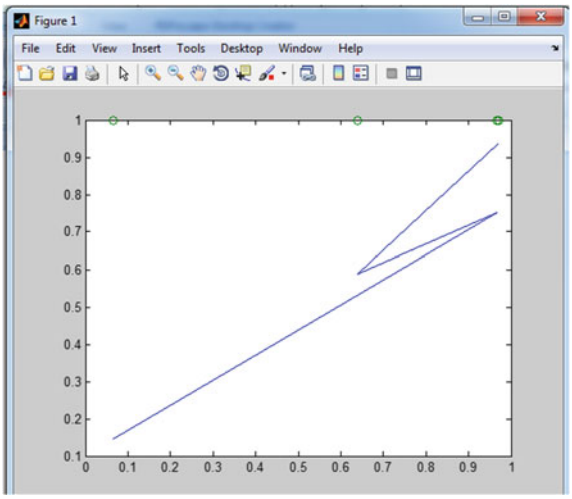


Fig. 19 Performance analysis of gray scale image

On the seventh day, almost the algae growth touches the maturity stage i.e., ready for harvesting. We can easily segregate the algae coils and also the healthy coils are identified clearly which is very important to count and measure the algae to raise the production. By using the convolutional neural network method, we can easily extract the image features without doing many calculations, and the following Fig. 22 shows the ROC curve of the CNN for extracting algae features. The programs are written in MATLAB for testing and analyzing the results. For testing and comparing, the capabilities of these said techniques are shown in Table 4.

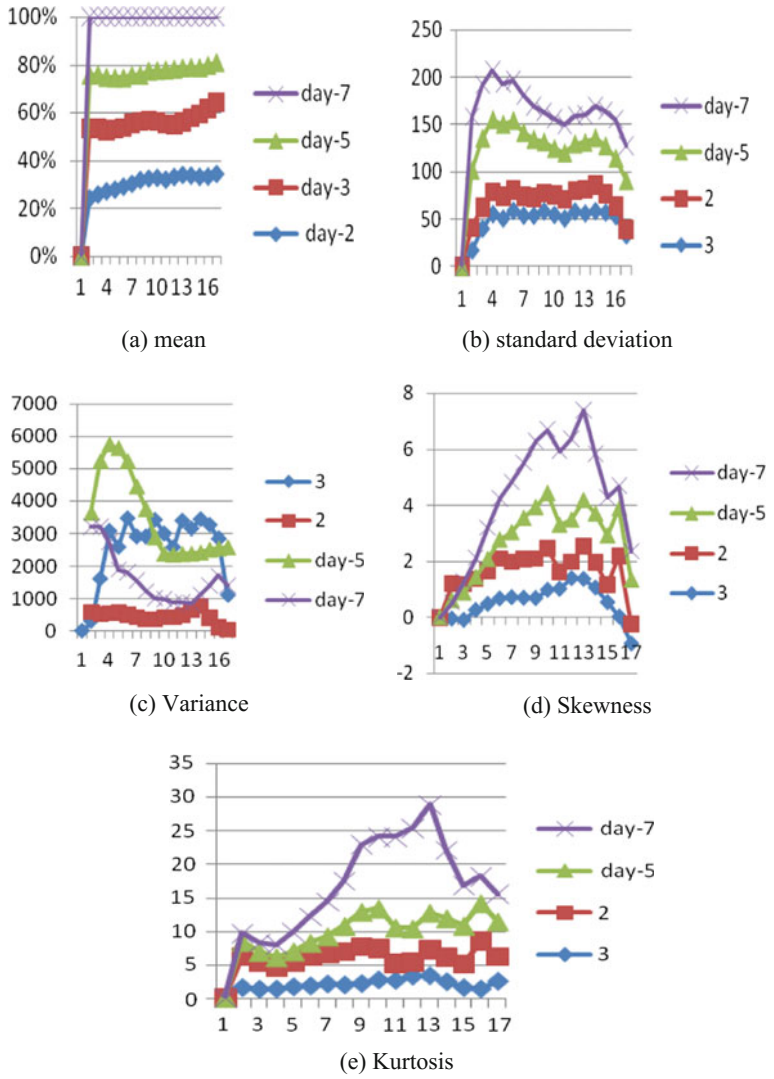


Fig. 20 Shows the performance analysis graph of first-order statistics

Table 2 Sample data for the first-order statistics features of the texture of the image

Mean				Std				Var			
Day-2	Day-3	Day-5	Day-7	Day-2	Day-3	Day-5	Day-7	Day-2	Day-3	Day-5	Day-7
177.56	215.75	160.72	178.44	24.04	17.40	60.30	56.53	578.00	302.73	3636.27	3195.35
173.56	185.63	149.16	162.06	22.72	40.18	72.40	56.62	516.13	1614.38	5241.36	3205.67
168.44	155.13	139.16	156.03	23.33	55.51	75.91	52.45	544.13	3081.58	5762.72	2751.32
165.13	150.63	125.03	150.59	23.71	50.98	75.12	43.57	562.25	2599.05	5643.32	1898.18
165.13	139.38	116.81	144.25	22.42	58.95	72.37	42.75	502.52	3475.05	5237.64	1827.68
167.50	138.88	111.03	134.47	20.48	54.02	66.73	39.59	419.60	2918.12	4452.48	1567.48
170.50	130.69	102.88	129.66	18.86	53.99	61.27	35.08	355.87	2914.50	3753.66	1230.94
172.25	129.13	107.75	119.53	18.98	58.54	53.85	31.94	360.20	3426.38	2899.42	1020.19
173.88	124.00	111.78	118.56	20.72	54.86	48.84	31.50	429.32	3009.33	2384.95	992.51
168.75	123.00	117.47	116.41	20.43	50.92	48.59	29.59	417.53	2593.20	2360.71	875.54
171.69	113.38	118.47	112.91	22.24	58.30	48.81	29.76	494.76	3399.05	2382.39	885.83
172.38	111.69	114.25	107.78	25.69	56.21	48.85	29.15	660.12	3159.96	2386.19	849.98
176.25	125.94	109.56	109.69	27.78	58.69	49.27	33.52	771.80	3444.86	2427.16	1123.64
183.94	144.50	106.34	117.22	20.06	57.19	50.15	37.42	402.60	3271.07	2515.20	1400.05
200.50	171.50	106.56	121.25	10.35	53.43	50.52	41.30	107.20	2854.67	2552.32	1705.94
232.19	202.63	110.25	128.09	6.38	33.27	50.86	37.51	40.70	1107.18	2586.97	1406.99

Skewnes				Kurtosis			
Day-2	Day-3	Day-5	Day-7	Day-2	Day-3	Day-5	Day-7
1.22	−0.03	−0.58	−0.09	4.64	1.59	2.16	1.29
1.26	−0.07	−0.31	0.38	3.95	1.37	1.57	1.37
1.09	0.28	0.06	0.69	3.31	1.39	1.39	1.90
1.16	0.49	0.42	1.11	3.65	1.71	1.61	2.89
1.40	0.69	0.68	1.47	4.45	1.85	1.96	3.83

(continued)

Table 2 (continued)

Skewnes				Kurtosis			
Day-2	Day-3	Day-5	Day-7	Day-2	Day-3	Day-5	Day-7
1.25	0.73	1.05	1.78	4.47	2.12	2.65	5.09
1.35	0.71	1.48	1.95	4.95	2.00	3.83	6.72
1.39	0.70	1.84	2.35	5.55	2.19	5.15	9.98
1.46	0.99	1.97	2.26	4.68	2.73	5.99	10.62
0.60	1.03	1.67	2.63	2.57	2.65	5.32	13.47
0.57	1.41	1.51	2.87	2.06	3.32	5.02	14.98
1.14	1.38	1.65	3.21	3.87	3.37	5.52	16.11
0.87	1.07	1.75	2.13	3.65	2.47	5.78	10.17
0.60	0.56	1.76	1.35	3.50	1.60	5.71	6.00
2.13	0.04	1.72	0.78	7.20	1.40	5.56	4.00
0.68	−0.92	1.57	1.01	3.68	2.54	5.21	4.08

Table 3 Sample second-order statistical methods

	Contrast	Correlation	Energy	Homogeneity
Day-1	0.3302	0.8545	0.2224	0.8967
Day-2	0.3345	0.8910	0.2001	0.8836
Day-3	0.3453	0.8926	0.1675	0.8847
Day-4	0.3987	0.9371	0.1554	0.8554
Day-5	0.4022	0.9547	0.1406	0.8358
Day-7	0.4294	0.9766	0.1203	0.8337

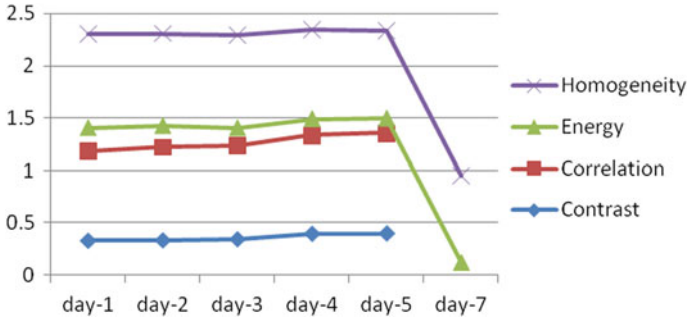


Fig. 21 Performance analysis graph of second-order statistics

Fig. 22 ROC curve for image feature extraction using CNN

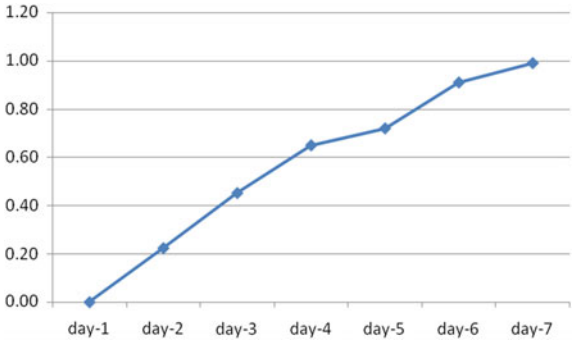


Table 4 Performance evaluation

Techniques	Images tested	Images recognized	Accuracy (%)
First- and second-order statistical features	220	182	82.73
Convolutional neural network	220	202	91.82

7 Conclusion

In this chapter, we tried to explain from the basics of algae image classification in detail. We have applied the first-order statistical measurements such as mean, standard deviation, variance, median, mode skewness, and kurtosis for training and testing images, and the performance are analyzed. We have extracted the second-order statistical measurements of the given image such as ASM, Entropy, ISM and correlation features and those features are fed into the various neural networks and then evaluated the performance. The Convolution neural network under deep learning method is used to classify the images for measuring the growth of the algae in ponds and the ROC curve is drawn. We are getting the outstanding results when we apply CNN method. Further experiments should be done to extract the features for image classification under deep learning in detail.

References

1. Castelluccio M, Poggi G, Sansone C, Verdoliva L (2016) Land use classification in remote sensing images by convolutional neural networks. Available online: http://arxiv.org/abs/1508.00092. Accessed on 12 Apr 2016
2. Couprie C, Farabet C, LeCun Y, Najman L (2013) Indoor semantic segmentation using depth information. In: Proceedings of the international conference on learning representation. Scottsdale, Arizona, 2–4 May 2013
3. Elaksher AF (2008) Multi-image matching using neural networks and photogrammetric conditions. In: The international archives of the photogrammetry, remote sensing and spatial information sciences, vol XXXVII. Part I, B3a. Beijing 2008
4. Farabet C, Couprie C, Najman L, LeCun Y (2013) Learning hierarchical features for scene labeling. IEEE Trans Pattern Anal Mach Intell 35:1915–1929
5. French M, Recknagel F, Jarrett GL (1998) Scaling issues in artificial neural network modelling and forecasting of algal bloom dynamics. In: Abt SR, Young Pezeshk J, Watson CC (eds) Proceedings of the international water resources engineering conference, ASCE, vol 1. Memphis, Tennessee, 3–7 Aug 1998, pp 891–896
6. Gonzalez RC, Woods RE (2008) Digital image processing. Gatesmark Publishing
7. Gonzalez RC, Woods RE, Eddins S (2009) Digital image processing using matlab. Gatesmark Publishing
8. Hu F, Xia GS, Hu J, Zhang L (2015) Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens 7:14680–14707
9. Jähne B (2002) Digital image processing. 5th rev. Springer, London
10. Joseph HWL, Huang Y, Dickman M, Jayawardena AW (2003) Neural network modeling of coastal algal blooms. In: Ecological Modelling, vol 159. Elsevier Science B.V., pp 179–201
11. Junna C, Ji G, Feng C, Zheng H (2009) Application of connected morphological operators to image smoothing and edge detection of algae. In: International conference on information technology and computer science, pp 73–76
12. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: NIPS, p 4
13. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; Curran Associates, North Miami Beach, FL, USA, pp 1097–1105

14. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. IEEE Proc. 86:2278–2324
15. Lee HS, Lee JHW (1995) Continuous monitoring of short term dissolved oxygen and algal dynamics. Water Res 29(12):2789–2796
16. Lek S, Guegan JF (1999) Artificial neural networks as a tool in ecological modeling, an introduction. Ecol. Model. 120:65–73
17. Merchant RE, Andre CA (2001) A review of recent clinical trials of the nutritional supplement Chlorella pyrenoidosa in the treatment of fibromyalgia, hypertension, and ulcerative colitis. Altern Ther Health Med 7(3):79–91. Review. PubMed PMID: 11347287
18. Mohammed TS, Al-Taie NI (2012) Artificial neural networks as decision-makers for stereo matching. GSTF Int J Comput 1(3)
19. Parsons TR, Takahashi M, Hargrave B (1984) Biological oceanographic processes. Pergamon, Oxford
20. Penatti OA, Nogueira K, dos Santos JA (2015) Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, Boston, MA, USA, 7–12 June 2015
21. Phung SL, Bouzerdoum A. MATLAB library for convolutional neural network. Technical report. ICT Research Institute, Visual and Audio Signal Processing Laboratory, University of Wollongong. Available at: http://www.uow.edu.au/~phung
22. Quigley M, Batra S, Gould S, Klingbeil E, Le QV, Wellman A, Ng AY (2009) High-accuracy 3D sensing for mobile manipulation: improving object detection and door opening. In: Proceedings of the IEEE international conference on robotics and automation (ICRA), Kobe, Japan, 12–17 May 2009, pp 2816–2822
23. Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: an astounding baseline for recognition. In: 2014 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 512–519
24. Recknagel F, French M, Harkonen P, Yabunaka K (1997) Artificial neural network approach for modeling and prediction of algal blooms. Ecol Model 96:11–28
25. Sandesh BK, Shalini C, Brinda BR, Kumar MA (2005) Digital image processing—an alternate tool for monitoring of pigment levels in cultured cells with special reference to green alga Haematococcus pluvialis. Biosens Bioelectron 21:768–773
26. Shanmugam K, Haralick RM, Dinstein IH (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 3:610–621
27. Socher R, Huval B, Bath B, Manning CD, Ng AY (2012) Convolutional-recursive deep learning for 3D object classification. In: Advances in neural information processing systems. Curran Associates, North Miami Beach, FL, USA, pp 665–673
28. Thomann RV, Mueller JA (1987) Principles of surface water quality modeling and control. Harper and Row, New York
29. Tutorial on deep learning [Online]. Available at: http://deeplearning.net/tutorial/lenet.html
30. Whitehead PG, Howard A, Arulmani C (1997) Modelling algal growth and transport in rivers: a comparison of time series analysis, dynamic mass balance and neural network techniques. Hydrobiologia 349:39–46
31. Wikipedia. https://en.wikipedia.org/wiki/Convolutional_neural_network

Review on Image Enhancement Techniques Using Biologically Inspired Artificial Bee Colony Algorithms and Its Variants

Rehan Ahmad and Nitin S. Choubey

Abstract When medical images are processed by morphological operations, they provide substantial amount of utilizable information. The technological advancement in the field of image analysis and medical imaging domain acquiesces the understating of detection and diagnosis of disease to enhance the quality of medical treatment. Application of image processing in medical imaging field administers the development of processing nebulous, skeptical, reciprocal, superfluous information, and data for a vigorous structural attribute. To understand any image, the human and artificial astuteness system matches the features extracted from an image. Image enhancement is a decisive stage in image processing system. It intents at convalescencing the ocular data and the informational trait of wry images. After the acquisition of an image, if it is of poor quality, it requires enhancement. Various available techniques can be applied for enhancement; some are providing good results with limitation of computing time. A new intelligent algorithmic approach, based upon biologically inspired approaches, is suggested for image enhancement. In this ambience, this article describes about one of the most commonly used algorithms known as artificial bee colony algorithm, and its various types, used for image enhancement in different subdomains of medical imaging, are covered here.

Keywords Artificial bee colony algorithm · Modified ABC algorithm · Multiscale retinex (MSR) using the artificial bee colony (ABC) algorithm · Swarm intelligence · Algorithmic structures of swarm intelligence ABC with neural network · ABC with SOM-based k-means clustering · ABC with BPNN · CABC with forward NN · HABC with KELM · Basic ABC · ABC with FCONN · ABC with LS-SVM · EABCO · ABC with radial basis function NN · ABC with ASF · ABC with FCM · ABC with Otsu method

R. Ahmad · N.S. Choubey (✉)
Faculty of NMIMS, Mukesh Patel School of Technology Management and Engineering, Shirpur, Maharashtra, India
e-mail: Nitin.choubey@nmims.edu; nschoubey@gmail.com

R. Ahmad
e-mail: rehan.ahmad@nmims.edu; electron3011@gmail.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_11

1 Introduction

When fishes, birds, insects such as termites and bees, etc., behave collectively for a particular purpose, then this gathering is called as swarm. The behavior of each ambassador of this class behaves independently without any supervision. And due to their neighborhood perception, they all have a stochastic behavior.

Swarm intelligence (SI) is the terminology used, when swarm works in a self-organized and decentralized but collectively manners. The most common prototypes are social insects such as bees, ants, and bird flock. The acumen of the swarm depends upon the web of synergy between agents and the environment. Computer scientist, engineers, research scholars of bioinformatics are exploring new horizons of SI. The ability to handle complex problems such as dividing labors among fellow nest mates, building new nests, finding new food sources. All these have important correlation with real-world application in engineering domain. Two techniques are suggested based on the behavior of ant, called as ant colony optimization (ACO) by Dorigo and Colorni [1], and on bird flocking, called as particle swarm optimization (PSO) by Kennedy and Eberhart [2]. Both the techniques have been explored on the basis of several parameters by the researchers and their several variants have been implemented in real world for solving complex problems.

The self-organized division of labor features explained by Bonabeau and Dorigo [3] and the satisfaction principles by Millonas [4], represented by honey bee, is a prime requirement of SI. The survey on artificial bee colony algorithm covering different fields was first written by Karaboga and Akay [5]. The survey affirms that, based upon the intelligent behavior of honey bees, many algorithms have been devised for many applications. These studies comprise applications such as collective decision, foraging, nest site selection, and task allocation. Some of the algorithms based on bee swarm intelligence use virtual bee, the bees, ad hoc bees, honey bee marriage, hive pattern of bee, functioning system of bee, colony optimization process of bees, and artificial bee colony (ABC) algorithm.

2 Artificial Bee Colony Algorithm Approach

2.1 Prevailing Characteristics of Swarm Intelligence

Many categories of swarms do exist in this world. All are not intelligent or their brilliance level differs from crowd to crowd of swarm population. Self-organization is a distinctive trait which is the basis of collective behavior among agents, discovered by Bonabeau [3], Tereshko and Loengarov [7], Bonabeau also interpreted the four characteristics of self-organization in swarms:

(i) Positive feedback: It advocates the formation of suitable structures. Few examples are reinforcement and recruitment such as establishing a trail and pursuing in some ant species.

(ii) Negative feedback: A negative feedback mechanism is required to avert the saturation of available foragers. It also balances the positive feedback.
(iii) Fluctuations: Few activities between swarms for creativity such as random task, random walk. Randomness is often symbolic for developing structures for finding new solutions.
(iv) Many interactions: For the transmission of information over the entire network, swarm exchanged information with each other.

A swarm to be called intelligent, when it fulfills the below-stated guidelines of Millonas [4]:

(i) The proximity principle states that a swarm should have the capability to perform time and space computations.
(ii) The swarm should have the capability to identify the quality factors in the environment.
(iii) The swarm should have the capability to avoid activities in excessively narrow channels.
(iv) The stability principle states that the swarm should not change its mode of behavior due to change in the environment.
(v) The swarm should have the ability to adopt change in behavior as per the requirement.

2.2 Food Exploration Procedure of Bees

Intelligent behavior of honey bees was discovered due to the foraging behavior. The foraging etiquettes of honey bees depend on three main elements: the food source, the employed honey bees, and the unemployed honey bees. The functioning and behavior of honey bees depend on two parameters: the election of supply of food source and the termination of low food depository.

(i) Fountain of Food:

The food source contents depends on various factors which include the richness (energy concentration), easiness of energy extraction, and proximity to the nest, Seeley [6, 7].

(ii) Employed foragers:

Employed foragers (bees) are related to a food source which are currently under mining process by them. The information carried by employed bee about the food source is distance, direction, and size of food source present. They share this information through the waggle dance to onlooker members [5–7].

(iii) Unemployed foragers:

These are always searching for food source, categorized as scouts foragers and onlooker foragers. The scouts are new volunteers for food search; around the

surrounding of the nest, at the same time onlooker bees are waiting for food source exploration, based upon the details distributed by employed foragers Seeley [6]. Scouts bees are in 5–10% in number, to the total population of bees of the nest. The entire hive divided in number of sections. Dancing area is a part of nest with high importance where the source of information regarding quality of food sources is shared with onlooker's bee through the dance, known as waggle dance. The onlookers are observing dance pattern, and then they decide to work for a fountain of food. The profitability of fountain of food is directly reciprocal to the recruitment of onlooker bees [5–7].

2.3 Organization of ABC Algorithm

This algorithm mainly comprises three types of honey bees: employed bees, which are constantly looking for new fountain of foods; onlooker bees, which observes the waggle movement of employed bees in the nest to decide selection of fountain of food; and scout bees which constantly looking for new food connection. Onlooker and scouts are also known as unemployed bees. The location of new food sources are explored by scout bees. These food sources are milked together by onlooker bees and employed bees, and because of this constant exploration process, food sources will lay off. And due to this, the employed bee turns to scout bee and keeps looking for new food sources.

In ABC process, a food represents the solution of a task, and its quality (fitness) represents the amount of a food of the related solution. The basic algorithm represent the, amount of food available is identical to the employed bee, the employed bee at a time is a part of only one food source. The architecture of the ABC algorithm [5, 8] is depicted in Fig. 1.

The scout bees initiate the control parameters setting as well as food source population. The employed bee starts searching for sources of food having large quantities than earlier one. First of all the fitness value of new food sources will be checked. Then, the information of food sources are shared by employed bee with onlooker bees through waggle dance at hive.

The onlooker bees select the food sources based on the optimal probability value. For the evaluation of probability, roulette wheel selection method is used.

Numbers of trials are used to find the optimal solution, if not found, then employed bees become scout bees, and the solutions of employed bees are terminated. Then, scout bee searches for new solutions randomly. The food source whose value becomes poor due to foraging process, or new food source with little values, due to such process negative feedback arises which can be balanced by positive feedback. These steps can be recycled till a level of convergence is achieved [8].

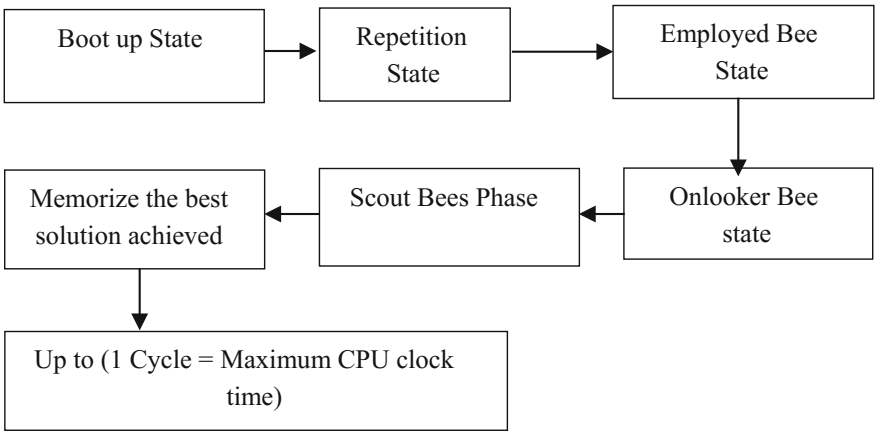


Fig. 1 Step-by-step procedure diagram of ABC algorithm process [1]

3 Applications of Biologically Inspired Artificial Bee Colony Algorithms and Its Variants in the Field of Health Care

Biomedical imaging display acquired images for purposes such as analysis, therapeutic, and diagnosis, in its pure digital format. Physiological images can be hoarded through ultrasound; magnetic resonance imaging (MRI); high-end sensors; and light technology such as endoscopy, OCT, X-rays, CT scan, mammogram hybridized with latest computer technology for the analysis, monitoring, and diagnostics of human organs and tissue during therapy and health evaluation process. For information gathering, medical imaging system can be categorized as morphologic or structural (such as CT, MRI, and OCT) and functional (such as PET, SPECT). Biomedical image processing remodels the second signal in order to construct the 3D images.

Whenever an image is transformed from one format to other, such as duplicated, flip through electronics devices, transferred, and advertised, the trait of the output image is of secondary than that of the input. Due to lack of wisdom about dealing with picture of degraded quality, it is troublesome to forecast in priory about the effectiveness of a specific improvement method. Image enhancement by Frank [9] intents to upgrade human perception level and deciphering information in an image to provide more useful input for automated image processing procedures. Image enhancement techniques can be categorized as follows:

1. Spatial-domain techniques which deal with pixels of an image.
2. Frequency-domain techniques use Fourier analysis or any other frequency-domain methods of image processing.

3. Few methods which are hybrid version of spatial- and frequency-domain techniques.

While applying enhancement process [9], care should be taken about the selection of best suitable technique. Therefore, the enhancement method(s) are application specific and often empirically developed. While using one of the image enhancement techniques for preprocessing process, few quantitative measures should be taken into account for the selection of most suitable enhancement technique.

Followings articles states the domain in which ABC algorithms and its variants for image processing had been applied in different sections of healthcare domain, for different types of images.

3.1 Magnetic Resonance Imaging [MR Images]

Dynamic contrast-enhanced (DCE) MR image is widely used in the clinical applications to diagnose breast cancer in identification of prognosis for women affected with malignant tumors. Computer-assisted evaluation (CAE) systems are used to help radiologists in the detection of cancer at the early stage of diagnosis. The success of a CAE system relies on selection of an appropriate segmentation function for separating malignant and benign lesions. Artificial neural network, support vector machine, and artificial bee colony algorithms are used for predicting lesion type on breast DCE-MR images applied by Sathya and Geetha [10]. A NN system is developed for cataloguing of suspicious lesions in breast MR images. The ANN classifier model has a three-layered NN, which consists of input layer, a hidden layer, and an output layer. The number of nodes in the input layer represents the number of input variables. The neural network classifiers are optimized by an artificial bee colony algorithm. This ANN classifier is a multilayer feedforward neural network. The selection of NN is due to its nonparametric statistical property. ABC is used for training an ANN to find the optimal set of weights of a neural network in the training process. The smallest error functions can be found out using ABC algorithm [10].

Magnetic resonance imaging is a diagnostic technique to obtain high-quality breast images. The prime advantage of it is noninvasive process; it possesses a high spatial resolution and an excellent contrast of soft tissues. MR images are widely used in applications such as diagnosis of tissue deformities study of breast pathology. Most of the breast images, affected by cancer, are captured through magnetic resonance imaging technique, because of its beauty of exploration of high amount of information of captured image. The performance of artificial bee colony (ABC) algorithm is compared with the enhanced SOM-based K-means clustering algorithm for segmentation of breast DCE-MR images by Sathya and Geetha [10]. This algorithm also provides a quantitative comparison of the performance of these image segmentation techniques with breast DCE-MR images. The artificial bee colony clustering is used for clustering the preprocessed image.

An ABC algorithm is used for clustering; bees represent a potential clustering solution as a set of k-cluster centers, for clustering the breast DCE-MR images. ABC clustering algorithm-based segmentation method offers good classifying results for the breast data. Another clustering method known as the enhanced self-organizing feature map-based k-means clustering algorithm is improved through the search of an optimized space in which to operate the clustering. It allows the ability to make the clustering methods to retain more information from the original image than the crisp segmentation methods. The SOM-based k-means clustering-based segmentation algorithm partially corrects the misclassified pixels. Hence, the clustering results of artificial bee colony clustering algorithm are superior to enhanced SOM-based K-means clustering algorithm [11].

Image can be easily affected by all kinds of noises at the time of acquisition and transmission, noise-affected images are considered of low-quality images, which sometimes may be discarded at higher-level image processing. Therefore, at the preprocessing stage of image, it is essential to conduct image denoising, in order to improve the signal-to-noise ratio of image, and highlights desired features of image. Image denoising is used to reinstate original image free from noise, along with to preserve the detailed information of the image. For this job a hybrid algorithm of artificial bee colony algorithm and BP neural network (ABC-BPNN) implemented by Wang and Zhang [12], opt the double circulation structure during the training process; after selecting the desired convergence speed and precision, it can alter the rules as per the requirement of structure and it automatically adjusts the number of neurons. The weight of the neurons is determined by bee colony algorithm. The output image is completely noise free [12].

Classification of brain images using magnetic resonance technology is a field of study in the field of neuroimaging. Normal and abnormal MR images of brain can be classified by forward neural network (FNN) to reduce features of FNN. Parameters are optimized using an improved artificial bee colony (IABC) algorithm by Zhang and Wu [13], which is renamed as scaled chaotic artificial bee colony (SCABC). This is improved based on parameters, fitness scaling, and chaotic theory. This method uses a discrete wavelet transform for features' extraction of an image. To reduce the size of the feature, a principle component analysis (PCA) procedure is implied. The *K*-fold stratified crossvalidation is employed to avoid over fitting. This algorithm is tested on the dataset of T2-weighted MR images consisting of 66 brain images. The results of SCABC are compared with traditional training methods such as genetic algorithm, elite genetic algorithm with migration, simulated annealing, and ABC. Least mean MSE and 100% classification are achieved by SCABC [13].

3.2 Breast Cancer

Mammography-based detection of breast cancer technique is in trend nowadays. Among the several detection parameters, microcalcification is used as a parameter;

it is available in the form of cluster and very small calcium structure present in the form of lines and circles present in the extra cell activity of breast tissues. Artificial bee colony algorithm is used to analyze and extract required information from a region under process using asymmetric approach by Jegdeesan and Suresh [14]. The improvement in sensitivity and reliability of the system for diagnosis of microcalcification clusters are the few merits of this method.

For the identification of breast cancer, circulating tumor cells (CTCs) in peripheral blood (PB) testing are used. Due to the presence of hundreds and thousands of indicative genes, identification of CTC is a complex process. The identification of CTC in breast cancer is used to solve gene selection by using a hybrid fuzzy online sequential particle swarm genetic (PSG) kernel extreme learning machine (FOP-KELM) classification by Mythili and Senthil Kumar [15]. This method estimates the mean values for each gene features and is compared with objective function of KELM to select and remove unimportant gene features. The selection of gene features from breast cancer gene dataset samples mean value is estimated for each feature and follows the procedure of KELM classification algorithm with equal kernel objective function. Artificial bee colony (ABC) algorithm is used to calculate the distance between two selected gene feature samples of breast cancer; these genes are classified and are grouped as individual cluster. This hybrid algorithm functions in three steps: In the first stage of the HABCA method, the gene samples from dataset are divided into the few samples, and in the second stage, distance value is estimated using ABC. And in the third stage, algorithm uses quadratic sums of squares (SSQ) function to cluster the features of dataset. To minimize the fuzzy membership value in ELM, and to optimize it, PSG algorithm is used. The impact of selected features from FOP-KELM has been explored using clustering technique. The HABCA algorithm extricates the CTC through the separation of tumor samples into a hierarchical tree structure in a top-down manner, and the distance between two gene tumor samples is found out by using basic ABC algorithm.

Because of the simplicity and highly efficient property of artificial bee colony algorithm, it can be used for classification of data in data mining process by Nazarian and Mashala [16]. If the reasons of breast cancer can be detected on time, its growth can be prevented and its diagnosis process can be simple by using the data related to it as a sample. On the basis of predefined parameter "precision," this method can be compared with other methods such as Naïve Bayes, neural network, decision tree C4.5, support vector machine [16].

For automated diagnostic systems of breast cancer tumors, a classifier uses a combined fuzzy clustering optimized neural network (FCONN) algorithm. FCONN is used for the classification of the breast cancer tumors using fuzzy c-means clustering (FCM) algorithm and neural network in an optimized way. The FCM is used for extraction of features, and ONN is used for classification. The Gbest-guided artificial bee colony (GABC) algorithm is used by Pourmandi and Addeh [17] for selection of the desired parameters of the classifier. As in neural network training, the hyper-parameters play vital roles for the accuracy recognition [17].

The hybrid artificial bee colony (ABC) algorithm with least squares support vector machine (LS-SVM), known as (ABC-SVM), is used by Mythili and Senthil Kumar [18] for identification of breast cancer in peripheral blood with microarray datasets. The most common breast cancer diagnosis technique is identifying circulating tumor cells (CTC) in the blood peripheral. A hybrid ABC-SVM procedure applied on available DNA microarray datasets from different sources. The objective is to extract gene signatures associated with pairwise differentiation between cell types and/or disease type. The comparison between cancer and control tissue shares knowledge about the discriminative factors of the primary disease. Hybrid artificial bee colony (ABC) approach attempts to explore the field by combining microarray gene expression data originated from tissue and PB. The ABC algorithm is used to obtain the optimal values of regularization parameter, c, and kernel RBF parameter, σ^2, embedded in LS-SVM toolbox and adopted a supervised learning approach to LS-SVM model for the identification and characterization of CTC. A hybrid classification method is used for the detection of CTC, between cancer blood and control PB in association with the primary and secondary disease. This procedure implicate circulating tumor cells in metastatic disease in metastatic breast cancer (MBC). The objective is to identify tissues in peripheral blood that can reveal the presence of cancer cells [18].

Breast cancer is one of the most important health problems in Western countries, and certainly, it is the most common type of cancer in women. The sources of breast cancer are breast tissue, most commonly from the inner lining of milk ducts or the milk tube. Mammogram is a technology used for diagnosing breast cancer. The enhanced artificial bee colony optimization (EABCO) is used by Sivakumar and Marcus [19] to identify the breast border and nipple position to locate the suspicious regions on digital mammograms, which uses bilateral subtraction between left and right breast image. The EABCO is used to detect the border and to find the nipple position. True-positive detection rate and the number of false-positive detection rate at various thresholds of the asymmetry images are the metrics used for performance evaluation of this technique [19].

3.3 Brain Images

For the classification of single-trial EEG data, brain–computer interface (BCI) is preferred. The artificial bee colony (ABC) algorithm is used by Hsu and Hu [20] for the choice of substantial features such as band power, autoregressive model, and coherence and phase locking values are taken out then grouped. Feature selection is used to improve the classification accuracy of single-trial electroencephalogram (EEG). The artifacts and background noise can be removed by surface Laplacian filter. Subfeatures are classified by support vector machine. Comparing with and without artifact removal and feature selection, using a genetic algorithm the obtained results indicate that this system provides promising results and is suitable for brain–computer interface applications [20].

For detection of brain disorder diseases, electrical signals generated by brain can be recorded, processed, and analyzed, these electrical signals are known as electroencephalogram (EEG) signals. To analyze the epileptic disorder in human brain through EEG signal by integrating, the best attributes of artificial bee colony (ABC) and radial basis function networks (RBFNNs) are used by Satpathy and Jagadev [21]. For classification of these signals, the RBFNNs have been trained by a modified version of ABC algorithm. In the modified ABC, the onlooker bees are selected based on binary tournament. Metrics such as Gaussian, multiquadric, and inverse-multiquadric are used for measuring the effectiveness of the method in numerous mixtures of healthy segments, seizure-free segments, and seizure segments. The test results show that RBFNN with inverse-multiquadric kernel trained with modified ABC is significantly better than RBFNNs with other kernels trained by ABC and modified ABC [21].

At time of brain–computer interface, for collecting the signals noninvasive techniques are preferred, because of several advantages. One of the most commonly used noninvasive techniques is electroencephalogram (EEG). Localized sources from the outer cortex of the brain are used to generate the EEG signal. Electrodes are used to acquire the signal. Each EEG electrode carries two signals, one original EEG signal and another is an unwanted signal (i.e., noise). Multiple electrodes deployed at several locations of brains to capture EEG signals; however, this placement plays vital role in the acquisition of EEG signals. Artificial bee colony algorithm with adaptive scale factor (ABC-ASF) optimization based optimal electrode selection strategy from which the acquired EEG signals encompass the major part of brain activities is involved in a task by Shreyasi and Pratyusha [22]. In ABC-ASF, on the basis of previous experience, the scale factor for mutation is traditional. At each generation, a set of scale factor value is separately assigned to each individual bee of the current generation according to the selection probabilities learned from their previous generations. This algorithm estimates the optimal position of electrode and gives promising results with respect to accuracy and run-time complexity [22].

To reduce the training time and to interrupt the classification model feature process are adopted in medical image processing. Artificial bee colony (ABC) is used for feature selection in CT scan images of cervical cancer to detect whether the data given as an input are cancerous or not. To obtain region of interest (ROI) from the segmented image, a textural feature algorithm is used. The classification of segmented featured using two hybrid algorithms, ABC with k-nearest neighbor (k-NN) algorithm, ABC, with support vector machine (SVM) by Agrawal and Chandra [23].

One of the techniques of brain–computer interface (BCI) is P300. It emulates processing of brain in stimulus events. Based on the patient's condition and their present cerebral loads, P300 components occur at posterior to stimuli from 250 to 600 ms. This change affects the administration of BCI. An estimation model that estimates an appropriate interval for P300 feature extraction is used, and an interval type-2 fuzzy logic system trained by artificial bee colony algorithm is used by Kuo and Kuo [24] to find the latency of elicited P300 with a certain range by means of

steady-state visually evoked potential. A support vector machine classifier is preferred to divide extracted epochs into target and nontarget stimuli [24].

3.4 Brain Tumor

For extraction and detection of a specific region in medical imaging, segmentation plays a vital role in it. Artificial bee colony (ABC) algorithm is used to extract brain tumors from magnetic resonance imaging using the segmentation process by Hancer and Ozturk for diagnosis and treatment of medical issues. The process involves three phases, preprocessing (enhancement) of the original MR image, processing (segmentation) by ABC algorithm-based clustering method, and post-processing (extraction) of tumors. This algorithm is tested on nine MR images captured in different positions from a patient with the methodologies based on fuzzy c-means, genetic algorithms, and K-means clustering methods. The segmentation process with the ABC algorithm provides both visually and numerically best results [25].

Tumor segmentation of MR brain images based on artificial bee colony (ABC) algorithm and fuzzy c-means (FCM) algorithm is a segmentation approach implemented by Menon and Ramakrishnan [26]. ABC algorithm is used for segmentation, and the value in continuous grayscale is inspected using threshold estimation. ABC algorithm is also used to look for optimal threshold value. In order to get an efficient fitness function for ABC algorithm, the original image is decomposed by discrete wavelet transforms. A filtered image is reconstructed with low-frequency components by removing the noise from it. The FCM algorithm is used for clustering the segmented image which assists to identify the brain tumor. The low noise content is the parameter which surpasses over other methods. Image after clustering illustrates the strength of the tumor present in the brain. This method also provides the intensity of tumor [26].

Using image segmentation procedure, pulmonary parenchyma can be detected from multisliced CT images. 2D Otsu algorithm is mainly used for image segmentation but suffers with issues such as high computational time, nonlinear multimodal global optimization problem, high response time. To deal with all these issues, a method which is the combination of modified ABC algorithm known as adaptive ABC [AABC] and 2D Otsu method is used by Kumar and Sharma [27]. AABC is applied to determine the optimized values of threshold parameters, and Otsu is used to generate the optimized function [27].

Biomedical photoacoustic (PA) signal is characterized with extremely low signal-to-noise ratio which yields significant artifacts in photoacoustic tomography (PAT) images. An adaptive method to improve the quality of PA imaging based on empirical mode decomposition (EMD) and reconstruction is used for the data acquired by ultrasound transducers, is adaptively decomposed into several intrinsic mode functions (IMFs) after a sifting preprocess. Noise is randomly distributed in different IMFs, depressing IMFs with more noise while enhancing IMFs with less

noise can effectively enhance the quality of reconstructed PAT images. Searching optimal parameters by means of brute-force searching algorithms requires high computational time, because of which practical applications are limited. To find parameters within reasonable time, two of the heuristic algorithms, simulated annealing algorithm, a probabilistic method to approximate the global optimal solution, and artificial bee colony (ABC) algorithm, an optimization method inspired by the foraging behavior of bee swarm, are selected to search optimal parameters of IMFs used by Guo and Ding [28]. The effectiveness of this method was proved on the basis of simulation results for both simulated data and PA signals from real biomedical tissue, which bear the potential for clinical PA imaging denoising [28].

3.5 Heart Diseases

ABC algorithm in conjunction with feedforward neural network can be used by Gaddala and Vanamala [29] for diagnosis of heart disease. Heart disease problems from a repository are chosen to evaluate the performance of ABC algorithm. A total of 25% of data are used for testing purpose and 75% of data are used for training set [29].

Forecasting coronary heart disease using robotic technology is an upcoming trend in the field of medical science, and it promotes use of artificial intelligence in health care. Large amount of patient data can be processed with the help of data mining algorithms, and this information can be evaluated through various available soft computing techniques, to guess the presence of disease. The accuracy of this forecasting is computed by artificial bee colony optimization algorithm (ABC algorithm). The more the accurate guess, the more the chances of right detection of disease by Sowmya and Vijayabhanu [30].

For diagnosis of arrhythmias automatically ECG heartbeat classification, a computer-aided ECG analysis implies a modified artificial bee colony (MABC) algorithm by Dilmac and Korurek [31]. ECG dataset from MITBIH is used for testing, and the obtained results are compared with seventeen other classifiers' results to check the accuracy of this method. Classification success rate of 99.30% is achieved using the MABC algorithm. The unbalanced sample numbers are the prime parameters which affect the classification of different datasets [31].

For detection of coronary artery disease, an algorithm combination of artificial bee colony (ABC) and k-nearest neighbor algorithm is suggested by Babaoglu and Kiran [32]. A total of 134 healthy and 346 unhealthy patients were processed to determine the stress level. This algorithm provides two centroid vectors, one for healthy patients and another for unhealthy patients. ABC trains the dataset. The test part of the dataset is classified using k-nearest neighbor algorithm. This approach could be used as an alternative classification tool, as an alternative classifier for diagnosis of coronary artery disease consists of exercise stress test data [32].

A hybrid feature extractor is used to categorize the abnormal ECG signals, such as atrial premature beat (APB), nodal (junction) premature beat (NPB) along with the normal beat, right bundle branch block beat (RBBB), premature ventricular contraction (PVC), left bundle branch block beat (LBBB). Hybrid feature extraction can be applied using morphological-based feature extraction scheme and Haar wavelet-based feature extraction scheme, for classification of ECG signals. After the feature extraction process, a feedforward neural network (FFNN) classifier classifies the signals according to their occurrence. Artificial bee colony (ABC) combined with genetic algorithm is used for training the neural network. A crossover rate is decided to achieve higher accuracy by Suresh and Panigrahi [33].

Artificial bee colony (ABC) algorithm-based least squares support vector machine (LS-SVM) classifier, a hybrid algorithm, implemented by Jain and Rai [34], is used to segregate ECG signals. A radial basis function kernel for heart monitoring systems is used for alarming an emergency call due to sudden cardiac disorder of patients. An automatic ECG beat classifier is used to share the bandwidth features extracted from analytic intrinsic mode functions (IMFs). Bandwidth features comprise amplitude modulation bandwidth (BAM), and frequency modulation bandwidth (BFM) obtained from ECG beats. By applying Hilbert transform on IMFs, bandwidth features BAM and BFM are extracted from ECG beats and applied as an input to the hybrid ABC-LSSVM classifier for automatic classification of normal and abnormal classes. In this classifier, the ABC algorithm is used to make the classifier automatic by optimizing the parameters of LS-SVM. This method assists automatic detection of abnormal beats, which can help in alarming emergency situations for heart patients. ABC with LS-SVM classifier that efficiently exploits bandwidth features for automatic classification of ECG beats decides the outcome of process [34].

3.6 Protein Structure and DNA Images

Forecasting the secondary structure of a protein is a challenging task in bioinformatics because of increasing computational complexity. To improve the exploitation property of artificial bee colony (ABC) algorithm, an internal feedback strategy-based ABC (IF-ABC) is used by Li and Li [35]. Improved ABC revised by internal feedback strategy is used to optimize protein secondary structures in AB off-lattice model. Internal states are fully utilized in each of the iterations to guide search process and to balance local exploration with global exploitation. Experiment was conducted on less than 200 samples of amino acids and provides promising results compared to nPERM, GAA, E-PSO, and I-TS techniques [35].

One of the problem statements of global optimization problem is forecasting the 3D architecture of a protein from the sequence of amino acid. A chaotic artificial bee colony (CABC) algorithm is enforced on 3D protein architecture forecasting, based on the 3D off-lattice structure of AB model, by Wang and Guo [36]. The

problem of premature convergence can be warded off by mixing global search and local search capability of ABC algorithm with the chaotic search algorithm. The CABC algorithm is applied to deal with multiextremism and multiparameter optimization problem for 3D protein structure prediction. The off-lattice AB model adopts the chaotic search algorithm to make the constant solution jump out of the local optimum, which allows the diversity of the population and avoid premature convergence of the algorithm. The obtained results of this algorithm when compared with the CABC algorithm provide a good global search capacity and better performance, compared to other available techniques [36].

A hybrid classification model consist of support vector machine (SVM) with ABC algorithm is used by Alshamlan and Badr [37] to measure the classification accuracy for selected genes. SVM is used as a classifier to solve classification issues that deal with high dimensional datasets. ABC algorithm is used for analyzing microarray dataset and for microarray gene expression profile. The ABC-SVM algorithm is an approach for solving gene selection and cancer classification problems and provides the highest classification accuracy along with the lowest average of selected genes. ABC-SVM algorithm consists of two stages: gene selection step (ABC) and cancer classification step (SVM). The ABC algorithm is used to adopt the predictive and informative genes from microarray dataset. The SVM classifier is trained and tested using the selected genes and provides high classification accuracy [37].

Predicting the biological structure of a protein is considered as an important task. The artificial bee colony (ABC) algorithm is used for prediction of structure for protein using three-dimensional hydrophobicpolar model with side chains (3DHP-SC) used by Cesar and Benıtez [38]. This method uses two approaches for the operation of ABC: a master–slave and a hybrid hierarchical method. The master–slave (MS–ABC) implies a global single-population system where a master process distributes the processing load into several slave processors, each one running in a different cluster-based processing environment. The master will initialize the food positions, and it generates new solutions in the algorithm cycles, by applying a greedy selection procedure, and distributing bees to the slaves. Slaves are bound for computing the fitness function of received bees. The computation of the fitness function is required for distributed computing. The hybrid hierarchical (HH–ABC) operates in two levels. In the upper level, multiple-population coarse-grained islands are formed. In the lower level, global single-population master–slaves are created. This combination is to gain advantage of the benefits of both models in a single approach [38].

The images of soft tissues in an organ can be captured using ultrasound imaging technique. Ultrasound devices have merits such as noninvasive process, cheap in price, nonharmful for all, adaptable. Along with this, the imaging techniques suffer from demerit of speckle noise due to inbuilt phase sensitive transducers. This noise degrades image quality and influences the diagnostic decisions in ultrasound systems. The structure of speckle noise is random, granular. A 2D FIR filter is used for noise removal. Artificial bee colony (ABC) algorithm is designed by Latifoglu [39] used to optimize the filter coefficients. During the optimization process, the filter

coefficients are tested with different numbers and connection types such as cascade and parallel for the optimum design of filter. To determine the image quality, mean squared error (MSE), peak signal-to-noise ratio (PSNR), and signal-to-noise ratio (SNR) are used as a metric. The performance of 2D FIR filters is designed based on ABC optimization tested on ovarian and liver tissues for the removal of speckle noise [39].

A hybrid algorithm which has modified DNA concepts and artificial bee colony (ABC) algorithm is used to classify hierarchical fuzzy. The mechanism of symbiosis applied to create a hybrid modified DNA-ABC algorithm for hierarchical fuzzy classification application is implemented by Feng and Li [40]. The partition number and the shape of the membership function are extracted by the symbiosis-based hybrid modified DNA-ABC optimization algorithm, due to which sufficient global exploration and adequate local exploitation for hierarchical fuzzy classification are possible. This algorithm is applied on five-benchmark University of Irvine (UCI) datasets, and the results prove the efficiency of the algorithm. The drawback of this algorithm is high calculation time to achieve high system performance [40].

Microarray or gene expression profiling is applied to compare and determine the gene expression level and pattern for different cell types or tissue samples. The more informative genes, that is responsible for causing a specific disease or cancer. Gene expression measures the activity degree for gene in a given tissue of the body. The accurate classification of genes is helpful for detection of disease and gives an idea to doctor to decide about curing procedure. ABC algorithm is used for analyzing a microarray gene expression profile. A feature selection algorithm, minimum redundancy maximum relevance (mRMR), combines with an ABC algorithm, mRMR-ABC, is used by Alshamlan and Badr to select useful genes from microarray profile. The accuracy of specific genes can be classified using mRMR-ABC. With small number of predictive genes when tested datasets using this method and compared with previously suggested methods, it has been found this method provides accurate cancer classification and gene selection [41].

DNA microarray analyzes level of millions of genes. The gene level indicates the amalgamation of different messenger ribonucleic acid (mRNA) molecule in a cell. This gene level is used for diseases diagnosis, tumor identification, selection of treatment to avoid happening of illness. This technique unable to identify the classes of diseases with respect to their gene level. In any classification technique, the process of recognition and learning becomes crucial due to availability of few samples and due to large number of genes. For classification, approximation, and prediction, a function artificial neural network (ANN) is used. Some of the most prominent techniques are multilayer perceptron (MLP), the radial basis function neural network (RBF), and support vector machine (SVM). A methodology for classifying DNA microarray performs a feature selection process based on a swarm intelligence algorithm to find classes of genes that describe a disease. Four datasets of different categories are used to validate the accuracy and to classify the samples of disease based upon relevance of genes. The hybrid version of ABC algorithm with Euclidean distance classifier is used by Garro and Rodríguez [42] for the

selection of best set of genes. Performance evaluation involves how an ANN, is trained with the set of genes are identified by the ABC algorithm, enhances the capability of forecasting and classification of the disease samples [42].

3.7 Liver Images and Lung

The segmentation of liver images is a crucial phase in liver disease diagnosis in medical image processing. A clustering method for segmentation of liver images uses artificial bee colony (ABC) algorithm Mostafa and Fouad [43]. The cluster centroid can be calculated by ABC algorithm. Liver image consist of few regions which obstruct the processing such as flesh regions which are small and thin regions around the liver area, sharp edges of organs, and small regions inside the liver are removed by morphological operations. The obtained regions are unified to give an initial estimate of the liver area. This process is extended using a region growing approach. To calculate the performance of this images a set of 38 images are taken and the similarity index is calculated. Segmentation accuracy offered by this method is of 93.73% [43].

The performance of K-means clustering method depends upon the selection of cluster centroids, in the local optimal solution. A hybrid version of artificial bee colony algorithm and K-means algorithm is used by Lin and Peng [44] for the diagnosis of liver function. This algorithm minimizes the dependency on the initial cluster centroids and the probability to be trapped by local optimal solution, by efficiently assigning data points to their appropriate cluster. The test results indicate that this algorithm is preferable over K-means clustering for diagnosis of liver function [44].

Lung cancer can be defined as uncontrolled development of cell in the tissues of lung. These uncontrollable cells inhibit the growth of healthy tissues. The growth of tissues is the basic reason of tumors. Detection of this unwanted growth of tissues is crucial in deciding the procedure for curing the disease. A lung cancer prediagnosis technique is the combination of feedforward backpropagation neural network (FFBNN) and artificial bee colony (ABC) designed and implemented by Balachandran and Anitha [45]. To reduce computational complexity, ABC algorithm minimizes the specifications of the dataset. The risk factors and the symptoms from the dimensional reduced dataset are the key factors for the training process. At prediagnosis process, the FFBNN parameters are optimized using ABC algorithm for high accuracy. More data are given to FFBNN-ABC to validate the given testing of data for accurate prediction of lung disease during the training phase [45].

3.8 Diabetics

Lack of insulin is one of the reasons of diabetics. Pancrea is the organ which produces insulin to help metabolize blood sugar and maintain blood glucose level within a healthy range. Diabetic patients are incapable to generate insulin and cannot remove glucose from the bloodstream. For the detection of diabetic, artificial bee colony (ABC) algorithm by Beloufa and Chikhi [46] uses a mutation operator for performance enhancement. The diversity of ABC can be maintained without compromising the solution quality; to achieve this, a blended crossover operator of genetic algorithm is applied. The performance of the proposed method is evaluated by parameters such as classification rate, sensitivity, and specific values using tenfold crossvalidation method [46].

Detection of glucose levels is an interesting domain for food industry, health, and nutrition. The basic application in food industry is determination of energy content and nutritional facts of the food and determination of the plant origin contents in the food of animal origin. Other few applications includes, glucose detection is to identify blood glucose levels in a diabetic patient. It is important to keep the blood glucose concentrations in normal levels for humans suffered from diabetes disease. A modeling approach based use of fuzzy reasoning mechanism to define a measured dataset obtained from an optical sensing circuit, using optical sensor to measure glucose content of an aqueous solution. The measured data consist of analog voltages representing the absorbance values of three wavelengths measured from an RGB LED in different glucose concentrations. To achieve a desired model performance, the parameters of the fuzzy models are optimized by using the artificial bee colony (ABC) algorithm by Saracoglu and Bagis [47]. Simulation results show that the computational requirements can be significantly eliminated by using the ABC algorithm-based fuzzy logic models. A modeling performance having the minimum mean squared error (MSE) of 0.0013 is achieved through this method [47].

Sr. No.	Type of ABC algorithm	Year of publication	Problem dealt	Dataset used	Area of applications	Reference
1	Basic artificial bee colony algorithm used with support vector machine and artificial neural network [ABC-SVM NN]	2013	For predicting lesion type on breast DCE-MR images	Diagnose breast cancer in identification of prognosis for women affected with malignant tumors	Detection of tumor in breast cancer	[10]
2	Artificial bee colony algorithm with BP neural network [ABC-BPNN]	2013	Noise removal	Low-quality images affected by noise	Brain images, breast images	[12]

(continued)

(continued)

Sr. No.	Type of ABC algorithm	Year of publication	Problem dealt	Dataset used	Area of applications	Reference
3	Artificial bee colony algorithm with forward neural network [ABC-FNN]	2011	Enhancement using classification	Brain images for neuroprocessing	Classification of brain images in neuroimaging	[13]
4	Hybrid artificial bee colony algorithm [HABC]	2015	Breast cancer tissues	Breast cancer tissues, breast cancer images	For the identification of breast cancer circulating tumor cells (CTCs)	[14]
5	Basic artificial bee colony algorithm	2012	Breast cancer cells detection	Microcalcification	Breast cancer detection	[15]
6	Gbest-guided ABC [GABC]	2015	Selection of breast cancer cells based upon different parameters	Different types, shapes, sizes of breast cancer cells	Breast cancer cell detection among the other available cells	[17]
7	Hybrid artificial bee colony algorithm with least squares SVM [HABC-SVM]	2016	Identification of circulating tumor cells	DNA microarray datasets	Identification of breast cancer in peripheral blood with microarray datasets	[18]
8	Enhanced artificial bee colony optimization (EABCO) algorithm	2012	Using digital mammograms identification of breast border and nipple position to locate the suspicious regions	Different breast cancer images	Breast cancer detection	[19]
9	Artificial bee colony optimization (ABCO) algorithm	2015	Classification of single-trial EEG	Removal of artifacts and background noise	Noise removal from EEG images	[20]
10	Artificial bee colony (ABC) algorithm	2015	Feature selection in CT scan images of cervical cancer	Computerized topographic images	Cancer detection	[23]
11	Artificial bee colony (ABC) algorithm	2016	Processing of brain in stimulus events	To find the latency of elicited P300 images	Brain–computer interface	[24]
12	Artificial bee colony (ABC) algorithm	2014	Extraction of brain tumors from magnetic resonance imaging	Brain images with tumor	Extraction and detection of a specific region in medical imaging	[25]
13	Artificial bee colony (ABC) algorithm and fuzzy c-means (FCM) algorithm	2015	Continuous grayscale inspection using threshold estimation and finding optimal threshold value	Brain images	For detection of tumor	[26]

(continued)

(continued)

Sr. No.	Type of ABC algorithm	Year of publication	Problem dealt	Dataset used	Area of applications	Reference
14	Adaptive artificial bee colony (AABC) algorithm	2012	Detection of pulmonary parenchyma	CT scanned brain images	For detection of tumor	[27]
15	Artificial bee colony (ABC) algorithm	2016	To increase the search time on the basis of certain parameter	Simulated data and PA signals from real biomedical tissue	To identify artifacts in photoacoustic tomography (PAT) images	[28]
16	Artificial bee colony (ABC) algorithm	2016	Detection of abnormalities in functioning of heart	Coronary images	Diagnosis of heart disease	[29]
17	Modified artificial bee colony (MABC) algorithm	2015	Diagnosis of arrhythmias of ECG, heartbeat classification	ECG dataset from MITBIH	Detection of abnormalities in heart	[31]
18	Artificial bee colony (ABC) and k-nearest neighbor algorithms	2013	For detection of coronary artery disease	134 Healthy and 346 unhealthy patients were processed	To determine the stress level	[32]
19	Artificial bee colony (ABC) algorithm based least squares support vector machine (LS-SVM) classifier	2016	To segregate ECG signals	Real-time ECG data from patient	Heart monitoring system	[34]
20	Artificial bee colony (ABC) algorithm with internal feedback strategy-based ABC (IF-ABC)	2014	Identification of protein structure	Experiment was conducted on less than 200 samples of amino acids	For forecasting the secondary structure of a protein	[35]
21	Chaotic artificial bee colony (CABC) algorithm	2013	Forecasting the 3D architecture of a protein from the sequence of amino acid	3D off-lattice structure of AB model	Forecasting the 3D architecture of a protein for disease detection	[36]
22	Artificial bee colony (ABC) algorithm with support vector machine (SVM)	2016	Measurement of classification accuracy for selected genes	Genes data and cancer data	Gene selection and cancer classification problems	[37]
23	Master–slave artificial bee colony (MS–ABC) algorithm	2010	Prediction of structure for protein using three-dimensional hydrophobicpolar model with side-chains	Patients protein data	Prediction of biological structure of a protein for disease detection	[38]

(continued)

(continued)

Sr. No.	Type of ABC algorithm	Year of publication	Problem dealt	Dataset used	Area of applications	Reference
24	Artificial bee colony (ABC) algorithm	2013	Removal of noise from ultrasound images	Ovarian, liver tissues for the removal of speckle noise	Ultrasound imaging	[39]
25	Minimum redundancy maximum relevance algorithm with artificial bee colony (mRMR-ABC) algorithm	2015	To select useful genes from microarray profile, cancer classification, and gene selection	Gene expression profile of patients	To determine the gene expression level, pattern for different cell types, and tissue samples for cancer detection	[41]
26	ABC algorithm with Euclidean distance classifier	2016	Classification of DNA microarray, feature selection		Gene levels are used for disease diagnosis, tumor identification	[42]
27	ABC algorithm with K-means clustering	2016	Diagnosis of liver function	Liver images of patient	Liver cancer detection	[44]
28	Feedforward backpropagation neural network (FFBNN) and artificial bee colony (ABC) algorithm	2013	Detection of unwanted growth of tissues	Tissue images of lung	For detection of lung cancer	[45]
29	Artificial bee colony (ABC) algorithm	2013	Detection of symptoms of diabetics	Tissue samples of patients	Diabetic detection	[46]
30	Artificial bee colony (ABC) algorithm	2016	Measurement of glucose content of an aqueous solution	Various fruits and food samples.	Glucose-level detection, determination of energy content	[47]

4 Conclusion

Use of artificial bee colony (ABC) algorithms and its variants covers domains where traditional algorithms for imaging, for diagnosis, and for other purposes fail to perform up to a critical mark. This algorithm overcomes their limitation; however in some cases, this algorithm has limitations in providing expected result alone, and then, it can be combined with other variants of ABC algorithm or can be combined with existing available techniques such GA, PSO, SI, and DE.

It is suggested by authors for image enhancement application use of hybrid version of GABC [global best-guided ABC] algorithm for the integration of global best (Gbest) information into the solution search equation to avoid limitations of traditional method. The basic ABC algorithm suffers with insufficiency regarding its solution search equation, which is good at exploration but poor at exploitation.

References

1. Dorigo M, Colorni A, Maniezzo V (1991) Positive feedback as a search strategy. Dipartimento di Elettronica, Politecnico di Milano. Technical report 91-016, Milan, Italy
2. Kennedy J, Eberhart R (1995) Particle swarm optimization. IEEE international conference on neural networks, vol 4. IEEE, Perth, WA, pp 1942–1948
3. Bonabeau E, Dorigo M, Theraulaz G (1999) Swarm intelligence: from natural to artificial systems. Oxford University Press Inc, New York, NY, USA
4. Millonas MM (1994) Swarms, phase transitions and collective intelligence. In: Langton CG (ed) Artificial life III. Santa Fe Institute. Studies in sciences of complexity, vol XVII. Addison-Wesley, MA, pp 417–445
5. Karaboga D, Akay B (2009) A survey: algorithms simulating bee swarm intelligence. Artif Intell Rev 31:61–85
6. Seeley TD (1995) The wisdom of the hive. Harvard University Press, Cambridge, MA
7. Tereshko V, Loengarov A (2005) Collective decision making in honey-bee foraging dynamics. Comput Inf Syst 9(3):1–7
8. Dervis K, Beyza G, Celal O, Nurhan K (2014) A comprehensive survey: artificial bee colony (ABC) algorithm and applications. Artif Intell Rev 42:21–57
9. Frank YS (2010) Image processing and Pattern recognition, fundamentals and techniques. Wiley, IEEE press, Inc., Hoboken, New Jersey, p 40
10. Sathya DJ, Geetha K (2013) Experimental investigation of classification algorithms for predicting lesion type on breast DCE-MR images. Int J Comput Appl 82(4):0975–8887
11. Sathya DJ, Geetha K (2013) Quantitative comparison of artificial honey bee colony clustering and enhanced SOM based K-means clustering algorithms for extraction of roi from breast dce-mr images. Int J Recent Trends Eng Technol 8(1):51–56
12. Wang J, Zhang D (2013) Image denoising based on artificial bee colony and BP Neural network. TELKOMNIKA 13(2):614–623
13. Zhang Y, Wu L, Wang S (2011) Magnetic resonance brain image classification by an improved artificial bee colony algorithm. Prog Electromagnet Res 116:65–79
14. Jegadeesan S, Suresh BP (2012) Mammogram image analysis for micro calcification detection using bee colony optimization. Int J Comput Sci Technol 3(3):1012–1016
15. Mythili S, Senthil Kumar AV (2015) CTCHABC-hybrid online sequential fuzzy extreme kernel learning method for detection of breast cancer with hierarchical artificial bee, vol 15. IEEE, pp 343–348
16. Nazarian M, Mashala AD, Haronabadi A (2013) Classification of breast cancer samples through using the artificial bee colony algorithm. Int J Comput Appl Technol Res 2(5): 522–525
17. Pourmandia M, Addeh J (2015) Breast cancer diagnosis using fuzzy feature and optimized neural network via the Gbest-guided artificial bee colony algorithm. Comput Res Prog Appl Sci Eng 01(04):152–159
18. Mythili S, Senthil Kumar AV (2016) Identification of breast cancer by artificial bee colony algorithm with least square support vector machine. Int J Comput Sci Inf Secur 14(3): 178–183
19. Sivakumar R, Marcus K (2012) Diagnose breast cancer through mammograms using EABCO algorithm. Int J Eng Technol 4(5):302–307
20. Hsu WY, Hu YP (2015) Artificial bee colony algorithm for single-trial electroencephalogram analysis, clinical EEG and neuroscience. EEG Clin Neurosci Soc (ECNS) 46(2):119–125
21. Satapathy SK, Satchidananda D, Jagadev AK (2016) ABC optimized RBF network for classification of EEG signal for epileptic seizure identification. Egypt Inform J (in press)
22. Shreyasi D, Pratyusha R, Konar A, Nagar AK (2014) Selecting the optimal EEG electrode positions for a cognitive task using an artificial bee colony with adaptive scale factor optimization algorithm. In: IEEE congress on evolutionary computation 2014, Beijing, China, pp 2748–2755 (2014)

23. Agrawal V, Chandra S (2015) Feature Selection using artificial bee colony algorithm for medical image classification. IEEE 978-1-4673-7948-9/15, pp 1–6
24. Kuo CH, Kuo YC, Chou HC, Lin YT (2016) P300-based brain-computer interface with latency estimation using ABC-based Interval type-2 fuzzy logic system. Int J Fuzzy Syst 02:1–13
25. Hancer E, Ozturk C, Karaboga D (2013) Extraction of brain tumors from MR images with artificial bee colony based segmentation methodology, pp 516–520
26. Menon N, Ramakrishnan R (2015) Brain tumor segmentation in MR images using unsupervised artifcial bee colony algorithm and FCM clustering brain tumor segmentation in MR images using unsupervised artificial bee colony algorithm and FCM clustering. In: IEEE international conference on communications and signal processing, Melmaruvathur, pp 0006–0009 (2015)
27. Kumar S, Sharma TK, Pant M, Ray AK (2012) Adaptive artificial bee colony for segmentation of CT lung images. In: Proceedings of international conference on recent advances and future trends in information technology (iRAFIT2012), pp 1–5 (Int J Comput Appl)
28. Guo C, Ding Y, Yuan J, Xu G, Wang X (2016) Adaptive photo acoustic imaging quality optimization with EMD and reconstruction. In: Proceedings of SPIE 10024, optics in health care and biomedical optics, vol VII, p, 100241S
29. Gaddala LK, Vanamala AK, Surekha Y (2016) Clinical support system to predict heart disease using artificial bee colony algorithm of artificial neural network. Int J Innov Eng Technol (IJIET) 7(2):315–321
30. Sowmya N, Vijayabhanu N (2015) Artificial bee colony (ABC) Optimization for the prediction of coronary heart disease. Int J Trend Res Dev 2(5):2394–9333
31. Dilmac S, Korurek M (2013) A new ECG arrhythmia clustering method based on modified artificial bee colony algorithm, comparison with GA and PSO classifiers. pp 1–5 (ISBN: 978-1-4799-0661-1/13/2013 IEEE)
32. Babaoglu I, Kıran MS, Ulker E, Gundu M (2013) Diagnosis of coronary artery disease using artificial bee colony and K-nearest neighbor algorithms. Int J Comput Commun Eng 2(1): 56–59
33. Suresh LP, Panigrahi BK (eds) Proceedings of the international conference on soft computing systems, advances in intelligent systems and computing, vol 397, pp 1003–1011. doi:10.1007/978-81-322-2671-0_94
34. Jain SN, Rai CS (2014) Blind source separation of super and sub-Gaussian signals with ABC algorithm, ACEEE. In J Signal Image Proc 5(1). doi:01.IJSIP.5.1
35. Li B, Li Y, Gong L (2014) Protein secondary structure optimization using an improved artificial bee colony algorithm based on AB off-lattice model. Eng Appl Artif Intell 27:70–79
36. Wang Y, Guo GD, Chen LF (2013) Chaotic artificial bee colony algorithm: a new approach to the problem of minimization of energy of the 3D protein structure. Mol Biol 47(6):894–900
37. Alshamlan HM, Badr GH, Alohali YA (2016) ABC-SVM: artificial bee colony and SVM method for microarray gene selection and multi class cancer classification. Int J Mach Learning Comput 6(3):184–190
38. Cesar M, Benıtez V, Lopes HS (2010) Parallel artificial bee colony algorithm approaches for protein structure prediction using the 3DHP-SC model. In: Intelligent distributed computing, vol IV, SCI 315. Springer, Berlin, pp 255–264 (springerlink.com)
39. Latifoglu F (2013) A novel approach to speckle noise filtering based on Artificial Bee Colony algorithm: an ultrasound image application. In: Computer methods and programs in biomedicine, vol 111, pp 561–569
40. Feng TC, Li THS (2016) Advanced hierarchical fuzzy classification model adopting symbiosis based DNA-ABC optimization algorithm. Appl Math 7:440–455
41. Alshamlan H, Badr G, Alohali Y (2015) mRMR-ABC: a hybrid gene selection algorithm for cancer classification using microarray gene expression profiling. In: BioMed research international, Article ID 604910, vol 2015. Hindawi Publishing Corporation, pp 1–15

42. Garro BA, Rodríguez K, Vázquez RA (2016) Classification of DNA microarrays using artificial neural networks and ABC algorithm. Appl Soft Comput 38:548–560
43. Mostafa A, Fouad A, Abd Elfattah M, Hassanien A, Hefny H, Zhu SY, Schaefer G (2015) CT liver segmentation using artificial bee colony optimisation. Proc Comput Sci 60:1622–1630
44. Lin Z, Peng L, Pei-li Q (2016) Liver function diagnosis based on artificial bee colony and K-means algorithm. Int J u- and e-Service Sci Technol 9(1):123–128
45. Balachandran K, Anitha R (2013) An efficient optimization based lung cancer pre-diagnosis system with aid of feed forward back propagation neural network (FFBNN). J Theor Appl Inf Technol 56(2):263–271, 9p
46. Beloufa F, Chikhi MA (2013) Design of fuzzy classifier for diabetes disease using modified artificial bee colony algorithm. Comput Methods Programs Biomed 112:92–103
47. Saracoglu OG, Bagis A, Konar M, Tabaru TE (2016) ABC Algorithm based fuzzy modeling of optical glucose detection. Adv Electr Comput Eng 16(3):37–42

Certain Applications and Case Studies of Evolutionary Computing Techniques for Image Processing

A. Vasuki

Abstract The chapter gives an introduction to optimization based on evolutionary computational techniques and swarm intelligence. Evolutionary computational algorithms adopt the principles of biological evolution and use a population of solutions that evolves with every generation. The bio-inspired computing algorithms that mimic the behavior of swarms of birds and insects, referred collectively as swarm intelligence, are a subset of evolutionary algorithms. The behavior of swarms individually as well as collective behavior in a flock has been extensively studied and an insight into their integration with the optimization algorithm is given. The evolutionary optimization algorithms such as genetic algorithm, particle swarm optimization, ant colony optimization, bee colony optimization, cuckoo search, fish school search, firefly algorithm have been reviewed. The application of these algorithms to image processing has been outlined, and few case studies have been presented.

Keywords Evolutionary computation · Bio-inspired computing · Image processing · Genetic algorithm · Particle swarm optimization · Ant colony optimization · Bee colony optimization · Cuckoo search · Fish school · Firefly algorithm

1 Introduction

Optimization is finding the best possible solution within a set of constraints. The solutions are obtained by defining objective or fitness functions whose maxima or minima lead to the best solution. When there are multiple maxima or minima, one of them will be the global best, whereas the others will be local best solutions. The algorithm has to converge to the global maxima or minima and not get trapped in local maxima or minima.

A. Vasuki (✉)
Department of Electronics and Communication Engineering,
Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India
e-mail: vasuki.a.ece@kct.ac.in; avasuki@gmail.com

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_12

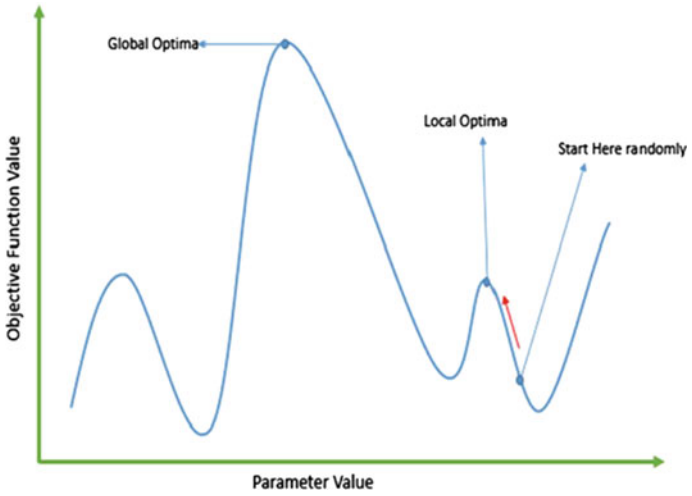


Fig. 1 Example of fitness function (*Source* Ref. [1])

The values taken by an objective function with respect to one of its input parameters is shown in Fig. 1. For example, if $f(x)$ is the fitness or objective function that is one-dimensional, x will be the input parameter. The curve shows distinct peaks and valleys that correspond to the maximum and minimum values of the function. The peaks are the local or global maxima as indicated in the figure, and the valleys are local or global minima. The problem is to start randomly at some point in the curve and move along to the global maximum or minimum, as the case may be.

Optimization can be single objective or multi-objective. The feasible solution might not satisfy all the constraints or might not give the maximum for all the fitness values. Evolutionary computational algorithms are developed from natural biological evolution, and a subset of the algorithms is based on the behavior of swarms of particles such as birds, bees, ants, and fish. Evolutionary computational algorithms introduce a randomness into the search for the optimum solution compared to the classical methods. They are based on searching a space that contains all possible solutions called the search space, and the search is done in parallel, thus making it efficient. Optimization methods assume importance in the context of finding the optimum solution in many engineering and scientific applications where all the parameters are constrained to lie between minimum and maximum limits.

In mathematical terminology, a mathematical function of a set of variables becomes the fitness or objective function. The function is also called as loss or cost function depending on the application. The values taken by the parameters determine the value of the function, either maximum or minimum. The values are chosen from a given set within allowed limits. The search space is a n-dimensional Euclidean space represented by R^n. If the fitness function is convex, there will not be local

minima. When there is more than one fitness function for a problem, it becomes multi-objective. If two functions are conflicting, a trade-off is essential. These algorithms are iterative and give the optimum solution in finite time. The maximum number of iterations reached or the value of the fitness function attained becomes the stopping criteria. Some of the application areas are electrical engineering, antenna design, control systems, operations research, and image processing.

2 Evolutionary Computational Techniques

Evolutionary computational algorithms are inspired by nature and biological evolution and a subset of them by the behavior of birds, animals, etc. A typical biological evolution is shown in Fig. 2.

Evolutionary computational algorithms are a class of optimization techniques that are based on bio-inspired computing [2]. The biological operations of reproduction, crossover, and mutation play a big role in the optimization process of finding the best solution to the given problem. The possible solutions become the population, and the fitness or objective function value indicates the quality of the solution. Depending on the definition of the fitness function, its maximum or minimum value might be the optimum solution. Through the process of natural selection, mutation, crossover, the population undergoes changes in each generation (each iteration creates one generation) and the fitness evaluated. The algorithms are iterative and converge in a finite time. Either the number of iterations or the fitness function value is fixed as the limiting condition for the algorithm to reach termination. The computational complexity must be small enough to make the algorithm practically feasible for real-time applications.

Evolutionary optimization algorithms are population-based, and a class of these algorithms mimic swarm behavior. Typical swarm of birds is shown in Fig. 3. The population represents the candidate solutions, which evolves over generations. The fittest of the population survive in the next generation and are selected for reproduction.

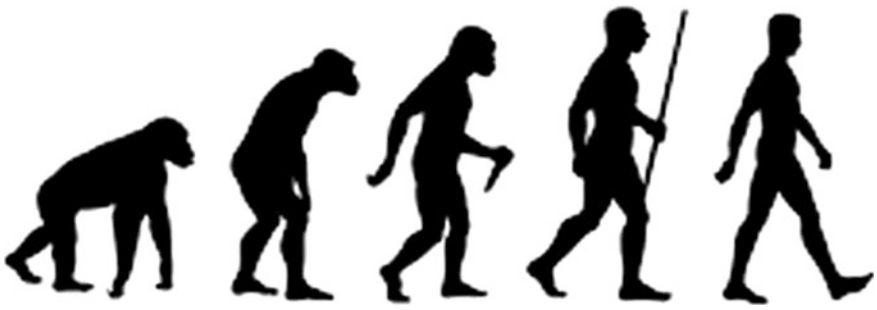


Fig. 2 Biological evolution (*Source* Internet)

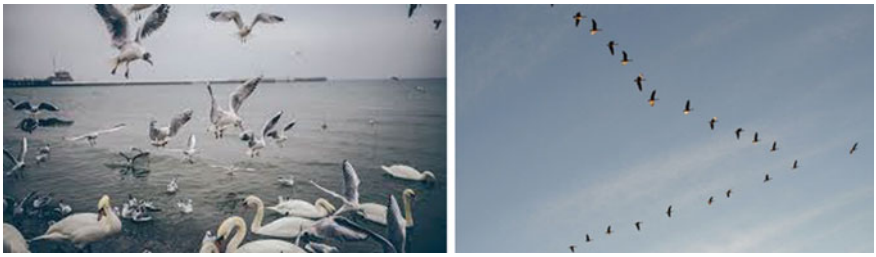


Fig. 3 Swarm of birds (*Source* Internet)

Biological process iteratively improves, striving towards optimization. It always finds the best solution with respect to the objective while maintaining the perfect balance with other components. This idea is adapted for computing and other engineering applications. Bio-inspired algorithm is a metaheuristics method that mimics nature in order to solve optimization problem. Numerous studies had been carried out extensively which have resulted in promising findings. This leads to exploration into new areas of application and computing. Some of the popular bio-inspired optimization methods are discussed in the following subsections.

2.1 Genetic Algorithm (GA)

GA is one of the earliest optimization algorithms developed based on biological evolution [3]. GA is a random search algorithm first developed by John Holland in the 1960s. GA is an evolutionary optimization algorithm based on the Darwin theory of 'Survival of the Fittest.' The symbolic representation of genetic algorithm is shown in Fig. 4.

Fig. 4 Symbolic representation of GA (*Source* Internet)

GA finds the optimal or near optimal solutions for problems based on genetics. It has one or more objective functions that are either minimized or maximized within a set of constraints. GA finds the best solution in the shortest possible time for continuous or discrete unconstrained or constrained optimization problems. There is a search space that contains all possible solutions to the problem. The algorithm finds the best possible solution by searching in parallel. GA is an efficient algorithm that always provides a solution to the problem that evolves over time. More number of parameters can be accommodated, and it is suitable for multi-objective optimization problems. GA also provides solutions in finite time for NP-hard problems.

The algorithm starts with a population of randomly generated individuals that constitutes the search space. The population size has to be chosen based on the problem. The algorithm progresses iteratively, where, at each iteration, next generation of population is created. This is based on the properties of natural selection such as reproduction, mutation, and crossover. A fitness value is calculated for the objective function, and the algorithm terminates when either the optimum fitness value is reached or the number of iterations reaches a pre-defined maximum. The parameters are coded as finite length strings with the coding alphabet being either binary 1s and 0s or some other characters. The coded parameters become chromosomes in the algorithm.

In reproduction, the fitness value of the strings is used in selecting them for reproduction. The individuals with higher fitness values contribute in producing offsprings for the next generation. The crossover operator takes the parents from the mating pool, operates on them, and produces children. If there is no crossover, the existing chromosomes become the next generation, and the crossover rate is zero. If all the offsprings are created by crossover, the crossover rate is 100%. The crossover ratio is a parameter that takes a value between [0, 1] that is an indication of the percentage of offsprings that are created by crossover. Crossover ratio is defined as the ratio of the number of offsprings created by crossover to the total number of offsprings created.

Similarly, mutation operator may be applied on the population, which randomly changes the value of any member of a string. For example, in a string of binary values, changing a 1 to 0 or a 0 to 1 is mutation. Mutation ratio is a measure of the percentage of offsprings that is created by mutation, and the value varies between [0, 1]. Mutation ratio is defined as the ratio of the number of offsprings created by mutation to the total number of offsprings created. If mutation ratio is zero, the offsprings are created either by copying the chromosomes or by crossover. If mutation ratio is 1, then all the offsprings are created by mutation. Crossover and mutation are applied in order to prevent the algorithm from converging to the locally optimum solution. GA might not always converge to the globally optimum solution, and calculation of the fitness value might be computationally expensive. But GA is simple and efficient that makes it suitable for any problem where the solution can be represented as a string of numbers.

2.2 Particle Swarm Optimization (PSO)

PSO was introduced by Eberhart and Kennedy in 1995 [4] for optimization of continuous nonlinear functions. It is a population-based search algorithm that is based on swarm intelligence similar to swarm of birds, as shown in Fig. 5. The concept of PSO is based on the five basic principles of swarm intelligence given by Millonas in his paper [5]:

1. Proximity: swarm population should be able to carry out time and space computations
2. Quality: swarm population should be able to respond to quality factors in the environment
3. Diverse response: swarm population should not commit its activities along narrow channels
4. Stability: swarm population should not change its mode of behavior whenever the environment changes
5. Adaptability: swarm population should be able to change behavior mode whenever the computational price is worth it.

The algorithm generates an initial swarm population of size n, with random positions and velocities. The fitness or objective function has to be defined according to the problem to be optimized. The following three steps [6] are iteratively repeated until the stopping criteria is met or the maximum number of iterations is reached.

- Evaluate the fitness of each particle
- Update individual and global best position and fitness
- Update velocity and position of each particle.

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot [\hat{x}_i(t) - x_i(t)] + c_2 \cdot r_2 \cdot [g(t) - x_i(t)] \qquad (1)$$

Fig. 5 Swarm of birds (*Source* Internet)

where

i	particle number
$x_i(t)$	position of ith particle
$v_i(t)$	velocity of ith particle
$\hat{x}_i(t)$	personal best position
$g(t)$	global best position
r_1 and r_2	random values between 0 and 1
c_1 and c_2	random values between 0 and 2
w	random value between 0 and 1.2.

$w \cdot v_i(t)$ is the inertia component responsible for the particle movement in the same direction as it was moving earlier, and w is the inertia coefficient

$c_1 \cdot r_1 \cdot [\hat{x}_i(t) - x_i(t)]$ represents the cognitive component responsible for memory in the particle, and c_1 is the cognitive coefficient

$c_2 \cdot r_2 \cdot [g(t) - x_i(t)]$ represents the social component responsible for movement of the particle toward the best position, and c_2 is the social coefficient

The equation for updating of particle position is given by:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{2}$$

The velocity of the particle is bounded within minimum and maximum limits, and the search space is also assumed to be bounded. This restricts the movement of the particles in the search space so that it does not become unbounded. Figure 6 shows

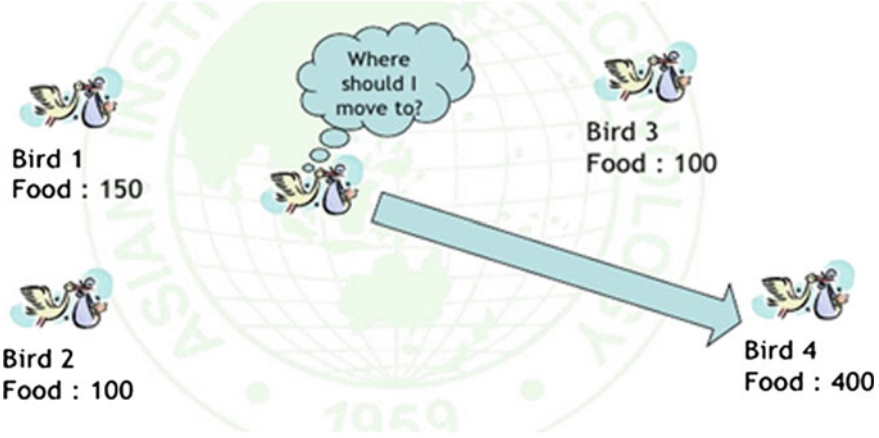


Fig. 6 Behavior of bird swarm (*Source* Internet)

the similarity of PSO algorithm to the behavior of a swarm of birds in foraging for food.

Each particle or bird is a candidate solution. The particles move in the search space looking for the best solution. The particles fly through the search space with a finite velocity and reach a position in each iteration. This position and velocity of the particles and the fitness value are updated at each iteration. The velocity and position of the particles are updated in a continual manner, and the particles move toward the best position. There is a local best and a global best. All the particles move toward the global best. The local best is the best position for each particle, and global best is the best position among all the particles. With each iteration, all the particles move toward the global best position. The iteration number is represented by i, and the particle number is represented by n.

The parameters for velocity and position updation are chosen based on the problem and the size of the search space. The new velocity and position are determined by the previous velocity and position and the movement parameter. If the velocity is too high, convergence will be faster, but it might not be the globally optimum solution. If the velocity is less, convergence will be slow. PSO has been found to yield good results in many industrial and image processing applications that require optimum solutions.

2.3 Ant Colony Optimization (ACO)

Ant colony optimization was introduced by Marco Dorigo in the 1990s. ACO is one of the swarm intelligence algorithms that has been developed, inspired by the collective behavior of ants foraging food [7, 8]. ACO is an optimization algorithm that mimics the behavior of ants finding the shortest path from their nest to food. A typical colony of ants moving together on a path is shown in Fig. 7. This behavior of ants can be applied in our algorithm for solving NP-hard problems in computer science such as finding the shortest path in a graph and traveling salesman

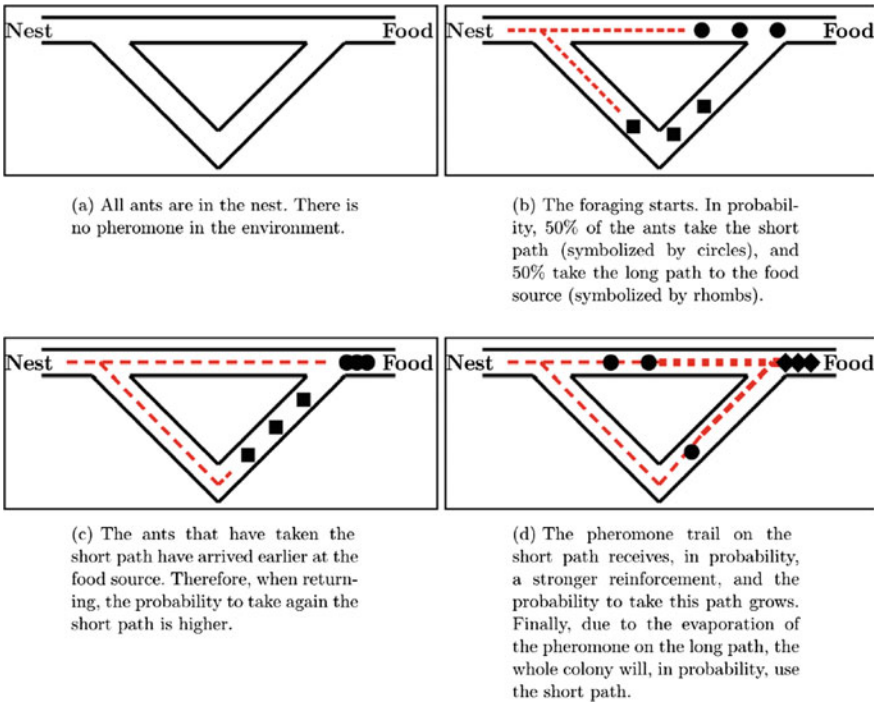
Fig. 7 Ant colony (*Source* Internet)

problem. Ants live in colonies and their behavior is governed by survival of the entire colony rather than individual survival.

Ants forage for food and when they find any, they carry it back to their nest. Ants leave a trail of pheromones while traveling from food to nest. The quantity and quality of the food determines the pheromones laid on the path. The trail is reinforced when the ants make multiple trips. Longer paths have lesser pheromones, whereas shorter ones have more pheromones. When there is pheromone on the trail, more ants follow the path to find food. This behavior is demonstrated in Fig. 8. As time elapses, pheromone gets evaporated. Pheromone evaporation has the advantage of avoiding convergence to a locally optimum solution.

The ant behavior has been modeled (Fig. 8), and the simulation results are given in Fig. 9 [7]. The number of ants in the population is assumed as n. It is assumed that there are two nodes: v_s (nest) and v_d (food source). There are two paths from v_s to v_d of length l_1 and l_2 ($l_2 > l_1$). An artificial pheromone value t_i is assigned to the two paths ($i = 1, 2$).

Starting from v_s, an ant chooses the ith path with probability $p_i = t_i/(t_1 + t_2)$. The ant changes the pheromone value as $t_i = t_i + (Q/l_i)$, where Q is a positive parameter and l_i is length of the path. All the ants are initially placed in v_s. As the ants start moving toward v_d, they find food and deposit pheromones. With time, pheromone



(a) All ants are in the nest. There is no pheromone in the environment.

(b) The foraging starts. In probability, 50% of the ants take the short path (symbolized by circles), and 50% take the long path to the food source (symbolized by rhombs).

(c) The ants that have taken the short path have arrived earlier at the food source. Therefore, when returning, the probability to take again the short path is higher.

(d) The pheromone trail on the short path receives, in probability, a stronger reinforcement, and the probability to take this path grows. Finally, due to the evaporation of the pheromone on the long path, the whole colony will, in probability, use the short path.

Fig. 8 Typical behavior of ant colonies in foraging food (*Source* Ref. [7], pp. 356)

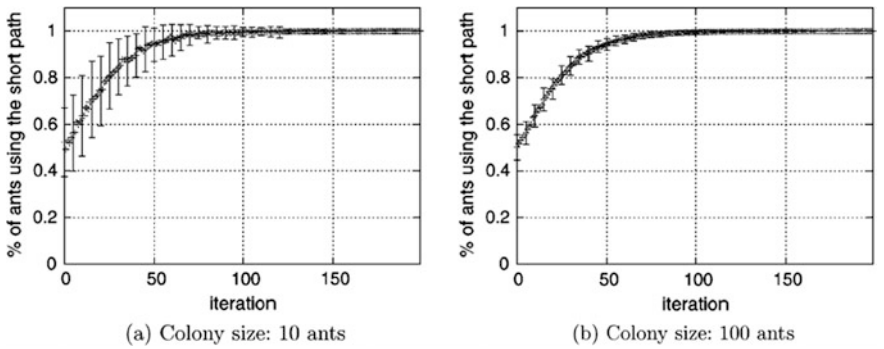


Fig. 9 Simulation results of ant colony behavior (*Source* Ref. [7], pp. 357)

evaporation takes place. Pheromone evaporation is modeled as $t_i = (1 - \rho) \cdot t_i$, $\rho \varepsilon [0, 1]$. The algorithm has been simulated in the reference paper, and the results are presented in Fig. 9. The assumed parameters are $l_1 = 1$, $l_2 = 2$, $Q = 1$, and the pheromone values are 0.5 on each path. The simulation results show that the fluctuations reduce with increase in the size of the ant colony, and finally, they converge to the shortest path. When the above algorithm is applied to practical problems, the differences between real ant behavior and assumed behavior have to be taken care of. The approximations have to be assumed based on the optimization problem.

2.4 Bee Colony Optimization (BCO)

It is a population-based optimization algorithm that was developed in 2001. The algorithm was developed by studying the behavior of honeybees in foraging food [9]. Figure 10 shows a beehive and a honeybee collecting food. A colony of bees

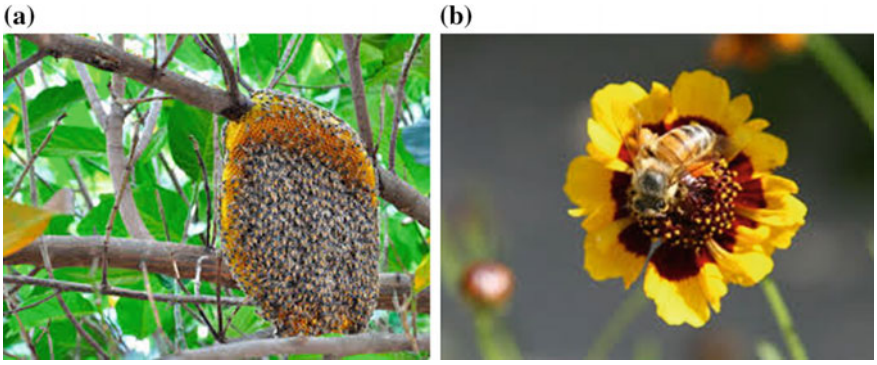


Fig. 10 **a** Beehive **b** Honeybee collecting food (*Source* Internet)

called scouts keeps moving in different directions to look for food (nectar or pollen in flower patch). In the Bee Colony algorithm, the quality of food is equivalent to the value of the fitness function. Higher the value of the fitness function, better is the quality of the food found by the bees. The scouts deposit the food in the hive and perform a waggle dance on the dance floor. This is a communication technique to inform other bees about the discovery and location of the food. The length of the dance is proportional to the quality of the food.

Each candidate solution to the problem is equivalent to a flower patch. The population of bees (agents) searches the solution space. Each time an agent visits a solution, it evaluates its fitness. For each solution, a neighborhood is delimited. Foragers are recruited by scouts to further look for solutions in the neighborhood of promising solutions. If a better solution is found by a forager, it becomes a scout. The area of the flower patch is shrunk so the local exploration is confined to the local best solution. If there is no improvement in fitness value during local exploration for a specified number of cycles, the flower patch is abandoned and new scout is randomly generated. This is continued till the globally optimum solution is found.

Artificial Bee Colony (ABC) optimization algorithm is a variation of the bee colony algorithm. It is based on the foraging behavior of honey bee swarm and was proposed in 2005 by Karaboga. There are three kinds of bees: employed, onlooker, and scout. There are as many number of food sources as the number of employed bees. When the employed bee finds another source of food with more nectar (fitness value), it puts in memory. A dance is conducted indicating this. The onlooker bees find the food based on the dance. The scout bees look for other sources of food.

2.5 Cuckoo Search (CS) Algorithm

Cuckoo search algorithm [10] is a swarm intelligence-based metaheuristic algorithm introduced in 2009 by Young and Deb [11]. Cuckoo is present in most of the countries in the world, and it has more than 100 species. Figure 11 shows two

Fig. 11 Cuckoo bird (*Source* Internet)

Fig. 12 Eggs laid by cuckoo bird (*Source* Internet)

different species of the cuckoo bird. Cuckoo bird undertakes Lévy flights, i.e., series of straight flight paths punctuated by 90° turns to explore the search space.

The parasitic brooding behavior of cuckoos is the inspiration behind this algorithm. The cuckoo bird lays its eggs inside the host nest of some other bird. Figure 12 shows the eggs laid by the cuckoo bird in a host nest. The cuckoo bird can lay eggs with many different colors and also with dots on the colors. Its eggs can match the appearance and color of the host eggs. The brood parasitism can be classified into three types—intra specific, cooperative, and nest takeover. The host bird has two options—it either abandons the nest and builds a new one or throws away the eggs of the cuckoo bird. Certain species of cuckoo removes the eggs of the host in order to hatch its own eggs. Certain other species mimic the color and pattern of the host eggs so that it will not be abandoned. Further, to enhance the parasitic behavior, some cuckoo lay eggs in timing with the host bird and the cuckoo eggs hatch earlier than the other bird.

To make the algorithm simple, three idealized rules are assumed:

(i) Each cuckoo lays one egg at a time and dumps it in a randomly chosen host nest.
(ii) The best nests with high quality eggs will carry over to the next generation.
(iii) The number of available host nests is fixed, and the probability that the host bird will discover the cuckoo eggs is $p \in [0, 1]$. As stated earlier, the host bird either abandons the nest and builds a new one or throws out the eggs of the cuckoo bird. The nest rebuilding ratio p_a is defined as the number of nests that are newly built to the total number of host nests n, and it varies between 0 and 1.

The objective or fitness function can be defined so that maximization leads to high quality solution that is globally optimum. Each egg in the nest is a possible solution, and each egg laid by the cuckoo bird is a new and better solution. The solutions are evaluated by the value of the fitness function, and potentially better solutions are evolved. The host nests form the initial population, and they are chosen randomly. The nests with poor fitness values are abandoned.

When generating new solutions, Lévy flight is performed according to the equation given below:

$$x_i^{t+1} = x_i^t + \alpha \oplus Lévy(\lambda) \tag{3}$$

where x is the solution, t is iteration number, i is the cuckoo, and α is the step size (>0), usually taken to be 1. The first term in the above equation represents the current location, and the second term is the transition to the next position. The second term in the equation is entry-wise multiplication. The Lévy flight is given by:

$$Lévy \sim u = t^{-\lambda} \quad (1 < \lambda \leq 3) \tag{4}$$

where the Lévy distribution gives the step size for the random walk through the search space. It has infinite mean and variance. Lévy walk around the existing solutions will lead to new solutions. However, to ensure that the solution does not get trapped in local optimum, a fraction of the new solutions should be generated by far field randomization. Lévy flight is efficient since large step lengths are possible with the power law step length distribution with a heavy tail [11].

The algorithm for cuckoo search can be summarized as follows:

- The objective function $f(\boldsymbol{x})$ is defined.
- An initial population of n host nests $x_i (i = 1, 2, \ldots n)$ is created.
- While t is less than the maximum number of iterations or stopping criteria not reached, do:

  – Get a cuckoo randomly by Lévy flight and evaluate its fitness F_i
  – Choose a nest j randomly among the n. If $F_i > F_j$, replace j by the new solution
  – A fraction of the worst nests is abandoned
  – Rank the solutions according to quality, retain the best ones, and find the current best

- Process results.

In applying cuckoo search algorithm to practical problems, it has been found that $n = 15$ and $p_a = 0.25$ are sufficient in most cases [11]. Cuckoo search is a population-based algorithm that uses selection similar to that in harmony search. Lévy flight is used in generating the step size for the random walk and it is efficient. The far field randomization generates new solutions that are far away from the current best solution, thus preventing local optima.

The number of parameters to be tuned is lesser than GA or PSO, hence it can be adapted to a wider class of optimization problems. This algorithm can be extended to a wider class of problems, where each nest contains multiple eggs representing multiple solutions. In Ref. [11], the CS algorithm has been compared with GA and PSO, and it has been found that CS is more efficient than the other two and also the success rate is higher. CS is suitable for multimodal function optimization, and several variants have been proposed in the literature since its invention in 2009.

2.6 Fish School Search Algorithm (FSS)

Fish school search (FSS) is a swarm optimization algorithm proposed by Filho et al. in 2008 [12]. FSS is a population-based search algorithm inspired by the behavior of fishes while looking for food. The fishes swim toward the food, eat, and gain weight. The weights of all the fishes cumulatively indicate the success of the fish school in getting food. The fishes move toward better places with each iteration, looking for food. The collective movements and behavior of the fish school is shown in Fig. 13.

The proposed novel algorithm is suitable for searching in high-dimensional, unstructured spaces. The fishes stay together in schools and exhibit collective behavior to increase their survivability and their chances of finding food. The drawbacks are the restricted individual movements and competition for food, if quantity available is less. Three main characteristics adopted in the proposed algorithm are: feeding, swimming, and breeding.

- Feeding refers to the ability of fishes to eat and grow strong. Fishes can also lose weight by swimming. Food is an indication of evaluation of candidate solutions in the search space.
- Swimming is the observable behavior of the fish. It is a collective, contained movement that is driven by feeding in the search space.
- Breeding is the natural selection process that helps stronger individuals to grow and weaker individuals will perish. This is an exploitation technique that will better adapt the candidate solutions in the search space.

The main features of the algorithm are as follows:

- The population consists of swarm of fish, and each fish is a possible candidate solution to the problem.
- The weight of the fish indicates the success of the fish in finding food; this is inherent memory.
- Evolution of the fish is through breeding and collective swimming.

Fig. 13 Fish school–collective movement (*Source* Internet)

The algorithm incorporates the following principles in order to reduce overall computation cost, learn adaptively, share local knowledge, and speed up the search by diversity:

- Simple computations in all individuals (fish)
- Various means of storing information (weight of fish)
- Local computations (swimming is composed of distinct components)
- Low communications between neighboring individuals (fish are to think local but also to be socially aware)
- Minimum centralized control (self-controlling of the fish school radius)
- Distinct diversity mechanisms (to avoid undesirable flocking behavior)
- Scalability (in terms of complexity of the optimization/search tasks)
- Autonomy (ability to self-control functioning).

The algorithm starts by randomly generating an initial population of fish with randomized weights and positions. Food is scattered in the aquarium at various locations in different quantities. The fishes swim towards the food, eat, and gain weight accordingly. The difference in the food concentration at the previous and current position of fish will determine the weight gained (or diminished) by the fish. The equation governing this is given by:

$$W_i^{(t+1)} = W_i^{(t)} + \frac{f(x_i^{(t+1)}) - f(x_i^{(t)})}{\max\{|f(x_i^{(t+1)}) - f(x_i^{(t)})|\}} \tag{5}$$

where $W_i^{(t)}$ is the weight of the ith fish, $x_i^{(t)}$ is the position of ith fish, and $f(x_i^{(t)})$ is the fitness function evaluated (quantity of food) at the ith fish or position. To ensure rapid convergence, additionally it is assumed that the weight of the fish can vary between 1 and W_{scale}, and the initial weight of the fish is $W_{\text{scale}}/2$.

The swimming operator is for all individual and collective feeding, breeding, and survivability. It is assumed that there are three classes of causes of swimming: (i) individual (ii) collective-instinct, and (iii) collective-volition. In *individual* movement, the fish assesses whether the food at another location is better than the present and swims towards it. The direction is chosen randomly. The step size of the fish is defined as step_{ind}. The fish moves according to the step size, and if necessary, the step size is multiplied by a random number between [0, 1]. If the food is out of the boundary of the aquarium or blocked by obstacles, or the quality of the food is not better, the fish does not swim. In *collective-instinct* movement, weighted average of the individual movements of the fish is computed according to the equation given below:

$$\overrightarrow{x}_i^{(t+1)} = \overrightarrow{x}_i^{(t)} + \frac{\sum_{i=1}^{n} \Delta \overrightarrow{x}_{\text{ind}\,i} \{f(x_i^{(t+1)}) - f(x_i^{(t)})\}}{\sum_{i=1}^{n} \{f(x_i^{(t+1)}) - f(x_i^{(t)})\}} \tag{6}$$

where $\Delta\overrightarrow{x}_{\text{ind}\,i}$ is the displacement of the ith fish in the individual cycle.

The *collective–volitive* movement is for overall evaluation of the variation in weight of the entire fish swarm. If the fish swarm is successful in finding food and has added weight, the radius of the school should contract and vice versa. The barycenter of the fish school is calculated as follows:

$$\text{Bari}(t) = \frac{\sum_{i=1}^{n} \overrightarrow{x}_i^{(t)} W_i^{(t)}}{\sum_{i=1}^{n} \overrightarrow{x}_i^{(t)}} \tag{7}$$

where the position and weight of all the fish in the swarm have been taken into account. The collective movement of the fish will be inwards (contraction) or outwards (dilation), taken in small steps with respect to the barycenter of the fish school. The contraction or dilation decision depends on the weight of the fish school, whether it has increased or decreased compared to the previous iteration. A parameter called volitive step, represented by step_{vol}, is defined for this movement. The new position of the fish in the swarm is calculated based on the Eqs. (8 and 9) given below. If the weight increases, Eq. (8) is used, and if the weight decreases, Eq. (9) is applied.

$$\overrightarrow{x}_i^{(t+1)} = \overrightarrow{x}_i^{(t)} - \text{step}_{\text{vol}} \cdot \text{rand}[x_i^{(t)} - \text{Bari}(t)] \tag{8}$$

$$\overrightarrow{x}_i^{(t+1)} = \overrightarrow{x}_i^{(t)} + \text{step}_{\text{vol}} \cdot \text{rand}[x_i^{(t)} - \text{Bari}(t)] \tag{9}$$

The **breeding** operator is used in producing the next generation of the fish school. The fish chosen for breeding must have reached a threshold value that is set according to the success of the fish. The fish that presents the maximum ratio of weight over distance with respect to the breeding candidate is the chosen one for breeding. The initial weight and position of the new child fish is given by:

$$\overrightarrow{W}_k^{(t+1)} = \frac{\overrightarrow{W}_i^{(t)} + \overrightarrow{W}_j^{(t)}}{2} \tag{10}$$

$$\overrightarrow{x}_k^{(t+1)} = \frac{\overrightarrow{x}_i^{(t)} + \overrightarrow{x}_j^{(t)}}{2} \tag{11}$$

Whenever a new child fish is born, the weakest or smallest among the fish in the school is removed, thus keeping the population size constant. The above process is repeated iteratively until one of the defined stopping criteria is met. The stopping criteria can be maximum number of iterations, minimum weight of the school, maximum radius of the school, and maximum fish number or breeding number, as the case may be.

2.7 *Firefly Algorithm (FA)*

The firefly algorithm is another bio-inspired optimization algorithm proposed by Xin-She Yang [13] inspired by the flashing behavior of fireflies. Different specimens of firefly flashing light are shown in Fig. 14.

The firefly is an attractive insect that rhythmically flashes light for short periods of time. The brightness of light and the pattern of flashing vary with the species. The flashes are for attracting other fireflies for mating as well as for attracting prey. According to the laws of physics, light intensity is inversely proportional to the square of the distance from the light source, i.e., $I \propto \frac{1}{r^2}$. Therefore, applying the law, the brightness of light from the firefly decreases exponentially with distance from the firefly. This makes the firefly visible to a limited distance. If an objective function can be defined that corresponds to the brightness of the firefly, an optimization algorithm can be developed based on the behavior of fireflies.

The following three assumptions are made in developing the firefly algorithm:

- All fireflies are unisexual, and they are all attracted to each other
- Attraction is proportional to the brightness of the firefly. The firefly with lesser brightness moves toward the one with greater brightness. If there is no brighter firefly, they move randomly
- The brightness of the firefly is determined by the landscape of the objective function. The value of the fitness function can be made proportional to the brightness of the firefly.

The FA is summarized as below:

- The objective function $f(\boldsymbol{x})$ is defined.
- An initial population of fireflies $x_i (i = 1, 2, \ldots n)$ is generated.
- Computation of $f(x_i)$ gives the light intensity at each firefly.
- Light absorption coefficient γ is defined.
- While $t <$ *maximum number of iterations*, do for all n fireflies:

– If $I_j > I_i$ move the ith firefly toward the jth firefly, as shown in Fig. 15.
– Attractiveness is given by $e^{-\gamma r}$ where r is the distance from the firefly. Evaluate the latest solutions and update the light intensity values.

- Rank the fireflies and choose the current best.
- Process the results.

The attractiveness can be made proportional to the light intensity and the fitness function can be maximized based on the brightness. The light intensity at a distance r from the firefly is given by $I(r) = \frac{I_s}{r^2}$ where I_o is intensity of light source, i.e., light intensity of the firefly. Based on absorption of the medium with absorption coefficient γ, light intensity varies with distance as $I = I_o e^{-\gamma r}$. Combining the effect of the above two mechanism, $I(r) = I_o e^{-\gamma r^2}$. The function can be approximated based on the rate of convergence.

Fig. 14 Fireflies at night (*Source* Internet)

The movement of the *i*th firefly toward the *j*th firefly is governed by the following equation:

$$x_i = x_i + \beta_o e^{-\gamma r_{ij}^2}(x_j - x_i) + \alpha(\text{rand} - \frac{1}{2}) \tag{12}$$

where β_o is the attractiveness of the firefly that is in direct proportion to the light intensity I_o, $r_{i,j}$ is the Cartesian distance between the fireflies i and j, α is a randomization parameter, *rand* is the random number generator between [0, 1]. For practical purposes, β_o can be taken as 1, α can be in the range [0, 1], and the randomization term can be approximated to normal or other distributions. The parameter γ determines the speed of convergence of the algorithm.

The algorithm has been simulated and tested for various test functions and compared with PSO and GA, and it has been found that FA is superior in performance (as per the results given in Ref. [13]).

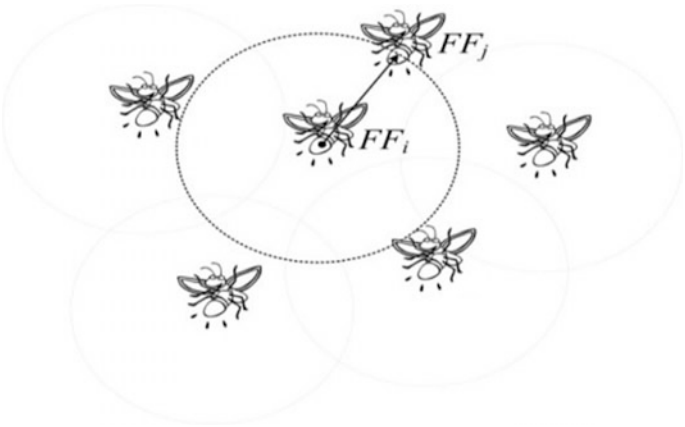


Fig. 15 Behavior of fireflies (*Source* Internet)

3 Applications and Case Studies

The application of evolutionary computational algorithms to image processing finds solutions to real-time problems that are optimal. There is a large set of possible solutions, and the technique is to find the best solution with the given constraints. So it becomes a constrained optimization problem. Image processing is a vast area with extensive applications. The image processing techniques that are used in practical applications are enhancement, filtering, segmentation, object detection and tracking, video processing, object recognition, deblurring, etc. Some recent developments that have taken place in finding an optimized solution to the image processing application and case studies are discussed in this section.

3.1 Swarm Intelligence for Detecting Interesting Events in Crowded Environments

Swarm intelligence has been applied for detecting anomalies in crowded environments [14, 15]. Anomaly detection is important for safety and security in public places, and this proposed automatic system can assist where it is tiresome to manually monitor videos of long duration. Anomaly is assumed as an event with low probability based on past history. Example of an anomaly in appearance will be a truck inside a crowd of humans attending a meeting or a truck sitting on a road divider. Example of anomaly in motion is dispersing of humans at an unusual speed from a crowded place or a bus veering off the road while driving.

Histograms of Oriented Swarms (HOS) along with Histograms of Oriented Gradients (HOGs) are used in anomaly detection in crowded scenes. It captures the scene dynamics based on appearance and motion. Swarms have been used for anomaly detection in videos in the proposed algorithm [14]. Swarms exhibit individual behavior, and they also follow a set of rules within their group. A swarm of agents flying over a crowded scene can capture the dynamics of the scene. The swarm comprises of agents and prey. It has been proposed to construct a prey based on optical flow values over a specific time window with a swarm flying over it to obtain accurate information of motion. The optical flow values of the pixels inside the region of interest (ROI) are the prey. Agents track the prey, filtering the motion, thus avoiding false alarms or noise. The algorithm is applied on region of interest only along with temporal information. ROIs are extracted automatically from the video frames in order to reduce the computational cost. Interest points are identified in the foreground of the image, and ROI is the rectangular region surrounding the interest point. Motion as well as appearance features are used to detect anomalies.

Computational cost is an important factor in this application since it requires fast action after detection. This has to be adapted for different types of situations in real time. The effectiveness of the algorithm has been demonstrated on four benchmark datasets of surveillance that makes it appropriate for such applications.

3.2 Localization of License Plate Number Using Dynamic Image Processing Techniques and Genetic Algorithms

Detection of license plate is a critical part of automatic vehicle identification system. An algorithm based on image processing and GA has been proposed [16] for detection of license plate symbols that differs in design and plate style in different countries. 98.4% overall accuracy has been achieved for two different datasets under different conditions of illumination, orientation, scaling, plate location, etc. Connected Component Analysis Technique (CCAT) is used to detect objects in the image. The image processing phase extracts the objects that may belong to the license plate. GA produces the optimum symbol locations based on Geometric Relationship Matrix (GRM) of the symbols in the license plate.

In the first phase, the captured input color image of the license plate is converted to gray scale. The gray scale image is converted to binary accounting for temporal, spatial variations, and illumination conditions. Dynamic thresholding has been adopted by varying the threshold based on the average intensity value in the neighborhood of each pixel. If the pixel intensity is more than 90% of the local mean, it is background pixel; otherwise, it is foreground pixel. The window size used is 30×30 in selecting the pixel neighborhood. The morphological operations of dilation and erosion have been applied to eliminate noisy objects. An 8-point connected component analysis is done to locate all the N objects in the image and then filtered depending on the size of the object.

The second phase of identifying the symbols will be done by GA. The chromosomes are formed with the genes being the possible symbols in the license plate. The fitness function is defined based on the distance between the chromosomes. Crossover and mutation have been applied in producing offsprings. The algorithm stops if the best chromosome distance is less than 5 or the average objective distance is not improved for six successive generations. The method has been adapted to accommodate differences in license plates in various countries.

3.3 Biologically Inspired Tracker (BIT)

Tracking of objects in video is a difficult task because of changes in shape, size of objects, illumination difference, etc. Tracking by the Human Visual System (HVS) is the best in terms of accuracy, efficiency and robustness. The performance of the HVS has been mimicked in the biologically inspired tracker [17].

BIT has two components—appearance model and tracking model. In the bio-inspired appearance model, classical simple cells that have the characteristics of multi-orientation, multi-scale, and multi-frequency selection have been represented by Gabor filters. The cortical complex cells receive the inputs from the simple cells and perform linear feature integration. The tracking model, that has been developed based on the learning mechanism of neurons, combines generative and

discriminative models corresponding to view-tuned and task-dependent learning. Fast Gabor approximation and Fast Fourier transform have been applied in this work to speed up the process of tracking in real time.

Visual object tracking has applications in surveillance, robot vision, motion analysis, etc. The proposed tracker has been tested on two benchmark datasets and can track at a speed of 45 frames per second. The biologically inspired tracker has been found to be efficient, accurate, and robust.

3.4 Blind Image Blur Estimation via Deep Learning

Deep Neural Network (DNN) and General Regression Neural Network (GRNN) have been used in a deep learning technique [18] for blind image deblurring. Some of the types of blur occurring in images are defocus blur, Gaussian blur, motion blur, etc. If the cause of the blur is known, it is non-blind and if it is not known it is blind. The blur type is classified, then its parameters are estimated. A pre-trained DNN and GRNN have been applied for the first time in the proposed work for the problem of blur analysis.

The input samples are projected into a feature space by DNN, and the blur is classified into one of three classes—Gaussian, motion, and defocus. Motion and defocus blur are preprocessed before the second phase. The neural network for DNN has 1024 nodes in the input layer, 3 nodes in the output layer, and the hidden layer nodes are chosen empirically. GRNN estimates the blur parameters accurately in the second phase. GRNN has an input layer with three nodes, hidden layer, unnormalized output units, a summation unit and normalized outputs. The proposed method has been tested on two datasets with 5000 images from each set. This method outperforms the previous methods even for non-uniform blur.

3.5 A Robust Approach for the Background Subtraction Based on Multilayered Self-organizing Maps

Motion detection in video is a challenging task with applications in video surveillance, robot vision, human–computer interface, etc. A robust approach for the background subtraction based on multilayered self-organizing maps has been proposed [19] that has been achieved using artificial neural networks. It applies segmentation on moving and static elements using a multilayered approach. The algorithm does background subtraction with post processing filtering to identify moving elements using motion segmentation.

The proposed algorithm detects motion in two stages: (i) modeling of background representation and (ii) detection of foreground pixels. The modeling of background mathematically formulates and adapts to changes in the background.

Detection is a thresholding procedure where the current pixel is compared to the closest sample, the difference magnitude is compared to a threshold and classified as foreground or background depending on whether it is greater than or less than or equal to the threshold, respectively. The threshold changes adaptively over time. The algorithm has been tested with the benchmark dataset *Change Detection Challenge* and found to be robust against moving background, illumination changes, jitter, camera motion, etc.

3.6 Multiple-Object Tracking Using Particle Swarm Optimization

The paper [20] proposes particle swarm optimization algorithm with multiple swarms for multiple-object tracking based on histogram matching. Depending on the number of target objects that are tracked, several swarms are created. Initially, histograms of the gray levels are calculated and a feature model for the target objects is established. The difference between the histogram of each particle in the search space and the target object is defined as the fitness value. The target objects are tracked in multiple iterations. The results obtained confirm that the proposed algorithm can rapidly converge, allowing real-time tracking of targets. When the tracked objects go out of range, global search capability of PSO re-traces the target objects, proving it suitable for global optimization.

3.7 A Swarm Intelligence-Based Algorithm for Face Tracking

The paper [21] proposes a face tracking algorithm based on particle swarm optimization. PSO algorithm has been found to be adaptive, efficient and robust for object tracking. Initially, all potential solutions are projected into a high-dimensional space where particles are initialized. Every particle is a potential tracker and the trackers are updated with each iteration. The tracking window is updated based on the position, scale and appearance of the object. The fitness function is the robust face model based on color and corner features.

The particles are driven to search for solutions in the high-dimensional space. The tracking window size is varied based on the object size. When the algorithm converges, the face is tracked. Further, a multi-feature model is proposed for face description to enhance the tracking accuracy and efficiency. The proposed model and algorithm are independent of the object and can be used for tracking any object. Experimental results given in the reference [21] on face tracking demonstrate the superiority of the proposed algorithm in visual object tracking under dynamic environments with real-time performance.

4 Conclusion

Evolutionary computation collectively refers to optimization techniques that are inspired by biological evolution. Some of them are based on swarm intelligence that emulates the behavior of swarms such as birds, fish, and ants. Their individual as well as collective behavior gives an insight into optimization algorithm development to provide the optimum real-time solution for problems in image processing. It has been found that evolutionary computational algorithms that are nature inspired or mimicking the behavior of particle swarms produces results in finite time for practical problems. The applications of bio-inspired computing to image processing with few recently published case studies have been outlined. The experimental results in these applications are given in detail in the published references. These evolutionary algorithms have been found to produce optimal solutions in finite time even for NP-hard problems that are intractable.

References

1. Tutorial Point. Genetic algorithms—introduction. https://www.tutorialspoint.com/genetic_algorithms/genetic_algorithms_introduction.htm
2. Halim AH, Ismail I (2014) Bio-inspired optimization method: a review. NNGT Int J Artif Intell 1:1–6
3. Goldberg DE, Holland JH (1989) Genetic algorithms in search. Optim Mach Learn 3:95–99
4. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, pp 1942–1948
5. Millonas MM (1994) Swarms, phase transitions, and collective intelligence. In: Langton CG (ed) Artificial life III, Addison Wesley, Reading, MA
6. Blondin J (2009) Particle swarm optimization: a tutorial. http://cs.armstrong.edu/saad/csci8100/psotutorial.pdf
7. Blum C (2005) ant colony optimization: introduction and recent trends. Phys Life Rev 2 (4):353–373
8. Dorigo M, Blum C (2005) Ant colony optimization theory: a survey. Theor Comput Sci 344:243–278
9. Lucic P, Teodorovic D (2003) Computing with bees: attacking complex transportation engineering problems. Int J Artif Intell Tools 12:375–394
10. Fister I Jr, Fister D, Fister I (2013) A comprehensive review of cuckoo search: variants and hybrids. Int J Math Model Num Opt 4:387–409
11. Yang X-S, Deb S (2009) Cuckoo search via Lévy flights. In: Proceedings of world congress on nature & biologically inspired computing, IEEE Publications, USA, pp 210–214
12. Filho CJAB, Neto FB, de L, Lins AJCC, Nascimento AIS, Lima MP (2008) A novel search algorithm based on fish school behavior. In: IEEE international conference on systems, man and cybernetics (SMC 2008), pp 2646–2651
13. Yang X-S (2009) Firefly algorithms for multimodal optimization. Chap. 10: stochastic algorithms: foundations and applications, Springer, Berlin, pp 169–178
14. Kaltsa V, Briassouli A, Kompatsiaris I, Hadjileontiadis LJ, Strintzis MG (2015) Swarm intelligence for detecting interesting events in crowded environments. IEEE Trans Image Process 24:2153–2166

15. Kaltsa V, Briassouli A, Kompatsiaris I, Strintzis MG (2014) Swarm based motion features for anomaly detection in crowds. In: Proceedings of IEEE international conference on image process (ICIP), pp 2353–2357
16. Samra GA, Khalefah F (2014) Localization of license plate number using dynamic image processing techniques and genetic algorithms. IEEE Trans Evol Comput 18:244–257
17. Cai B, Xu X, Xing X, Jia K, Miao J, Tao D (2016) BIT: biologically inspired tracker. IEEE Trans Image Process 25:1327–1339
18. Yan R, Shao L (2016) Blind image blur estimation via deep learning. IEEE Trans Image Process 25:1910–1921
19. Gemignani G, Rozza A (2016) A robust approach for the background subtraction based on multi-layered self-organizing maps. IEEE Trans Image Process 25(11):5239–5251
20. Hsu C-C, Dai G-T (2012) Multiple object tracking using particle swarm optimization. In: *WASET–IJCECE*, vol 6, pp 744–747
21. Zheng Y, Meng Y (2009) A swarm-intelligence based algorithm for face tracking. IJISTA 7:266–281

Histopathological Image Analysis for the Grade Identification of Tumor

M. Monica Subashini

Abstract The proposed method is to analyze brain tumor to identify the grade of glioma from magnetic resonant image and histopathological images. The proposed work includes three phases. The first phase preprocesses the pathological images. This involves enhancement and contrast improvement of the images. Secondly, the processed histopathology image is subjected to feature extraction. Gaussian filters techniques and statistical feature extraction techniques are utilized for feature extraction. In the last phase, classifiers are developed to classify the low-grade and high-grade images based on extracted features. K-mean clustering network and SVD classifier are used for classification of low-grade/high-grade gliomas. MATLAB, a familiar tool, efficiently uses algorithms and techniques for identification of low-grade and high-grade glioma tumors.

Keywords Brain tumor · Gliomas · Histopathology · Feature extraction · SVD · K-clustering networks

1 Introduction

WHO (World Health Organization) is the widely used classification and grading of gliomas [1]. Gliomas are defined as brain tumors with histopathological, immunohistochemical, and ultrastructural features of glial differentiation. Grade I and grade II tumors are slow-growing tumors termed as low grade, histopathologically characterized by a low cellularity. Grade I lesions are circumscribed and hence surgically removable in most cases without additional therapy. The grade II gliomas can only be partially treated by surgery and eventually will recur [2]. Gliomas are classified as astrocytomas, oligodendrogliomas, oligoastrocytomas which are derived from astrocytes, oligodendrocytes, and ependymomas. The tumors' optimal management is still debated [3]. Although histological benign,

M. Monica Subashini (✉)
School of Electrical Engineering, VIT University, Vellore, Tamil Nadu, India
e-mail: monicasubashini.m@vit.ac.in

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_13

most of these tumors will transform into malignant tumors within 5–10 years of diagnosis.

Grade III and grade IV gliomas are fast-growing malignant lesions characterized by high cellularity [4]. Anaplastic astrocytomas, oligodendrogliomas, and oligoastrocytomas are the frequent occurring grade III tumors. Glioblastoma is the prototype of the malignant (grade IV) gliomas. Generally, high-grade tumors show a prominent enhancement, while the grade II gliomas lack enhancement. As per the views of few researchers, one-third of malignant gliomas are not showing any contrast enhancement [5]. Pathological diagnosis has to be correlated with clinical and radiological data. More reliable methods are needed to improve the classification and grading of gliomas [6]. The proposed model would aid physicians in diagnosing the tumors with a simple, user-friendly approach.

1.1 Types of Tumor

Tumors are primary and secondary in types. Primary tumors are named based on the cell type or the brain part in which they originate. Mostly, primary brain tumors originate from glial cells. Since it is from glial cells, they are named as glioma. The common types of tumor found in adults are the following:

1.1.1 Astrocytoma

Astrocytes are the star-shaped glial cells in brain, and the tumor arisen is known as astrocytoma. The grade can be of any of the four types. Cerebrum is the most common location for these tumors especially in adults.

Grade I/II astrocytoma is otherwise known as low-grade glioma.
Grade III astrocytoma is also called as high-grade/anaplastic astrocytoma.
Grade IV astrocytoma is known as glioblastoma/malignant astrocytic glioma.

1.1.2 Meningioma

Meninges are the part of brain where this type of tumor arises. It is benign and develops gradually in a slow pace. The type can be of grade I, II, or III.

1.1.3 Oligodendroglioma

Middle-aged adults are more prone to this tumor. A fatty substance which covers the nerves for protection is the originating location for this type of tumor.

1.1.4 Medulloblastoma

Cerebellum is the origin of this tumor. The other name is primitive neuroectodermal tumor. The stage is IV.

1.1.5 Grade I or II Astrocytoma

Children are more prone to this low-grade tumor which occurs anywhere in the brain. The most common type is juvenile pilocytic astrocytoma. Grade I is the type [7].

1.1.6 Ependymoma

The tumor arises from cells that line the ventricles in the spinal cord or the central canal of the cord. Children and young adults are the victims. The types are grade I/II, or III.

1.1.7 Brain Stem Glioma

This tumor originates in the lowest part of the brain. It is a low-grade or high-grade tumor. The most common type is diffuse intrinsic pontine glioma.

1.2 Problem Statement

- The main approach of this project is to identify the brain tumors by feature extraction and classify based on the extracted feature from the histopathological images.
- The first phase is to extract the features from histopathological images and in the last phase, the tumors are classified by use of extracted features with the help of classifiers.
- After extracting the features, SVD and K-mean clustering classifier techniques are applied to classify the image into high-grade or low-grade of gliomas.

1.3 Literature Survey

Many researchers are working toward tumor grade identification from magnetic resonant images which is noninvasive. Gray-level co-occurrence matrix, nearest

neighborhood network, and support vector machines are the common techniques applied for grade identification [8, 9]. Though the grade is predicted, confirmation of disease is from pathological images (invasive). The histopathological images are processed to identify the grade with the help of automated systems to assist the physician when the patients/images are bulk in number. Generally, feature extraction techniques are applied to extract the information and then proceed with classification using neural networks/fuzzy logic.

As discussed, researchers [10] have applied automatic nuclei segmentation for feature extraction and support vector machine, random forest method for classification. Recent works [11, 12] focus on pathological image analysis in which researchers extracted color gray level features and Tamura's features. They have applied nearest neighborhood method, for classifying normal and cancerous images.

1.4 Proposed Work

The proposed work includes three phases. The first phase preprocesses the pathological images. This involves enhancement and contrast improvement of the images. Secondly, the processed histopathology image is subjected to feature extraction. Gaussian filters techniques and statistical feature extraction techniques are utilized for feature extraction. In the last phase, classifiers are developed to classify the low-grade and high-grade images based on extracted features. K-mean clustering network and SVD classifier are implemented for classification of low/high-grade tumors (gliomas).

2 Methodology

The proposed method of glioma grade identification from histopathological images is shown in Fig. 1.

2.1 Database

Histopathological images were downloaded from online sources [13]. Figures 2a, b and 3a, b are the sample images considered for the proposed work of feature extraction and classification. The processed glioma histopathological images [13] are categorized into two distinct classes as high grade and low grade. A total of 80 images were utilized in the research. The database was divided into training set and testing set. Training set [Appendix] included 31 images in each grade (low/high) and nine images for the testing purpose.

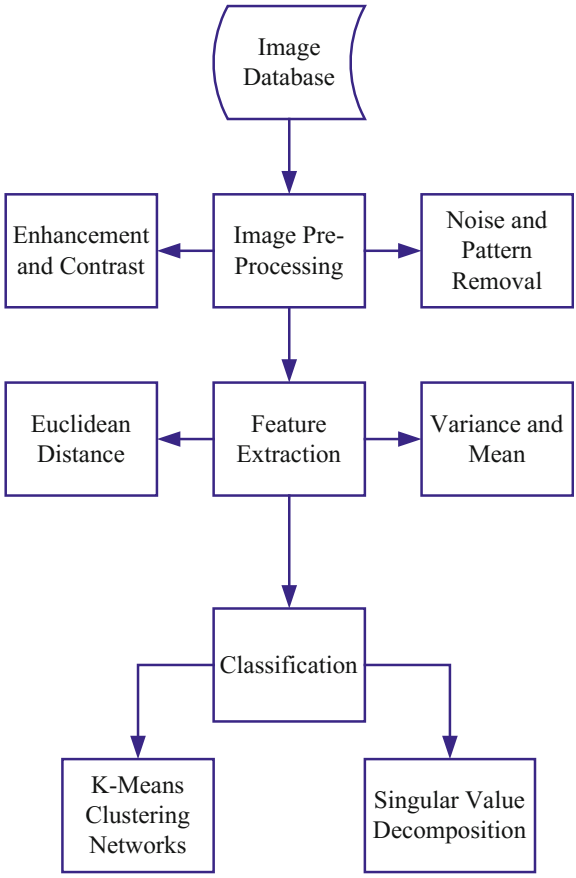


Fig. 1 Typical block diagram of the complete grade identification process

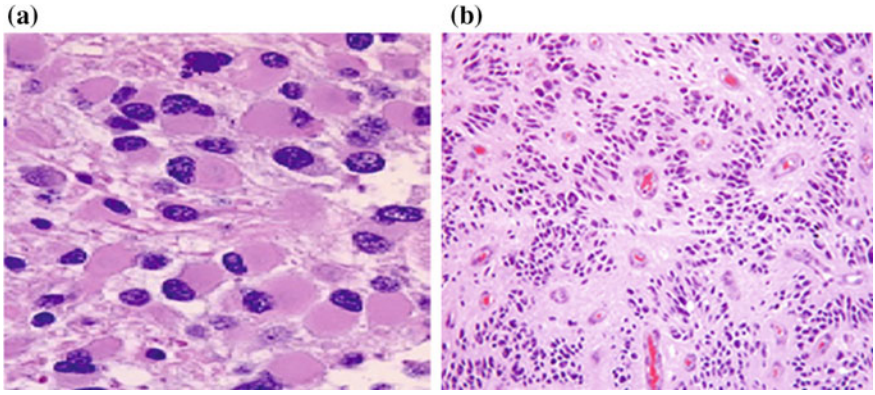


Fig. 2 **a** Pilocytic astrocytoma—grade I. **b** Low-grade astrocytoma—grade II

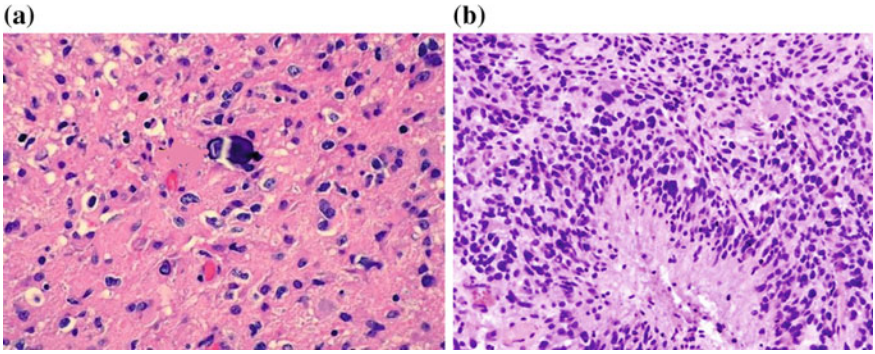


Fig. 3 **a** Anaplastic astrocytoma—grade III **b** Glioblastoma multiform—grade IV

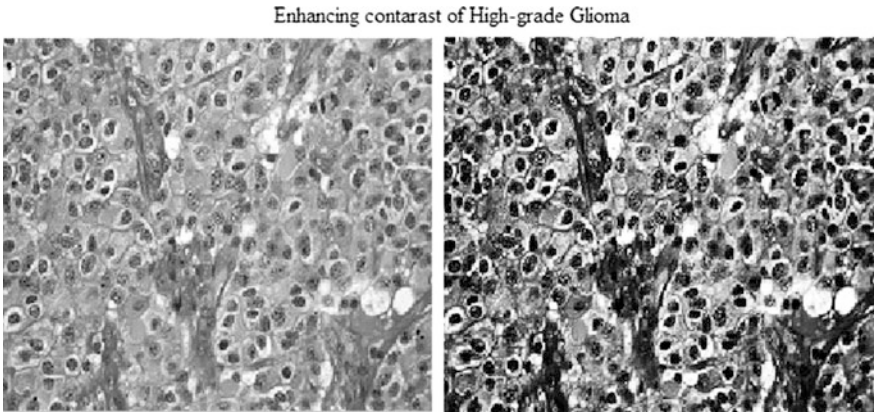


Fig. 4 Typical representation of enhancement of high-grade gliomas

2.2 Preprocessing

Preprocessing of the images includes enhancing the contrast of image, pattern, and noise removal. The contrast of the image was improvised using histogram equalization. The preprocessed images are then subjected to the feature extraction.

2.2.1 Histogram Equalization

In histogram equalization, image contrast is enhanced by transforming the intensity values such that the image histogram is flat. Figures 4 and 5 represent the contrast-enhanced images of high-grade and low-grade gliomas, which are subjected to FFT techniques for noise and pattern removal [14].

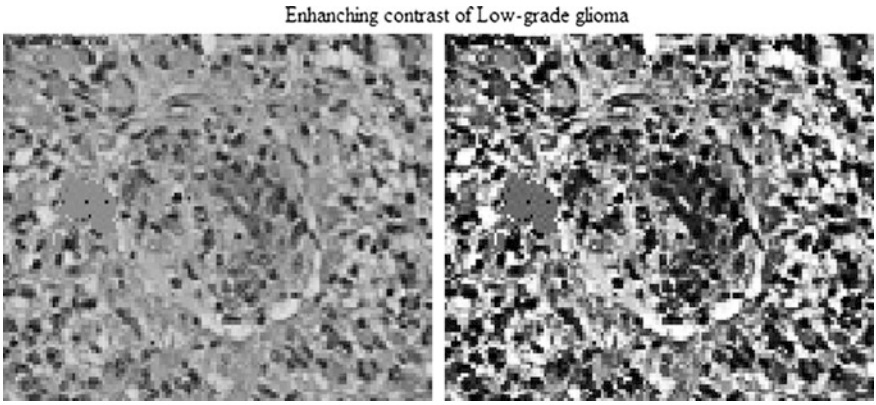


Fig. 5 Typical representation of enhancement of low-grade gliomas

The contrast of the pathological images is increased when the data (image) is depicted by contrast values that are close. The intensities are distributed fairly on histogram by this adjustment [15].

The areas of the lower local contrast also gain a higher contrast level through this method. The task is thus accomplished by effective histogram equalization techniques. These techniques are useful on images with backgrounds, foregrounds that are bright/dark at the same time. In particular, histopathological images have common foreground and backgrounds.

The calculation is not computationally intensive. The implementation of histogram equalization deals with image change or the palette change.

- 'I,' the original image;
- 'M,' the histogram equalization mapping operation;
- 'P' is a palette; and
- P (M (I)), is the operation.

A new palette is defined as $P' = P(M)$ in which image I remain unchanged and histogram equalization is implemented as palette change. If no change in palette P, image I is modified as $I' = M(I)$. This implementation is through image change. Palette change is comparatively better in most cases since it preserves the original data.

2.2.2 Fast Fourier Transform Techniques (FFT)

A fast Fourier transform (FFT) is an efficient algorithm for the computation of discrete Fourier transform (DFT) and its inverse. This technique is used for removing pattern and noise from images. Time or space domain is converted to frequency domain by Fourier transform, whereas a fast Fourier transform quickly computes those transformations by factorizing discrete Fourier transform matrix.

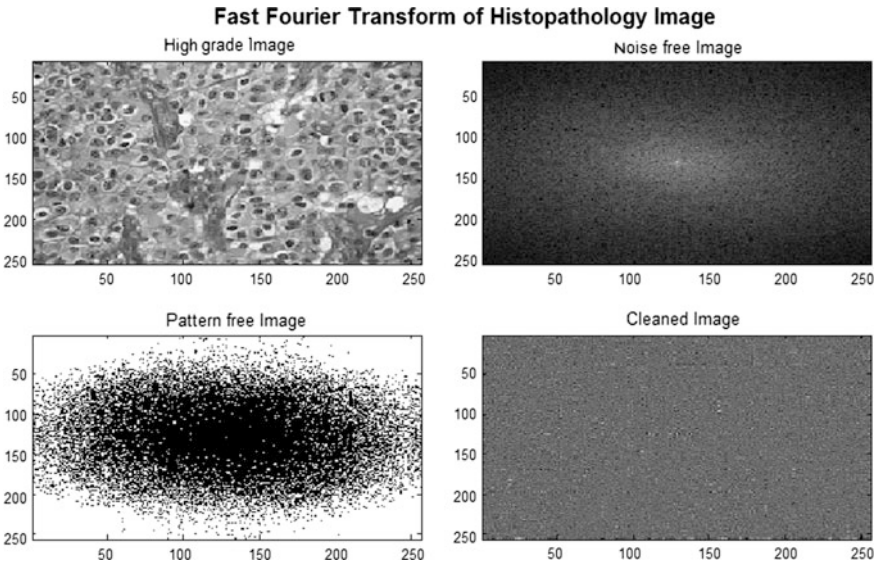


Fig. 6 Typical representation of noise and pattern-free image of high-grade gliomas (FFT)

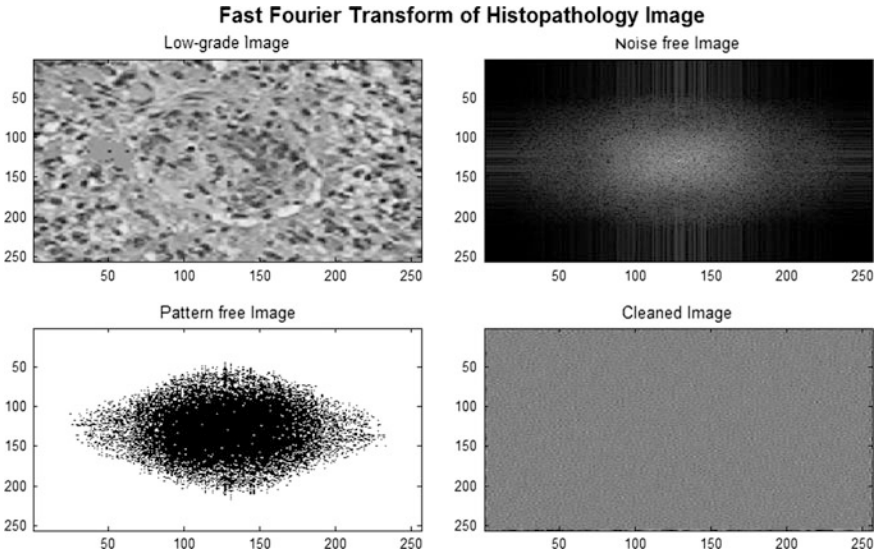


Fig. 7 Typical representation of noise and pattern-free image of low-grade gliomas (FFT)

This leads to the wide usage of the fast Fourier transforms in removing pattern and noise in engineering, science, and mathematics [16].

Figures 6 and 7 show high-grade and low-grade glioma images after removing of noise and pattern. An FFT computes DFT, and it produces similar results when

evaluating the DFT definition directly; FFT is faster and accurate even in presence of round-off error [17].

Let $x_0 \ldots x_{N-1}$ be complex numbers. Discrete Fourier transform can be defined by Eq. (1).

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, \ldots, N-1. \tag{1}$$

There are 'N' outputs X_k. Each output requires a sum of N terms as per Eq. (1) [18].

3 Feature Extraction

The preprocessed image is subjected to feature extraction. Gaussian filter is used for the extraction of statistical features. The features are mean, variance, and standard deviation.

3.1 Gaussian Filter Techniques

Gaussian filter's impulse response is a Gaussian approximation (function). Gaussian filters usually do not overshoot for a step function, and they minimize the fall and rise time. This property is related to the minimum possible group delay occurrence. It is considered as time domain filter which is ideal, just as the same as an ideal frequency domain filter. The extracted features, Euclidean distance, variance, and mean values using Gaussian technique are shown in Table 1.

In most of the discrete cases, the standard deviations of the images are related by Eq. (2)

$$\sigma \cdot \sigma_f = \frac{N}{2\pi} \tag{2}$$

Table 1 Range of features extracted from training and testing images

Grade of tumor	Training set	Test set	Euclidian distance	Mean	Variance
High grade	31	9	14 ± 0.5461	180 ± 22	90 ± 31
Low grade	31	9	4 ± 0.4652	60 ± 26	32 ± 18
Total sample images	80				

N, the total number of samples and standard deviations are expressed by the number of the samples. The design of cutoff frequency of a Gaussian filter depends on standard deviation in frequency domain as shown in Eq. (3), and all the quantities are expressed in their physical units.

$$f_c = \sigma_f = \frac{1}{2\pi\sigma} \tag{3}$$

The cutoff frequency (in physical units) is calculated using Eq. (4) if σ is measured in samples. 'F_s' is the sampling rate.

$$f_c = \frac{F_s}{2\pi\sigma} \tag{4}$$

The response value at this cutoff frequency of the Gaussian filter equals 0.607. The Gaussian filter response is reduced to 0.5 in the power spectrum and 0.707 in the amplitude spectrum. The response of the filter is analyzed based on an arbitrary cutoff value ($1/c$), and the frequency is calculated using Eq. (5).

$$f_c = \sqrt{2\ln(c)} \cdot \sigma_f \tag{5}$$

The Gaussian filter width of size n has a standard deviation of $\sqrt{(n_2 - 1)/12}$ since the moving average corresponds to probability distribution which is uniform. Equation (6) yields a standard deviation with the usage of successive m moving averages of sizes n_1, n_2 … and n_m.

$$\sigma = \sqrt{\frac{n_1^2 + \cdots n_m^2 - m}{12}} \tag{6}$$

The Gaussian filter is noncausal as the filter window is symmetric across the origin when considered in time domain. Hence, the Gaussian filters sometimes become physically unrealizable. This consequence does not affect if the signal frequency is lesser than the filter bandwidth.

4 Classification

K-mean clustering network and singular value decomposition are applied for the classification of the extracted features to identify the grade of tumor (low/high).

4.1 K-Mean Clustering Network

The primary application of K-means network is clustering, or classification. This technique is also applicable for calculating the Euclidean distance which is a most important feature in classification shown in Table 1 and would be used for the purpose of supervised classification. K-means clustering network for classification is a vector quantization method, and it aims in partitioning or clustering 'n' observations into clusters or classes. The calculated nearest mean is the key in segregating the observations to their cluster. Figures 8 and 9 represent the cluster image of high-grade/low-grade gliomas.

The basic steps involved in K-mean clustering network classification:

- Application of K-means clustering for training the data in each class;
- R prototypes per class are used for this training;
- Class label assignment to each of the K × R prototypes; and
- New feature vector 'x' is clustered to class of the closest prototype.

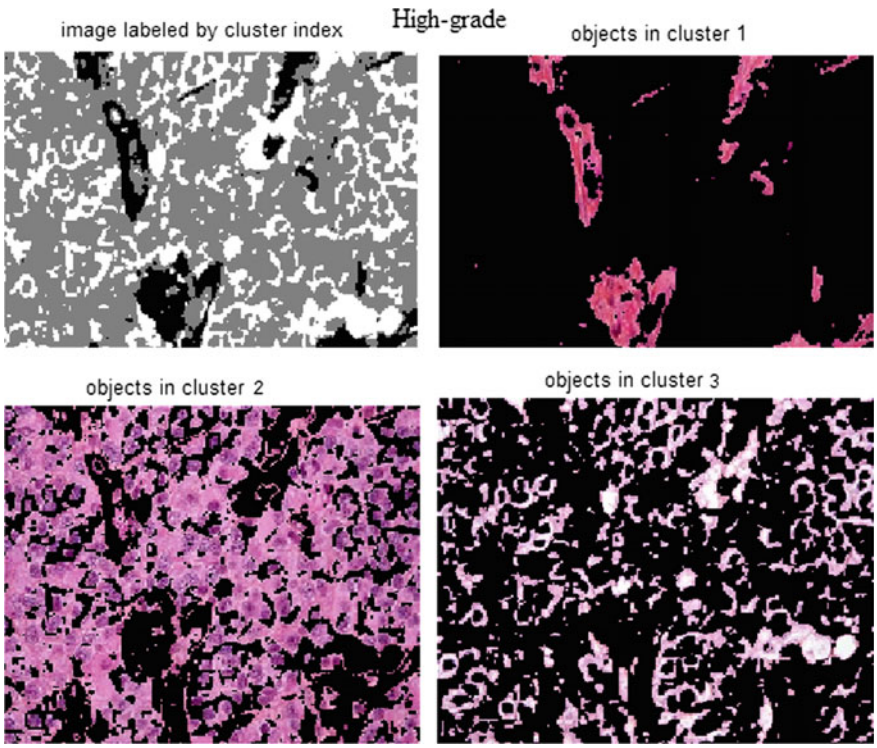


Fig. 8 Typical representation of high-grade image with respect to K-mean cluster, *top left* represents input image, *top right* represents object in cluster 1, *bottom left* represents object in cluster 2, and *bottom right* represents object in cluster 3

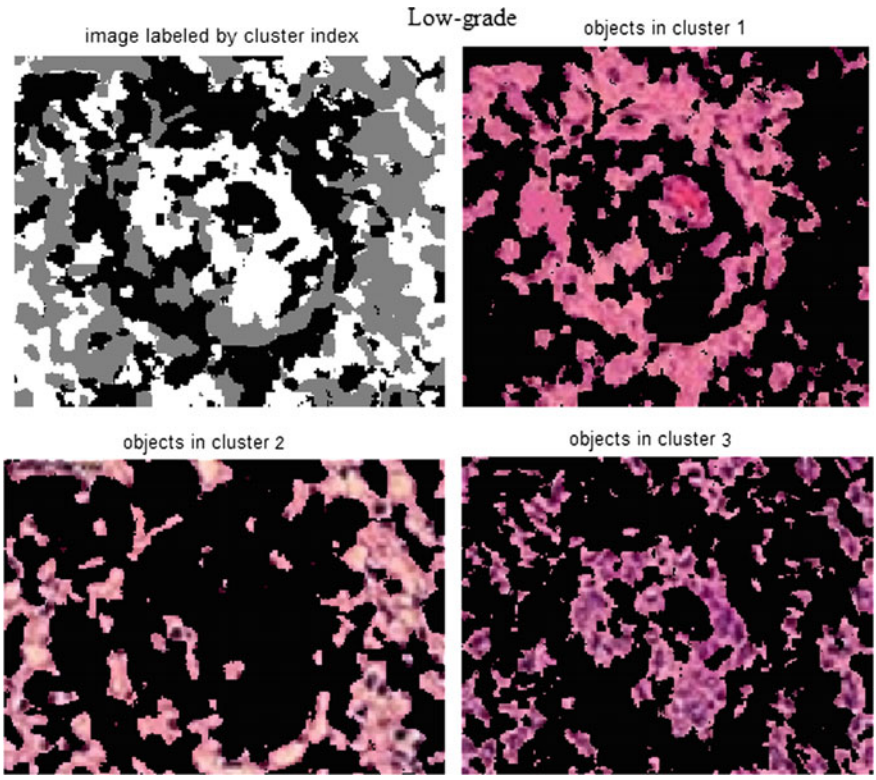


Fig. 9 Typical representation of high-grade image with respect to K-mean cluster, *top left* represents input image, *top right* represents object in cluster 1, *bottom left* represents object in cluster 2, and *bottom right* represents object in cluster 3

The image is reduced to a set of features like Euclidean distance, variance, standard deviation, and mean. These features are classified into distinct classes [19]. Classification may be supervised or unsupervised [20] learning method.

4.2 Singular Value Decomposition

SVD is used for the classification of high-grade and low-grade gliomas on the basis of error value between several singular values with respect to the original image. SVD transforms matrix A into a product USV^T. This allows the refactoring of a digital image into matrices [21]. The refactored three matrices represent the image with small set of values. Those singular values of refactoring preserve the information of the image. Images provide different singular values, and the classification is based on these values. The results are evaluated by compression ratio and quality measurement.

4.3 SVD Approach for Classification

Image compression reduces the amount of data representing an image. Compression is achieved by removing three basic redundancies: (1) coding redundancy, (2) interpixel redundancy, and (3) psychovisual redundancies [22]. The values obtained after compression of images are considered for classifying the tumor image into high/low grade. This process is implemented to extract the most prominent feature for discrimation.

The property of SVD relates to the rank of matrix A to be equal with number of its nonzero singular values. The singular values of a matrix are decreased quickly with the increase of rank. This property reduces noise or compression of matrix data. The smaller singular values or the higher ranks are eliminated to compress the data [23]. The image when SVD transformed carries the first singular value which has good amount of information from the original image [24]. A few singular values are utilized to represent the image with less difference from the original image [23].

The SVD image classification process is performed based on Eqs. (7)–(9)

$$A = \mathrm{USV}^{\mathrm{T}} = \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathrm{T}} \tag{7}$$

The outer product expansion also gives the similar result,

$$A = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_r u_r v_r^{\mathrm{T}} \tag{8}$$

The SVs with smaller values are dropped when compressing the image [25]. The sum after the first k terms is truncated to obtain the closest matrix of rank k.

$$A_k = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_k u_k v_k^{\mathrm{T}} \tag{9}$$

The storage for A_k is found to be $k\,(m + n + 1)$. The integer 'k' is less then n, and A_k, close to the original image. For specific choices of the integer k, the storage for A_k is less than 20%. Figures 10 and 11 typically represent the graph of error between the compressed singular value and the original image of high-grade and low-grade glioma.

4.3.1 Image Compression Measures

The performance of the completed SVD image compression is validated with compression factor which was computed and quality of the compressed image. The compression factor is given in Eq. (10) using compression ratio CR.

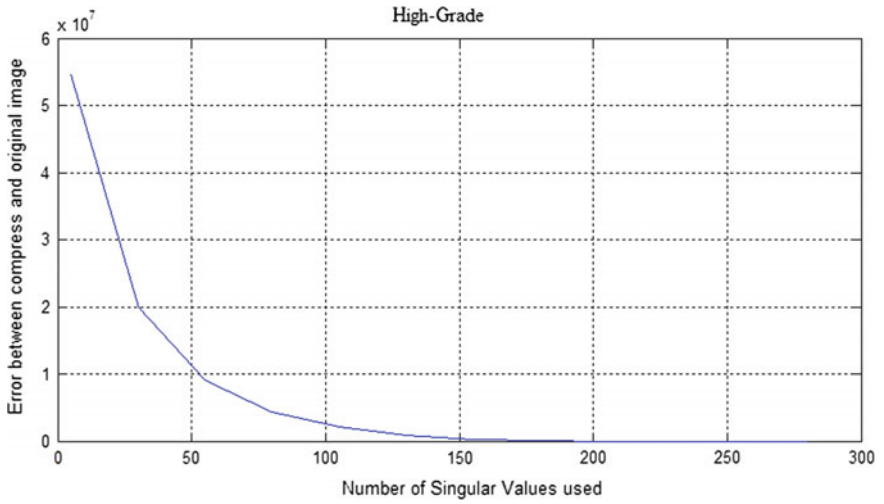


Fig. 10 Typical representation of mean square error and compressed singular value for high grade

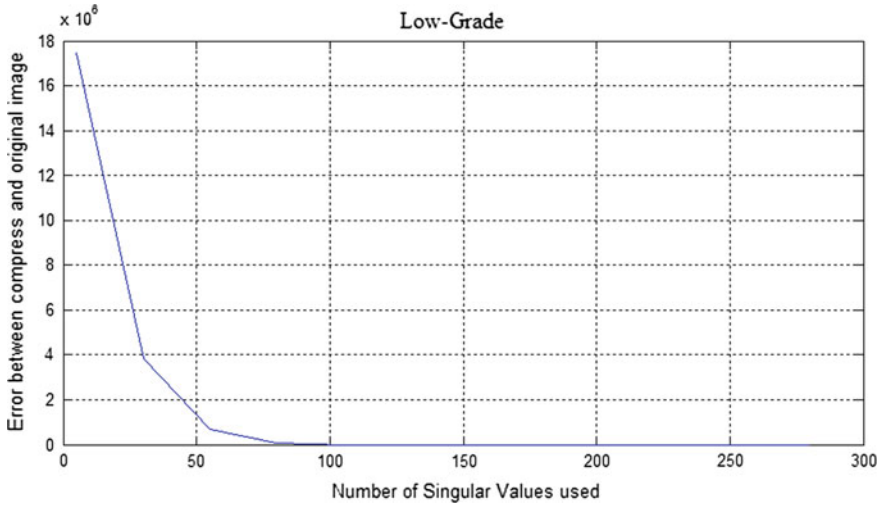


Fig. 11 Typical representation of mean square error and compressed singular value for low grade

$$\mathrm{CR} = m * n/(k(m+n+1)) \tag{10}$$

The mean square error (MSE) [26] as per Eq. (11) is computed to measure the quality of the image after compression.

$$\text{MSE} = \frac{1}{mn}\sum_{y=1}^{m}\sum_{x=1}^{n}\left(f_A(x,y) - f_{A_k}(x,y)\right) \tag{11}$$

The mean square error and computed singular value leads in classification of grades since MSE and SVs are different for high-grade and low-grade images.

5 Discussion and Result Analysis

The K-mean clustering network and SVD classifiers are trained using feature vectors extracted from the training set. The validation set is used to optimize the classifier parameters to obtain the 'optimized classifier.' Optimization was done manually.

A number of combinations of parameter values are tried, and the best performing ones are selected. The optimized K-mean network classifier correctly classified 83.8 and 92.3% of the validation and test sets, respectively (Table 2). Table 2 also shows the kappa statistics for the test and validation sets as 0.782 and 0.897, respectively. The kappa statistic is a chance-corrected measure of the agreement between the classifier predictions and the true classes of samples. The optimized SVD classifier correctly classified 91.9% and 95.4 ± 1.94% of the validation set and test set, respectively (Table 3). Table 3 also shows the kappa statistics for the test and validation sets as 0.891 and 0.938 ± 0.026, respectively. In the case of SVD classifier for high grade, there are 180 singular value used for decomposition, and for the low grade there are 100 singular value used for decomposition which is shown in Figs. 10 and 11.

5.1 Performances of Classifier

The feature extraction and classification times are measured of the developed algorithms. All the tests are run on 2 GHz Core i3 Dell studio with 4 GB of RAM. The feature extraction time is comprised of the time to perform the color standardization (a MATLAB program) and the time to extract the feature for the unknown sample. The feature extraction is performed in part by a MATLAB

Table 2 K-mean clustering network classifier performance on the validation and test sets

	Validation set	Test set
Correctly classified samples	31	36
Incorrectly classified samples	6	3
Total	37	39
Accuracy	83.8%	92.3%
Kappa statistic	0.782	0.897

Table 3 Overall SVD classifier performance on the validation and test sets

	Validation set	Test set
Correctly classified samples	34	37.2 ± 0.756
Incorrectly classified samples	3	1.8 ± 0.756
Total no. of images	37	39
Accuracy	91.9%	95.4 ± 1.94%
Kappa statistic	0.89	10.938 ± 0.026

Test set results are the mean ± standard deviation of the performance over 40 runs

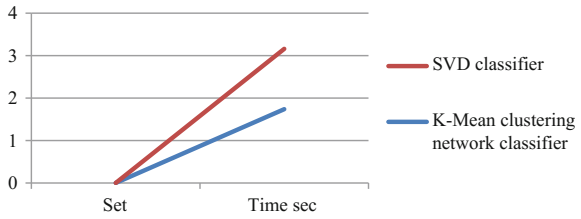


Fig. 12 Typical representation of the performance of SVD and K-mean clustering network classifier

program. The color enhancement took an average of 1.0 s per image. The MATLAB-based feature extraction took, on average, 1.42 s per image. The classification times are the average total time to train the training set and the test set which averaged over 30 runs each. The classification times were comparatively negligible. The SVD took, on average, 0.047 s, and the K-mean clustering network classifier took, on average, 1.72 s to train and classify the entire set of test images. The performance of SVD classifier is good in comparison with the K-mean clustering network classifier in the aspect of classification timing. The accuracy of SVD classifier is also high when compared with the K-mean clustering network in classifying low grade to high grade of gliomas. Figure 12 represents the performance of SVD and K-mean clustering network classifier.

6 Conclusion

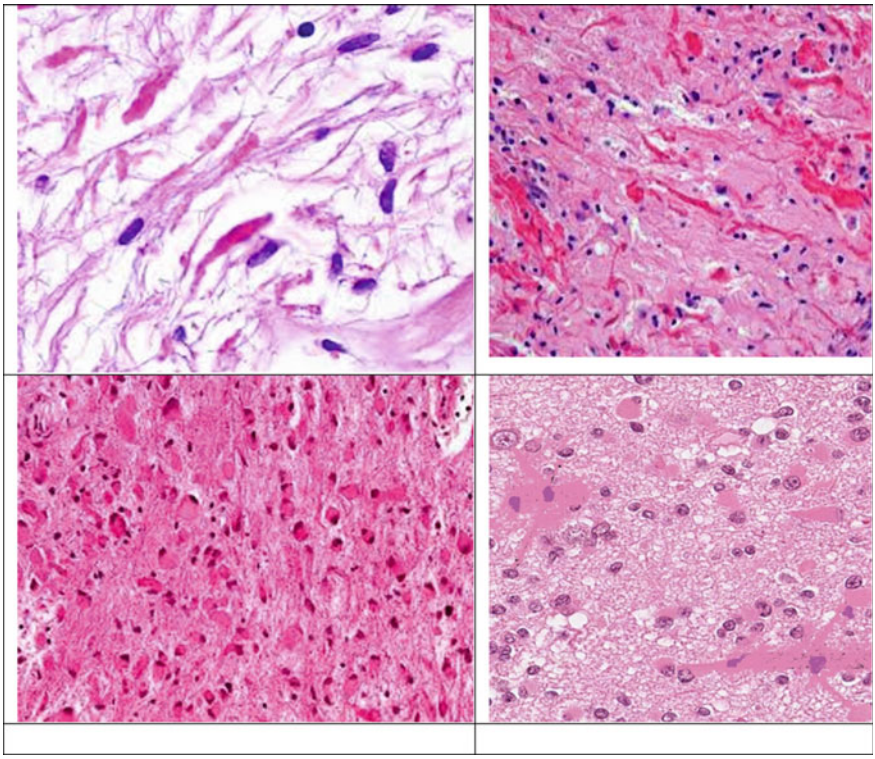
The proposed method provides a diagnostic tool for glioma grade classification. The pattern- and noise-removed images are feature extracted. The set of features representing images are classified using SVD and K-means clustering network. Based on the obtained results, SVD is the most suitable method since this approach is simple and robust to implement. The image compression and classification is done efficiently by SVD. The SVD outperforms K-mean clustering networking in providing good compression ratio. The statistical variations of the image are well adapted by SVD. The methodology followed can be made more accurate by

increasing the number of training set images in the database. The physicians can diagnose brain tumor grade when images are huge in number for analysis. The proposed system includes benefits for modern diagnostic equipment which focus toward speed diagnosis, most accurate, good sensitivity, specificity, less delay, low cost, and user-friendly for doctors.

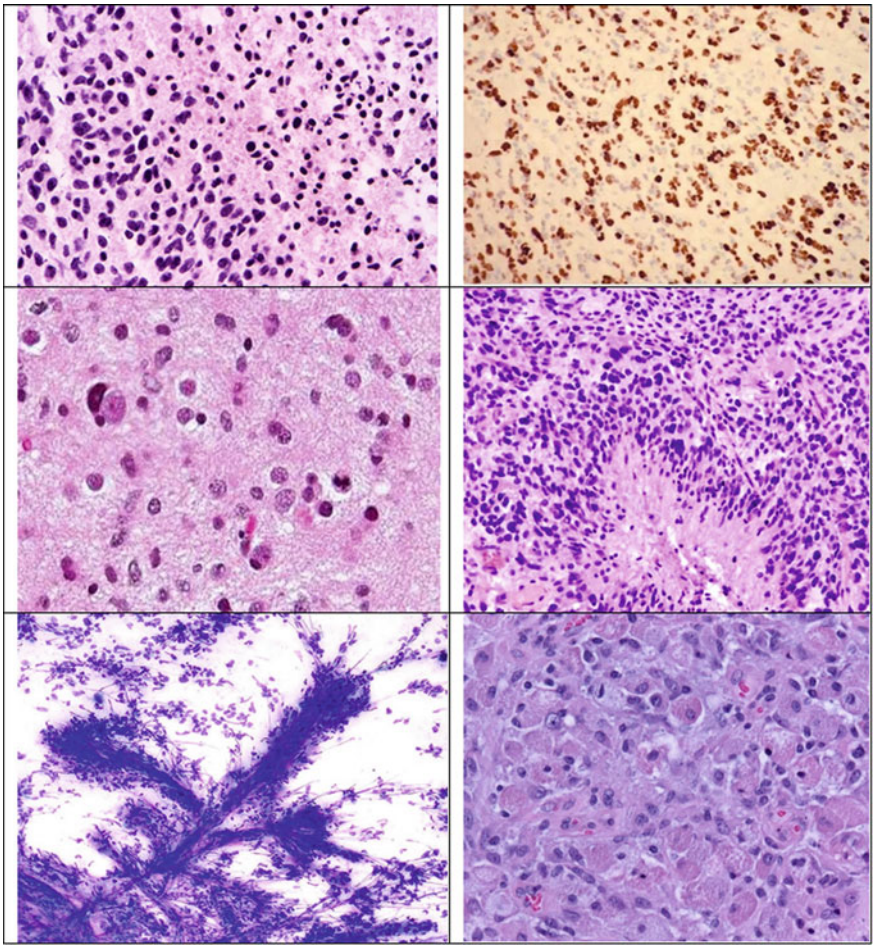
Acknowledgements We thank our Management and School of Electrical Engineering, VIT University for the resources provided to complete this work.

Appendix

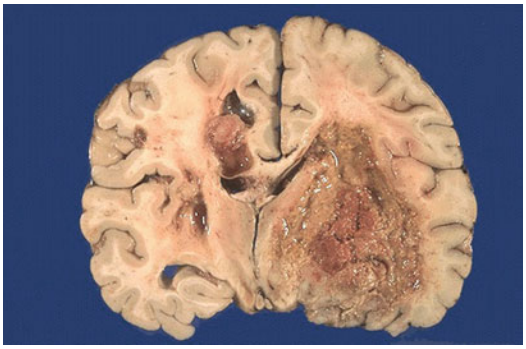
Dataset: Low grade astrocytoma pathological images—training set. *Source* http://library.med.utah.edu



High grade astrocytoma pathological images training set



High grade astrocytoma brain image. *Source* http://library.med.utah.edu



References

1. World Health Organization. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs297. 15 Nov 2010
2. Roniotis A, Marias K, Sakkalis V, Stamatakos G, Zervakis M (2010) Comparing finite element and finite difference techniques as applied to the development of diffusive models of glioblastoma multiforme growth. Presented at the IEEE engineering in medicine and biology society conference, Buenos Aires, Argentina
3. Roniotis A, Marias K, Sakkalis V, Tsibidis G, Zervakis M (2009) A complete mathematical study of a 3D model of heterogeneous and anisotropic glioma evolution. In: Proceedings of IEEE engineering in medicine and biology society conference, vol 1, pp 2807–2810
4. Buckner JC, Aldape KD, Ballman K, Scheithauer BW, Burger PC et al (2004) Immunohistochemical detection of EGFRvIII and prognostic significance in patients with malignant glioma enrolled in NCCTG clinical trials. J Clin Oncol 22, 14(Suppl): 1508
5. Sakkalis V, Roniotis A, Farmaki C, Karatzanis I, Marias K (2010) Evaluation framework for the multilevel macroscopic models of solid tumor growth in the glioma case. In: Proceedings of IEEE engineering in medicine and biology society conference, vol 1, pp 6809–6812
6. Behin A, Hoang-Xuan K, Carpentier AF, Delattre JY (2003) Primary brain tumours in adults. Lancet 361:323–331
7. Cuadra M, Pollo C, Bardera A, Cuisenaire O, Villemure J, Thiran J-P (2004) Atlas-based segmentation of pathological mr brain images using a model of lesion growth. IEEE Trans Med Imaging 23(10):1301–1314
8. Chavan NV, Jadhav BD, Patil PM (2015) Detection and classification of brain tumors. Int J Comput Appl 112(8)
9. Ajaj K, Syed NA, Image processing techniques for automatic detection of tumor in human brain using SVM. Int J Adv Res Comput Commun Eng 4(4)
10. Fukuma K, Surya Prasath VB, Kawanaka H, Aronow BJ, Takase H (2016) A study on feature extraction and disease stage classification for Glioma pathology images. In: IEEE international conference on fuzzy systems (FUZZ-IEEE)
11. Kumar R, Srivastava R, Srivastava S (2015) Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. J Med Eng 2015
12. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B (2009) Histopathological image analysis: a review. IEEE Rev Biomed Eng 2:147–171
13. Image Source: http://library.med.utah.edu
14. Wiskott L, Fellous J, Kruger N, von der Malsburg C (1997) Face recognition by elastic bunch graph matching. IEEE Trans Pattern Anal Mach Intell 19(7):775–779
15. Popescu BP, Vehel JL (2002) Stochastic fractal models for image processing. IEEE Signal Proc Mag 19(5):48–62
16. Martinez M, Kak AC (2001) Pca versus lda. IEEE Trans Pattern Anal Mach Intell 23(2):228–233
17. Ahmed S, Iftekharuddin K, Vossough A (2011) Efficacy of texture, shape, and intensity feature fusion for posterior-fossa tumor segmentation in MRI. IEEE Trans Inf Technol Biomed 15(2):206–213
18. Jahne B (2002) Digital image procession. Springer, Berlin
19. Wang T, Cheng I, Basu A (2009) Fluid vector flow and applications in brain tumor segmentation. IEEE Trans Biomed Eng 56(3):781–789
20. Wickramaratna J, Holden S, Buxton B (2001) Performance degradation in boosting. In: Kittler J, Roli F (eds) Proceedings of second international workshop *on* multiple classifier systems, Springer, Berlin, Germany, pp 11–21
21. Lee CH, Schmidt M, Murtha A, Bistritz A, Sander J, Greiner R (2005) Segmenting brain tumor with conditional random fields and support vector machines. In: Proceedings of international conference on computer vision, pp 469–478

22. Nyul LG, Udupa JK, Zhang X (2000) New variants of a method of MRI scale standardization. IEEE Trans Med Imaging 19(2):143–150
23. Haralick R, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst Man Cybern SMC-3(6):610–621
24. Quddus A, Basir O (2012) Semantic image retrieval in magnetic resonance brain volumes. IEEE Trans Inf Technol Biomed 16(3):348–355
25. Tosun AB, Gunduz-Demir C (2011) Graph run-length matrices for histopathological image segmentation. IEEE Trans Med Imaging 30(3):721–732
26. Mrak M, Grgic S, Grgic M (2003) Picture quality measures in image compression systems. In: *IEEE EUROCON*, Ljubljana, Eslovenia, Sept 2003

Super-Resolution via Particle Swarm Optimization Variants

Maria Aparecida de Jesus, Vania V. Estrela, Osamu Saotome and Dalmo Stutz

Abstract Super-resolution (SR) reconstructs a high-resolution (HR) image from a set of low-resolution (LR) pictures and restores an HR video from a group of neighboring LR frames. Optimization tries to overcome the image acquisition limitations, the ill-posed nature of the SR problem, to facilitate content visualization and scene recognition. Particle swarm optimization (PSO) is a superb optimization algorithm used for all sorts of problems despite its tendency to be stuck in local minima. To handle ill-posedness, different PSO variants (hybrid versions) have been proposed trying to explore factors such as the initialization of the swarm, insertion of a constriction coefficient, mutation operators, and the use of an inertia weight. Hybridization involves combining two (or more) techniques wisely such that the resultant algorithm contains the good characteristics of both (or all) the methods. Interesting hybridization techniques include many local and global search approaches. Results for the SR reconstruction of still and video images are presented for the PSO and the HPSO algorithms.

Keywords Super-resolution · Image registration · Fusion · Image restoration · Mosaicking · Motion estimation · Particle swarm optimization · High-resolution imaging · High-resolution video

M.A. de Jesus · V.V. Estrela (✉)
Telecommunications Department, Universidade Federal Fluminense (UFF), Rio de Janeiro, Brazil
e-mail: vania.estrela.phd@ieee.org

M.A. de Jesus
e-mail: majesus1977br@gmail.com

O. Saotome
Instituto Tecnologico de Aeronautica (ITA), CTA-ITA-IEEA, Sao Jose dos Campos, São Paulo 12228-900, Brazil
e-mail: osaotome@gmail.com

D. Stutz
Instituto Politecnico do Rio de Janeiro (IPRJ), UERJ, Nova Friburgo, Brazil
e-mail: stutz@iprj.uerj.br

J. Hemanth and V.E. Balas (eds.), *Biologically Rationalized Computing Techniques For Image Processing Applications*, Lecture Notes in Computational Vision and Biomechanics 25, DOI 10.1007/978-3-319-61316-1_14

1 Introduction

Although monitoring cameras are omnipresent, people's concern for details still calls for better pictures, due to limited equipment costs and constraints, weather conditions, as well as target shooting distances.

Super-resolution (SR) refers to methods to upscale, upsize, and restore pictures or video sequences. Starting from low-resolution (LR) images, SR recovers a high-resolution (HR) one that offers enriched visual results while eradicating additive noise, handling the imaging detector sizes and optical sensor constraints. It is related to image fusion, registration, and mosaicking [1–3], and [4]. Some SR applications are described below.

(i) Freeze frame and region of interest (ROI) zoom for perception and analysis.
(ii) Resolution-enhanced automatic target recognition.
(iii) In reconnaissance, several images from the same region can help to render a better resolution image.
(iv) To enrich the resolution and create multimodal versions of pathological areas in medical imaging (ultrasound, CT, MRI, etc.) by merging data from several limited resolution images.
(v) Fluorescence microscopy.
(vi) Video standard conversion, e.g., from PAL-M to HDTV signal and from 4K to 8K.
(vii) Forensics.

SR approaches try to model the quality loss when using LR cameras and then to solve an associated ill-posed inverse problem, which does not possess a forthright solution. SR imaging typically involves regularization, optimization, and extensive computations. SR works successfully when the LR images involve somewhat different views of the same object, that is, all the object knowledge exceeds the knowledge from a single frame. Motion estimation (ME) can help to upscale an image and infer the correlation between frames or patches. If an object is steady and appears identical in all frames, no additional knowledge is available. If there is movement or fast transformations, then the target will appear distinctly in different frames. This redundancy in LR frames assists the HR frame reconstruction.

Many factors limit the resolution of imaging systems due to the diffraction limit or resolution constraints associated with the geometry of the optical elements (c.f. Fig. 1). Among the geometrical restrictions resulting from image acquisition CCD cameras, one can cite the size and the shape, and the pitch of pixels can lead to errors. Video super-resolution (VSR) seeks to reconstruct an HR video sequence from an LR one [5, 6]. There are two types of VSR algorithms:

(i) Multiframe complementary information (MCI) techniques use redundancy from different frames; and
(ii) ME methods using data on moving objects such as motion vectors (MVs) or displacement vectors (DVs). VSR performance hinges on the accuracy of the ME framework.

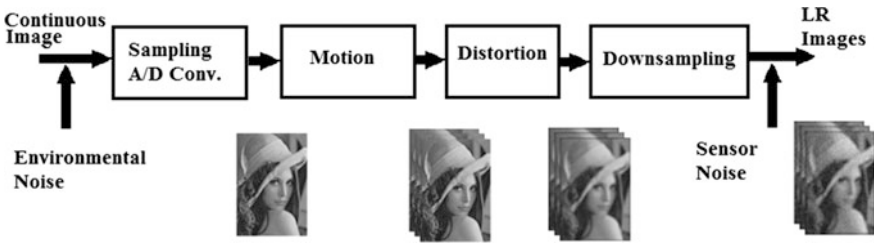


Fig. 1 Imaging degradation process

Sparse reconstruction is an emblematic ill-posed inverse problem where the measurement error and the sparsity terms are used as two conflicting terms to be handled simultaneously inside an objective function. This framework brings in new challenges to the optimization task. Classical optimization and analysis techniques may perform SR disappointingly because it is an expensive optimization problem. This motivates the application of computational intelligence methods [4, 7, 8].

The VSR seeks an HR video that meets certain specifications. This problem can be recast as an optimization task, which pursues the optimal result consistent with quality requirements. For a given imaging degeneration model, optimization techniques can be used to discover the global minimum (or maximum) using a proper fitness function or cost functional.

This chapter examines particle swarm optimization (PSO) applied to SR imaging and VSR schemes relying on an image degeneration model explained in Sect. 2 [7, 9]. Experimental results using evaluation metrics show that PSO methods can improve both objective and subjective results.

2 The Image Degradation Model

Due to hardware limitations, the imaging system has imperfections and various types of degradations as shown in Fig. 1. The Point Spread Function (PSF) models some kinds of optical and mechanical distortions. As the image pixel results from integration over the sensor area, the restricted sensor density causes aliasing effects, limiting the image spatial resolution. These degradations are handled entirely or partly with different SR techniques (Fig. 2). This work uses the subsequent notation:

(i) Upper case bold letters $\boldsymbol{X}$ and $\boldsymbol{Y}$ symbolize lexicographically ordered vectors for HR and LR frames;
(ii) Lower case bold letters such as $\boldsymbol{x}$ and $\boldsymbol{y}$ stand for lexicographically ordered vectors for HR and LR image *patches* in that order;

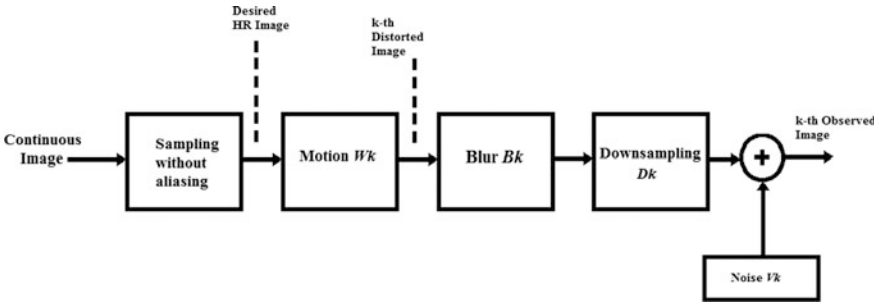


Fig. 2 The observation model relating HR to LR images

(iii) Underlined upper case bold letters show the result of a vector concatenation, e.g.,$\underline{\boldsymbol{Y}}$ is a vector concatenation of $\boldsymbol{Y_k}$, with $k = 1, \ldots, K$ where K stands for the number of captured LR frames by the camera; and
(iv) Plain upper case symbols denote matrices, and simple lower case symbols refer to scalars.

If $\boldsymbol{X}$ denotes the desired HR image (i.e., the digitally sampled image), and $\boldsymbol{Y_k}$ is the kth LR camera observation, then there are K LR frame versions of $\boldsymbol{X}$ where each LR observation vector $\boldsymbol{Y_k}$ is related to the HR image $\boldsymbol{X}$ by

$$\boldsymbol{Y_k} = D_k H_k F_k \boldsymbol{X} + \boldsymbol{V_k}, \tag{1}$$

where the knowledge on motion for the kth frame is encoded in F_k, H_k reproduces the blurring effects, D_k is the operator in charge of downsampling, and $\boldsymbol{V_k}$ represents the noise. If $\boldsymbol{Y}$ is the observed image, then rearranging these linear equations into a large linear system yields

$$\underline{\boldsymbol{Y}} = M\boldsymbol{X} + \underline{\mathrm{V}}. \tag{2}$$

Matrices D_k, H_k, F_k, and M are extremely sparse, unknown, and have to be estimated from the existing LR observations, which worsen the linear system ill-conditioned nature. Thus, regularization is always advantageous and frequently essential.

3 PSO Super-resolution

SR reconstruction has been a very active research area, and many techniques have been suggested to handle this matter. Figure 3 depicts the SR rationale.

Before stating some PSO variants, a genetic algorithm (GA) will be stated, because some of its reasoning appears in hybrid PSO methods.

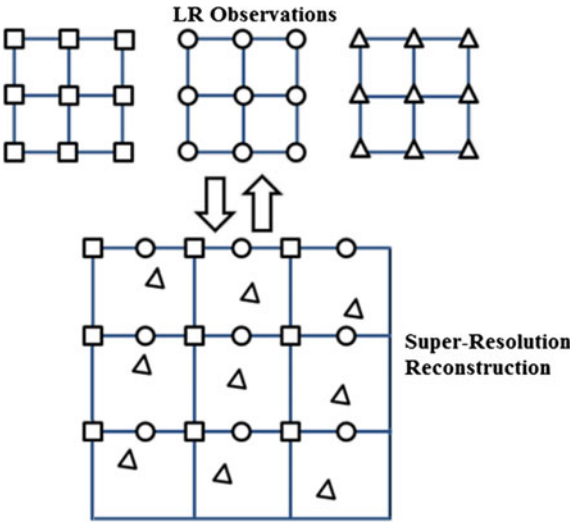


Fig. 3 SR rationale

3.1 Genetic Algorithm (GA)

A GA [10, 11] is a soft optimization strategy to find solutions to optimization problems. They are global search heuristics methods motivated by evolutionary procedures such as selection, mutation, inheritance, and crossover (aka recombination).

GAs consider a population consisting of abstract representations (or chromosomes) of candidate solutions (or individuals) evolves to improved solutions. Habitually, solutions consist of binary strings having 0s and 1s, but other forms of encodings can be used. The evolution procedure begins with a random initial population of individuals and occurs in generations. The fitness of every single individual in the population is computed for each generation. Individuals are stochastically taken from the existing population using their fitness values as the selection criterion. Then, recombination is done along with perhaps randomly mutated individuals to produce a new population for the subsequent algorithm iteration. The algorithm ends once the maximum number of generations or an acceptable population fitness level is reached. If the procedure finishes because of reaching the maximum number of generations, then a proper result may be found.

In SR imaging, GAs can deal with local resolution complications [12, 13], even if it entails huge computational time and it lacks fine-tuning possibilities.

3.2 Classical Particle Swarm Optimization (CPSO)

CPSO is a simple population-based optimization scheme, which involves minimal computational effort. It employs a search motivated by a model of social influence and learning. Individuals emulate the success of their neighbors.

Consider a present diffuse population of size S known as a swarm. Each swarm member is dubbed a particle, and it as a point belonging to the search space. A particle group has a tendency to cluster at an optimized position (maximum or minimum). Therefore, to accomplish CPSO, each particle corrects itself by comparing its previous similarity measure (SM) to its neighbor SMs to reach the best result [14].

If f: $\mathbb{R}^n \rightarrow \mathbb{R}$ is the cost functional to be minimized, then the real number output corresponds to the result of the optimized fitness function, for the given candidate solution [15].

Given an unknown gradient ∇f of f, an optimum solution $\boldsymbol{x}^*$ for which $f(\boldsymbol{x}^*) \leq f(\boldsymbol{y})$, for all $\boldsymbol{y}$ in the search space, is sought. Alternatively, the maximization of a function $h = -f$ can be performed.

At iteration k, $\boldsymbol{x}_i \in \mathbb{R}^n$ is the position vector of the ith particle in the search space whose velocity is $\boldsymbol{v}_i \in \mathbb{R}^n$, $\boldsymbol{p}_{besti}$ is the best $\boldsymbol{x}_i$ for particle i, $\boldsymbol{g}_{best}$ is the global best known position among all particles of the entire swarm, and w means the weight. c_1 and c_2 are acceleration constants whose pdfs are uniformly distributed in the interval sandwiched between 0 and 1.

In the CPSO-based SR method, $\boldsymbol{x}_i$ is the transform parameter vector that has to be estimated, $f\left(\boldsymbol{x}_i^k\right)$ is the functional or cost function to be optimized, the $\boldsymbol{p}_{besti}$ is the maximum cost functional $f\left(\boldsymbol{x}_i^k\right)$ for each particle, and $\boldsymbol{g}_{best}$ corresponds to the best cluster.

CPSO-based SR Algorithm

(1) Initialize the swarm randomly with initial values for variables $\boldsymbol{x}_i^0$, $\boldsymbol{v}_i^0$, $\boldsymbol{p}_{besti}$, $\boldsymbol{g}_{best}$
(2) For each particle i, repeat until the whole population has been analyzed:
(3) Set initial values for variables $\boldsymbol{x}_i^0, \boldsymbol{v}_i^0, \boldsymbol{p}_{besti}, \boldsymbol{g}_{best}$
(4) $\boldsymbol{v}_i^{k+1} = w\boldsymbol{v}_i^k + c_1 \,\text{rand}\,[\boldsymbol{p}_{besti} - \boldsymbol{x}_i^k] + c_2 \,\text{rand}\,[\boldsymbol{g}_{best} - \boldsymbol{x}_i^k]$..
(5) $w^{(k+1)} = w^k + dw$.
(6) $dw = \frac{(w_{\min} - w_{\max})}{T}$.
(7) $x_i^{(\boldsymbol{k}+1)} = \boldsymbol{x}_i^k + \boldsymbol{v}_i^{k+1}$.
(8) If $f\left(\boldsymbol{x}_i^k\right) > f(\boldsymbol{p}_{besti})$ then $\boldsymbol{p}_{besti} = \boldsymbol{x}_i^k$.
(9) If $f\left(\boldsymbol{x}_i^k\right) > f(\boldsymbol{g}_{best})$ then $\boldsymbol{g}_{best} = \boldsymbol{x}_i^k$.
(10) If it converges, then stop.
(11) If the maximum number of iterations is not reached, then go to step 2.

The stopping criterion can be maximum number of iterations, and/or a solution with adequate *MI* value. The parameters w, c_1, and c_2 are chosen by the developer and adjust the performance and effectiveness of the CPSO method.

4 SR Challenges

SR uses subpixel accuracy MC to match areas in adjacent frames to merge them and to combine details wisely, which sometimes cannot be done successfully due to lack of novelty in the LR frames [16–18].

Generic image priors may help regularize the solution properly but are not sufficient. Regularization becomes especially crucial when there is an insufficient number of measurements, and/or only one LR frame is observed [5]. Recently, example-based (EB) methods helped to regularize the SR reconstruction problem and to break the SR limit caused by inadequate measurements [19]. EB methods find the prior knowledge by sampling other images locally and are effective when observations are insufficient. Some issues must be addressed: (i) the choice of the optimal patch size given the target image; (ii) different databases with different statistics are necessary; and (iii) how to use the EB prior more efficiently.

Projection onto convex sets (POCSs) express the SR problem as a manifold delimited by multiple constraining convex sets whose intersection contains the sought after image [20]. The concept of a convex set brings in flexibility and adds in different types of constraints or priors, even nonlinear and nonparametric restrictions. It can handle motion blur, but it may require post-processing. The POCS can be further extended for robust, object-based SR reconstruction. The advantage of the POCS technique easily incorporates any kinds of constraints and priors difficult for stochastic approaches. POCS downsides are the substantial computation, slow convergence, and the solution nonuniqueness that depends on the initial guess, the necessity of priors on the motion parameters and system blurs, and inability to estimate parameters at the same time.

The appropriate model formulation is the basis to obtain a suitable solution. Computational intelligence [21] is a substitute to traditional optimization that offers high precision, and lower time cost in demanding applications. Experimental outcomes using the evaluation metrics defined in Sect. 5 [22] confirm improvements in both objective and subjective results. Other ways to resolve the SR problem comprise machine learning (ML) [23] and compressive sensing (CS) [21, 23]. Next, some drawbacks are listed and briefly discussed.

Absence of Motion or Change adds little innovation, and the overall image quality will resemble ordinary spatial upsizing.

Abrupt Motion Detection is problematic to track and results in low-quality frames attributable to: (i) motion blur produced by the camera, and (ii) the use of compression relying on delta-frames. Abrupt motion generates many alterations among adjacent frames, thus increasing the effects of the codec quantization on the required bit rate with a corresponding loss of details.

Heavy Compression (aka low bit-rate compression) in several circumstances can be intolerable for SR. There are two ways of implementing lossy video codecs: (i) via delta-frames, and (ii) with the help of key frames only [24]. If the video has suffered high compression by a key-frame codec, then this means that each frame was compressed independently, and many particulars are missing. This customarily leads to blocking artifacts, which are usual in heavily compressed JPEG images for the reason that the object changes a great deal during motion. Hence, to accurately replicate motion and acquire minutiae becomes unmanageable. Other frames may cause blocking artifacts along with poor picture quality. Similarly to MC, SR uses differences between frames, which results in gross and useless changes that mean no improvement for the HR frame (loss of details).

Image Registration is impacted the accomplishment of high-quality SR reconstruction where the LR frames can be seen as complementary spatial samplings of the HR image to be fused. When LR observations have heavy aliasing artifacts, the resulting HR frame will show low quality. The performance of standard image registration methods decreases proportionally to the decline in the resolution of the observations, resulting in more registration errors. LR picture registration and HR image estimation are fundamentally interdependent processes. Accurate subpixel ME improves HR image estimation. Conversely, high-quality HR image can simplify ME. Therefore, LR image registration affects the HR image reconstruction. Still, with restricted observations, the joint estimation of registration parameters and HR image may introduce overfitting. Stochastic approaches dealing with the HR image estimation and image registration simultaneously are auspicious; still, the motion models can restrict performance. Optical flow ME works when used in more intricate scenarios. Nonetheless, the scarce measurements for local MEs make SR algorithms vulnerable to errors. The 3-D SR problem brings in several extra challenges.

Computational Loads are severe due to the large number of unknowns involved in costly matrix manipulations, which obstruct real-time implementations. The number of calculations goes up dramatically for nontranslation models, which can be bettered by massive parallel computing. The FPGA technology can ease real-time SR systems. For videos with random movements, favorable results stem from parallel computing, for example GPUs and other unusual hardware deployments.

Robustness is essential since the parameters describing the image degradation cannot be found perfectly, and sensitivity to outliers may originate visually disturbing pieces. These imprecisions cannot be regarded as being Gaussian noise, which is the customary supposition when employing the l2-norm.

Performance Limits clarify SR camera development, assisting the investigation of factors such as model errors, the frame quantity, and zooming factors, but an ambitious examination of the performance bounds for all SR techniques can be difficult. Initially, SR reconstruction is considered a hard task demanding many interdependent constituents. Second, the most informative prior for the SR task is unknown. Last, good metrics are still necessary. It is appropriate to extend models with a comprehensive analysis of SR performance by adding factors such as ME,

decimation factor, the number of frames, and prior knowledge. Although it is cumbersome to get a consistent performance evaluation for different SR techniques, some benchmarks and realistic datasets can simplify algorithm assessment and understanding.

5 Image Quality Assessment (IQA) in SR

Image Quality Assessment (IQA) procedures can be classified as subjective where human observers perform the image quality assessment and objective, which uses objective metrics. Since ultimately all the images have to be appraised by the human visual system (HVS), subjective evaluation is the only true metric. In reality, however, subjective evaluation is not only problematic and costly, but it cannot be transformed into real-time computer programs with output feedback. Consequently, it is more feasible to use objective IQA metrics to analyze the image quality. According to the original image accessibility (ground truth), traditional objective IQA approaches can be [25, 26, 27]:

(i) Full-reference (FR) metrics rely on the original and distorted images;
(ii) Reduced-reference (RR) metrics require part of the original image and the distorted image; and
(iii) No-reference (NR) metrics need only the distorted image.

The most common IQA metrics are described below.

Mean Square Error (MSE): It provides the squared error between the original and the SR image. It will be closer to zero when the ground truth and registered image are equal. It increases when there is a rise in dissimilarity. The MSE can be calculated as follows:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} [\boldsymbol{I}(i,j) - \boldsymbol{I}'(i,j]^2}{MN}, \tag{3}$$

where $\boldsymbol{I}(i, j)$ and $\boldsymbol{I}'(i, j)$ are the original and the SR images, M and N are correspondingly the numbers of rows and columns of each image.

Peak Signal-to-Noise Ratio (PSNR): The *PSNR* is expressed in decibels (dBs) as

$$\mathrm{PSNR} = 20 \log_{10} \frac{\mathrm{Max}_I}{\sqrt{\mathrm{MSE}}}, \tag{4}$$

with Max_I corresponding to the maximum possible pixel value of the image. Using a byte per pixel leads to $\mathrm{Max}_I = 255$. PSNR is high when the ground truth and the SR images are similar.

Correlation Coefficient (CC): It represents how correlated (in the least squares sense) the two data sets are. It is defined as:

$$\mathrm{CC} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[\boldsymbol{I}(i,j)-\mu_I][\boldsymbol{I}'(i,j)-\mu_{I'}]}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}[\boldsymbol{I}(i,j)-\mu_I]^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}[\boldsymbol{I}'(i,j)-\mu_{I'}]^2}} \tag{5}$$

where μ_I and $\mu_{I'}$ are the mean of images of $\boldsymbol{I}(i, j)$ and $\boldsymbol{I}'(i, j)$, respectively. The maximum absolute value of *CC* is 1 and signifies perfect correlation.

Structural Similarity Index (SSIM): It measures the similarities between images, and it emphasizes any image alteration as a mixture of correlation loss, the amount of luminance misrepresentation, and contrast interferences. It is given by

$$\mathrm{SSIM} = \frac{(2\mu_{\boldsymbol{I}}\mu_{\boldsymbol{I}'}+c_1)(2\sigma_{\boldsymbol{II}'}+c_2)}{\left(\mu_{\boldsymbol{I}'}^2+\mu_{\boldsymbol{I}}^2+c_1\right)\left(\sigma_{\boldsymbol{I}'}^2+\sigma_{\boldsymbol{I}}^2+c_2\right)}, \tag{6}$$

where $\sigma_{\boldsymbol{I}}$ and $\sigma_{\boldsymbol{I}'}$ are the variances of $\boldsymbol{I}(i, j)$ and $\boldsymbol{I}'(i, j)$, and $\sigma_{\boldsymbol{II}'}$ the covariance between $\boldsymbol{I}(i, j)$ and $\boldsymbol{I}'(i, j)$. c_1 and c_2 are constants that prevent the denominator to be zero. SSIM $\in$ [0, 1], and it approaches 1 as the images become more similar.

The *MSE* along with the *PSNR* are widely used *FR* metrics because they are easy to implement and have clear physical meanings. Still, they fail to be consistent with the perceived visual quality in many cases. The *SSIM* index takes into account the fact that the human visual system (HVS) perceives the local image structures when computing image quality. Most FR metrics detect and match the features of the original and distorted images to estimate the visual quality of the SR image. These IQA methods pose challenges for SR applications since image sizes change during the SR process.

Moreover, the original HR images are missing whenever the image SR methods are needed in real applications, which may compromise the FR metrics. So, the objective IQA metrics for SR images are desperately necessary.

6 Other PSO Variants

6.1 Some Types of Hybrid PSO (HPSO) Algorithms

PSO is an effective optimization technique that has been applied to an ample range of optimization problems. Nevertheless, its performance can be improved employing certain variants called hybrid PSOs (HPSOs) discussed in this section. The PSO changes can be done in one of the stages below-mentioned, or they can involve a combination of these strategies [28].

(i) **Initialization**: Initial conditions influence the performance of PSO. If the initialization is inadequate, then the algorithm may search an undesirable area, and it will lead to the wrong optimal solution. The performance of CPSO is sensitive to the initialization of the swarms.

(ii) **Constriction Factor**: The random weights controlling the CPSO parameters may cause a kind of explosion as the velocities of the particles, and positional coordinates grow toward infinity. This explosion has been traditionally restricted by an upper bound on the particle velocity, or to the step-size. Even so, the implementation of a good constriction factor can inhibit explosion and improve convergence to local optima.

(iii) **Inertia Weight (w)**: It balances the exploration–exploitation trade-off. A big value of w increases the exploration, and a small value strengthens the exploitation. Some used a constant inertia weight; others employed a linearly decreasing w and a third group used nonlinearly decreasing inertia weight.

(iv) **Mutation Operator**: It expands the performance of PSO and permits escaping from the local minima. Different variants of PSO using the mutation of the global best particle and the mutation of the local best particle were suggested to prevent the PSO from stagnation in local minima.

Nature-inspired procedures, e.g., genetic algorithms (GAs) [29], simulated annealing (SA) [30], differential evolution (DE) [31], evolutionary programming (EP) [32], artificial immune systems (AISs) [33], ant colony optimization (ACO) [34], can be applied to an extensive variety of global optimization problems. These algorithms are advantageous in the optimization of intricate problems and are typically population-based metaheuristics.

Mingling PSO with GA is quite popular as it is the case with combining PSO and the DE algorithms. The PSO–ACO2 combination is better for solving discrete optimization problems than PSO–ACO [35]. Algorithms such as tabu search (TS), SA, fuzzy logic can also be employed [19, 36, 37], and [38]. As far as local search methods go the Nelder–Mead simplex, Quadratic Programming (QP) and Interior Point (IP) can be combined with PSO.

PSO methods can be associated with other algorithms as follows. (1) One procedure works as a pre-optimizer for the preliminary population of the next algorithm; (2) The entire population is divided into subpopulations, which evolve using PSO and other algorithms; and (3) The unique operators of an algorithm are inserted as local search improvement for the other algorithm.

HPSO methods have been applied to solve an assorted variety of problems encompassing image processing, remote sensing, data clustering, and engineering problems to name a few. Hence, there is still possibility of new SR applications.

GA-PSO, DE-PSO, and PSO with adaptive mutation (AMPSO) are commonly used HPSO strategies to solve unconstrained global optimization problems. GA-PSO combines the PSO and GA algorithms; DE-PSO is a hybrid version of DE with PSO and AMPSO is a mixture of PSO and EP.

HPSO algorithms are still a motivating and promising field that can offer additional comprehensions regarding the behavior and prospective benefits and disadvantages of several metaheuristics. This chapter may motivate and help to develop new hybrid models or to employ the existing models to new applications.

6.2 A Hybrid PSO (HPSO) Algorithm for SR Imaging

PSO has been effectively for SR [38]. When conventional GA and PSO find problems to determine the global optimum, HPSO approaches relying on two GA concepts: subpopulation and crossover. They are incorporated into the CPSO method to improve the accuracy of that conventional GA and CPSO since these traditional procedures cannot find the global optimum when there is a huge number of parameters to be estimated.

The particles are split into $m = 1, \ldots, M$ subpopulations where each one has its personal best optimum $\boldsymbol{g}_{\boldsymbol{sub\text{-}best\text{-}m}}$. The PSO process is applied for each subpopulation. If $\boldsymbol{g}_{\boldsymbol{sub\text{-}best\text{-}m}}$ is better than $\boldsymbol{g}_{\boldsymbol{best}}$, then $\boldsymbol{g}_{\boldsymbol{best}}$ is replaced by the $\boldsymbol{g}_{\boldsymbol{sub\text{-}best\text{-}m}}$ where m is the subpopulation number.

The $\boldsymbol{g}_{\boldsymbol{sub\text{-}best\text{-}i}}$ are organized in ascending order of fitness function values. The top two $\boldsymbol{g}_{\boldsymbol{sub\text{-}best\text{-}i}}$ are taken as parents ($\boldsymbol{x}_i$, and $\boldsymbol{x}_j$) for a crossover with i and j as their corresponding subpopulation number. The offspring are produced for each by arithmetic crossover, using the relationships

$$\boldsymbol{x}_i' = r\,\boldsymbol{x}_i + (1-r)\boldsymbol{x}_j, \text{ and} \tag{7}$$

$$\boldsymbol{x}_j' = r\,\boldsymbol{x}_j + (1-r)\boldsymbol{x}_i, \tag{8}$$

with velocities

$$\boldsymbol{v}_i' = \boldsymbol{v}_i V, \tag{9}$$

$$\boldsymbol{v}_j' = \boldsymbol{v}_j V, \text{ and} \tag{10}$$

$$V = \left(\boldsymbol{v}_i + \boldsymbol{v}_j\right) / \left|\left|\boldsymbol{v}_i + \boldsymbol{v}_j\right|\right|, \tag{11}$$

where r is uniformly distributed between 0 and 1. The offspring replaces the worst particle in the same subpopulation.

HPSO-Based SR Algorithm

(1) Initialize the swarm randomly with initial values for variables $\boldsymbol{x}_i^0$, $\boldsymbol{v}_i^0$, $\boldsymbol{p}_{\boldsymbol{besti}}$, $\boldsymbol{g}_{\boldsymbol{best}}$
(2) For each particle i, repeat until the whole population has been analyzed:
(3) Set initial values for variables $\boldsymbol{x}_i^0, \boldsymbol{v}_i^0, \boldsymbol{p}_{\boldsymbol{besti}}, \boldsymbol{g}_{\boldsymbol{best}}$
(4) Generate M subpopulations
(5) Perform crossover
(6) If $f\left(\boldsymbol{x}_i^k\right) > f(\boldsymbol{p}_{\boldsymbol{besti}})$ then $\boldsymbol{p}_{\boldsymbol{besti}} = \boldsymbol{x}_i^k$
(7) Compute $\boldsymbol{g}_{\boldsymbol{sub\text{-}best\text{-}m}}$ for each subpopulation m

(8) If $f\left(\boldsymbol{x}_i^k\right) > f(\boldsymbol{g}_{best})$ then $\boldsymbol{g}_{best} = \boldsymbol{x}_i^k$
(9) If it converges, then stop.
(10) If the maximum number of iterations is not reached, then go to step 2.

6.3 PSO Drawbacks

CPSO works fast for optimizing complex multidimensional spaces with a broad range of applications, but it falls easily into a local optimum in high-dimensional spaces, and it converges slowly. This limitation is known as swarm stagnation (SS). There are two puzzling areas for future progress: SS and Dynamic Environments [5].

Prospective users can circumvent SS by certifying that particles continue moving. Although some methods have been suggested for this, the problem is to distinguish an optimal solution. Numerous mechanisms were proposed to restrict particle velocities and to guarantee maximal exploration, such as (i) slower particles that complete more fitness function evaluations while moving to the final solution, and (ii) to introduce mechanisms to increase the likelihood to find a global optimum. Hybrids using other nature-inspired paradigms addressed PSO's inherent challenges to reposition stagnant particles or slowing down those using subpopulations.

Dynamic Environments imply the existence of objective function values that diverge over time and are difficult to handle when the optimum position travels thru the problem space in steps larger than the best candidate solution of the group can follow naturally. CPSO cannot track the optimum because unless the goal moves to the space of another swarm member, the swarm will continue to move to its previous best, set at the previous goal position. Re-resetting the swarm or using the old swarm with new best values helps mitigate complications. The major limitation to effective optimization when using PSO in Dynamic Environments is when the optimum position has velocity superior to the particle velocity and therefore escapes the swarm. Dynamic problems require a balance between velocity constraint and the need to track optima. Slow-moving optima can be pursued naturally since the particles keep on moving about in an area related to the velocity of the particles. Complications happen when the optimal displacements are larger than the particle dislocations. The fundamental norm is to keep swarm diversity, which increases the likelihood that the optimum will shift to an adequate area.

7 Case Studies of the Use of PSO in SR

To some extent, the determination of the appropriate model is the basis for finding the solution. Bearing in mind computational intelligence, the image or the video sequence with the highest resolution corresponds to the optimal solution obtained by PSO. This established optimization method is on the forefront for its implementation simplicity, higher computational accuracy, and lower time cost in complex applications when compared to other heuristic methods. In PSO, a particle is a potential solution vector to the problem. Nevertheless, the parameter selection affects the performance intensely, and schemes that work well for all problems are absent. In many cases, the PSO parameters have to be adjusted several times to obtain adequate precision and escape the fixed parameters limitation during the whole optimization process.

This section illustrates the use of SR in still images and video sequences using the PSO and HPSO algorithms stated beforehand [39].

7.1 SR in Still Images

Geometrical SR restoration is possible if and only if the input LR images have been undersampled, and hence, they contain aliasing. Since there is aliasing, the high-frequency content of the sought reconstructed image is embedded in the LR content of each of the observed LR images. Given an appropriate number of LR observations, and if the set of observations differ in phase (i.e., the scene images are shifted by subpixel quantities [1, 18, 40–42]), then the phase evidence can help to isolate the aliased high-frequency part from the true low-frequency portion, and the full-resolution image can be precisely recreated.

There exist both single-frame and multiframe SR alternatives. Multiframe SR explores the subpixel shifts between multiple LR images of the target scene. It produces a better resolution image resulting from combining data from all the LR pictures, and the created HR images are superior scene descriptions. Single-frame SR approaches attempt to enlarge the image without blurring effects. These procedures employ other parts of the LR images, or other distinct images, to predict what the HR image should appear. Nowadays, SR methods work well with both grayscale and color images.

Figures 4, 5, and 6 show results of applying PSO and HPSO to the image Lena.

7.2 SR in Video

The SR main conception is to add new information to a target frame by detecting the subpixel displacements between this frame and its nearby ones. These subpixel

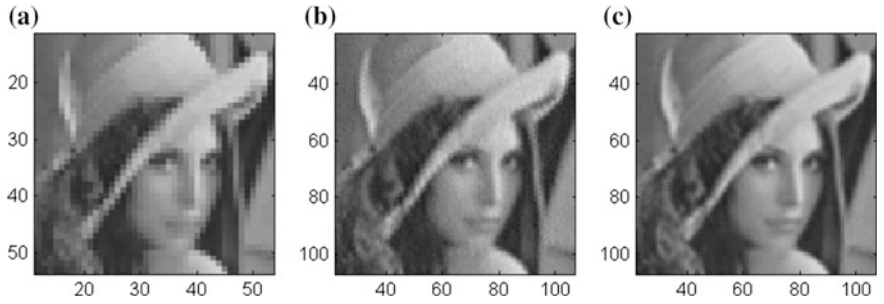


Fig. 4 **a** LR image; **b** PSO result; and **c** HPSO result

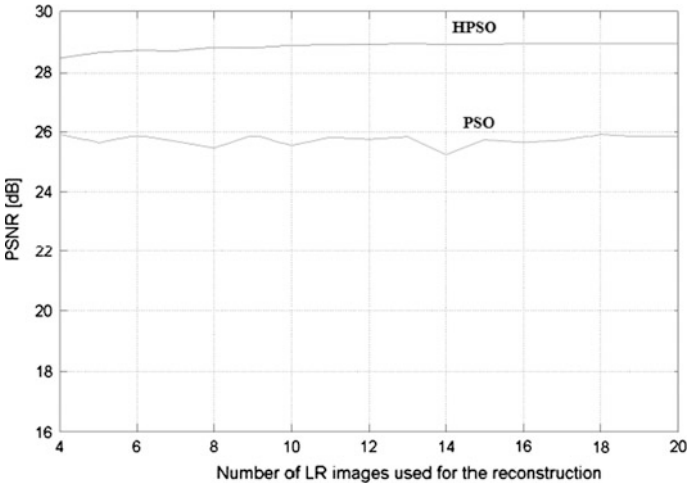


Fig. 5 PSNR curves for Lena as the number of LR images increase

shifts are created by the camera and/or by the objects moving in the scene. Figure 7 shows the motivation behind VSR; that is, an SR frame is generated by combining four consecutive LR frames.

The importance of a good ME in SR naturally arises. Independently of the SR method used, the underlying ME technique must be as accurate as possible [17, 24, 40, 41, 43, 44]. In literature, little attention has been paid to the performance of ME techniques when used in SR problems. The notation for an entire image is $\boldsymbol{g_l}$, representing the lth image in the LR sequence. $\boldsymbol{g_k}$ denotes the image whose resolution must be improved.

Let $\boldsymbol{g_l}(\boldsymbol{x})$ be a pixel located at $\boldsymbol{x} = (x, y)^{\mathrm{T}}$ in the lth LR image, which is $M \times N$. $\boldsymbol{g_l}$ is divided into blocks $\boldsymbol{bx}$ of size $m \times n$ such that $\boldsymbol{g_l}(\boldsymbol{bx_p})$ with $p = 1, \ldots, P$ and P is the total number of blocks.

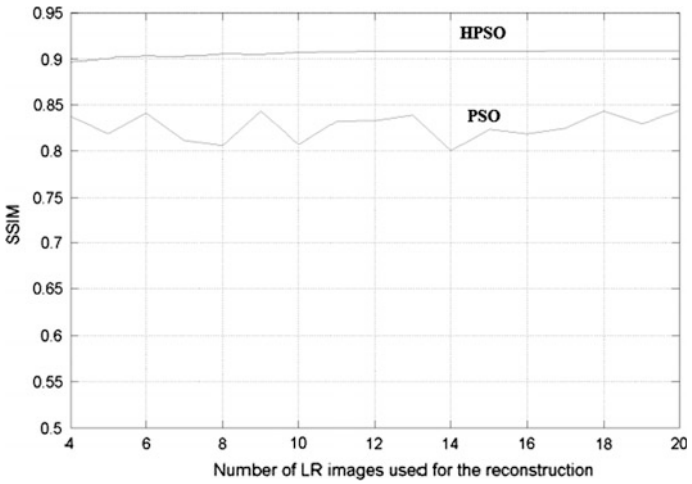


Fig. 6 SSIM evolution for Lena as the number of LR images increase

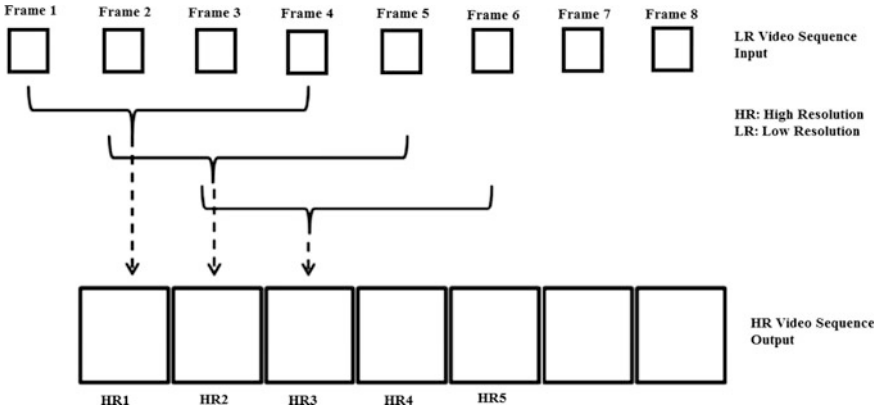


Fig. 7 In VSR processing, each HR frame is a combination of N successive LR images

Accurate ME comprises a very important part in SR problems. Pixels with bad and/or contradictory MEs due to occlusions and poor ME results degrade the SR reconstruction and consequently should not be considered.

Quantitative assessments between the enhanced and the original images can be done using the *PSNR*. Inaccurate MVs can be detected by applying the displaced frame difference (*DFD*) between the upsampled and the compensated frames given by

$$\mathrm{DFD}(\boldsymbol{g}_l, \boldsymbol{d}^{lr}_{l,k}, \boldsymbol{g}_k)(\boldsymbol{x}) = \left| \boldsymbol{g}_k(\boldsymbol{x}) - \boldsymbol{g}_l(\boldsymbol{x} - \boldsymbol{d}^{lr}_{l,k}) \right| \tag{12}$$

The *DFD* allows identifying pixels in $\boldsymbol{g}_k$ that are not anticipated by the MV estimates $\boldsymbol{d}^{lr}_{l,k}$ (between the LR images lth and kth) and $\boldsymbol{g}_l$. Large *DFD* values indicate nonpredictable/observable pixels in $\boldsymbol{g}_k$.

After estimating the motion between frames, it is possible to project the LR image on the sought after HR grid with the help of the MVs. An HR grid can be defined by locating the pixel values of the present image, while the remaining HR positions stay empty (zero MVs). A different frame having subpixel shifts with respect to the current image has new data placed onto the HR grid. An explanation of the SR algorithm steps follows:

(1) Align the LR image with the matching positions of the HR image.
(2) If there is subpixel motion, then new pixels from frames $l = k$ will be placed in empty locations of the HR grid.
(3) After filling all possible places, empty HR pixel values will be interpolated to obtain the final SR image.
(4) Use the estimated motion fields to render the HR reconstruction.
(5) Calculate the *PSNR* values between the original HR image $\boldsymbol{I}$ and the resulting SR image.

Figure 8 shows three regions R1, R2, and R3 belonging to a frame of the mobile sequence. This subsection explores the process of obtaining SR versions of these patches. Figure 9 presents 5 LR frames of the mobile sequence. Figure 10a shows the PSNR curves for the reconstructions done for R1, which is an area with texture

Fig. 8 Region 1 (R1) has a textured area with translational motion, Region 2 (R2) shows a textured area with independent object motion, and Region 3 (R3) contains a flat area with translational motion

Fig. 9 Mobile LR sequence obtained using from frame 7 to frame 11

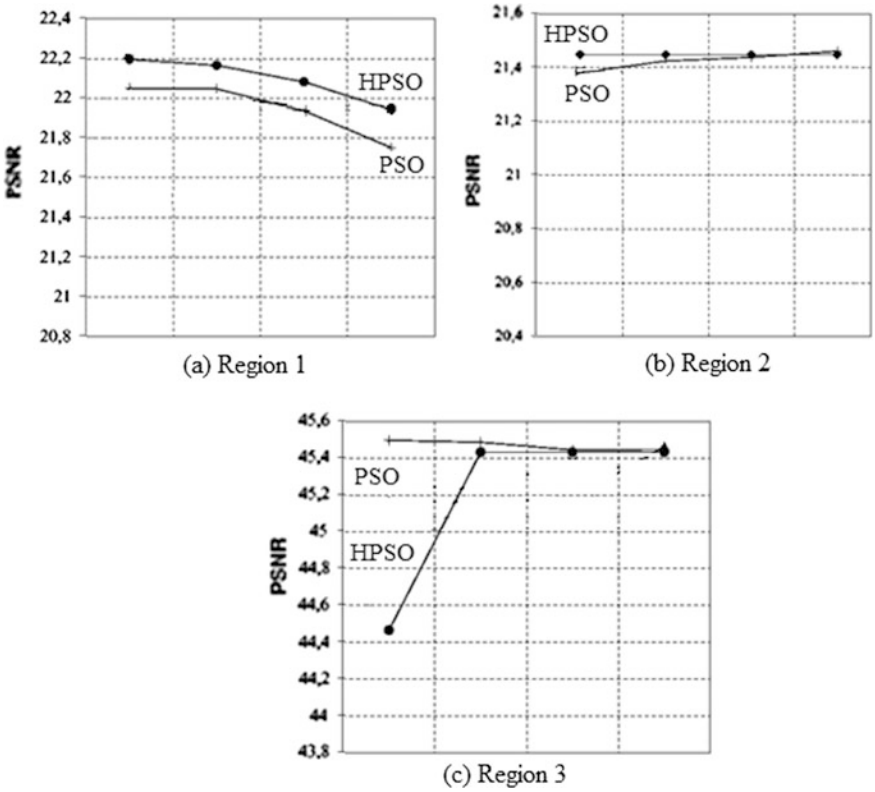


Fig. 10 *PSNR* values versus number of LR frames for the mobile sequence: **a** R1, **b** R2, and **c** R3

but also some extremely flat subregions. In this case, the PSNR values improve for the PSO and HPSO algorithms. Regions R2 and R3 contain objects with motions. The most remarkable conclusion on the R2 is that when the motion does not depend from the camera motion, increasing the accuracy of the ME can introduce errors to the global SR process. The advance achieved by the use of the DFD is perceptible in Fig. 10b where the motion complexity is high (rotation and translation). Lastly, R3 has a flat region with translational motion and the PSNR results are depicted in Fig. 10c.

8 Conclusions

Metaheuristic optimization algorithms like PSO have become a widespread choice for solving complex and intricate problems such as SR reconstruction, which are otherwise demanding to solve by traditional algorithms. This chapter attempts to examine the PSO algorithm and some of its hybrid versions. Hybridization means combining two (or more) methods in a sensible way such that the resultant algorithm contains the positive characteristics of both (or all) the procedures. Interesting techniques to use in hybridization include many local and global search approaches. Results from the use of SR reconstruction for still and video images are shown for the PSO and the HPSO algorithms.

References

1. Fernandes SR, Estrela VV, Magalhaes HA, Saotome O (2014) On improving sub-pixel accuracy by means of B-Spline, In: Proceedings of the 2014 IEEE international conference on imaging system and techniques (IST 2014), Santorini, Greece. doi:10.1109/IST.2014.6958448
2. Katsaggelos AK, Molina R, Mateos J (2007) super resolution of images. Morgan & Claypool Publishers, USA
3. Rivera LA, Estrela VV, Carvalho PCP (2004) Oriented bounding boxes using multiresolution contours for fast interference detection of arbitrary geometry objects. In: Proceedings of 12-th international conference in central Europe on computer graphics, visualization and computer vision (WSCG 2004), Plzen, Czech Republic
4. Su H, Wu Y, Zhou J (2012) Super-resolution without dense flow. IEEE Trans Image Process 21(4):1782–1795
5. Irani M, Peleg S (1990) Super resolution from image sequences. In: Proceedings of the ICPR, vol 2, pp 115–120
6. Tsai RY, Huang TS (1984) Multiframe image restoration and registration. In: Advances in computer vision and image processing, vol 1, chap 7, pp 317–339, JAI Press, Greenwich, Conn, USA
7. Anitha J, Vijila CKS, Hemanth DJ (2011) An overview of computational intelligence techniques for retinal disease identification applications. Int J Rev Comput 5(1):29–46
8. Coelho AM, Estrela VV, do Carmo FP, Fernandes SR (2012) Error concealment by means of motion refinement and regularized bregman divergence. In: Proceedings of the 13th International conference on intelligent data engineering and automated learning (IDEAL'12), Natal, Brazil, pp 650–657. doi:10.1007/978-3-642-32639-4_78
9. Anitha J, Vijila CKS, Hemanth DJ (2009) Comparative analysis of GA and PSO algorithms for abnormal retinal image classification. Int J Recent Trends Eng 2(3):143–145
10. Hemanth DJ, Anitha J (2013) Modified cross-over techniques in genetic algorithms for performance enhancement of retinal image classification system, In: Proceedings of the 3rd international conference on computational intelligence and information technology, pp 28–34, IET, Mumbai, India, doi: 10.1049/cp.2013.2587
11. Jestin VK, Anitha J, Hemanth DJ (2011) Genetic algorithm for retinal image analysis. Int J Comput Appl 1:48–52
12. Rouet IM, Jacq II, Roux C (2000) Genetic algorithms for a robust 3-D MR-CT registration. IEEE Trans Inf Tech Biomed 4(2):126–136

13. Wyawahare MV, Patil PM, Abhyankar H (2009) Image registration techniques: an overview. Int J Sig Process Image Process Pattern Recogn 2:11–28
14. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the 1995 IEEE international conference on neural networks, IEEE, pp 1942–1948. doi:10.1109/ICNN.1995.488968
15. Li Q, Sato I, Murakami Y (2007) Affine registration of multimodality images by optimization of mutual information using a stochastic gradient approximation technique. In: Proceedings of the 2007 IEEE international geoscience and remote sensing symposium (IGARSS 2007). doi:10.1109/IGARSS.2007.4422814
16. http://www.infognition.com/articles/. Retrieved on 21 Nov 2016
17. Jesus MA, Estrela VV, Saotome O (2016) Super-resolution in a nutshell. In: Proceedings of the Brazilian technology symposium 2016 (BTSym 2016). ISSN 2447-8326. V. 1
18. Silva Filho P, Rodrigues M, Saotome O, Shiguemori EH (2014) Fuzzy-based automatic landmark recognition in aerial images using ORB for aerial auto-localization. In: Bebis G et al (eds) Advances in visual computing, vol 8887, LNCS, 467-476, Springer, Berlin
19. Feng K, Zhou T, Cui J, Tana J (2014) An example image super-resolution algorithm based on modified k-Means with hybrid particle swarm optimization. optoelectronic imaging and multimedia technology III. In: Dai Q, Shimura T (eds) Proceedings of SPIE ,vol 9273, 92731I. doi:10.1117/12.2073216
20. Altunbasak Y, Patti AJ, Mersereau RM (2002) Super-resolution still and video reconstruction from MPEG-coded video. IEEE Trans Circuits Syst Video Tech 12:217–226
21. Singh P, Sharma J (2016) A review paper on image denoising by low rank matrix decomposition and genetic algorithm. Int J Eng Sci Res Technol 5, 7:166–681. doi:10.5281/zenodo.56923
22. Reibman AR, Bell RM, Gray S (2006) Quality assessment for super-resolution image enhancement. In: Proceedings of IEEE international conference on image processing. USA, pp 2017–2020
23. Yang J, Wright J, Huang T, Ma Y (2010) Image super-resolution via sparse representation. IEEE Trans Image Process 19(11):2861–2873
24. Brandi F, de Queiroz RL, Mukherjee D (2008) Super-resolution of video using key frames and motion estimation. In: Proceedings of IEEE international conference on image processing, USA, pp 321–324
25. Al-Najjar YAY, Soong DC (2012) Comparison of image quality assessment: PSNR, HVS, SSIM, UIQI. Int J Sci Eng Res 3(8). ISSN 2229-5518
26. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Proc 13(4):600–612
27. Lukeš T, Hagen GM, Křížek P, Švindrych Z, Fliegel K, Klíma M (2014) Comparison of image reconstruction methods for structured illumination microscopy. In Proceedings of SPIE 9129
28. Imran M, Hashima R, Khalidb NEA (2013) An overview of particle swarm optimization variants. In: Proceedings of the 2012 Malaysian Technical Universities conference on engineering and technology (MUCET 2012), Part 4: information and communication technology, Procedia Eng 53:491–496
29. Holland JH (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. University of Michigan Press, Ann Arbor, MI
30. Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220(4598):671–680
31. Price K, Storn R (1995) Differential evolution—a simple and efficient adaptive scheme for global optimization over continuous spaces. Technical report, International Computer Science Institute, Berkley
32. Fogel LJ, Owens AJ, Walsh MJ (1965) Artificial intelligence through a simulation of evolution. In: Maxfield M, Callahan A, Fogel LJ (eds) Biophysics and cybernetic systems, In: Proceedings of the 2nd cybernetic sciences symposium, pp 131–155

33. Timmis J, Neal M, Hunt J (2000) An artificial immune system for data analysis. BioSystems 55(1):143–150. doi:10.1016/S0303-2647(99)00092-1
34. Dorigo M, Stutzle T (2004) Ant colony optimization (ACO)
35. Holden N, Freitas AA (2008) A hybrid PSO/ACO algorithm for discovering classification rules in data mining. J Artif Evol Appl 2008. doi:10.1155/2008/316145
36. Anitha J, Vijila CKS, Selvakumar AI, Hemanth DJ (2012) Performance improved PSO based modified kohonen neural network for retinal image classification. J Chin Inst Eng 35(8): 979–991
37. Hemanth DJ, Vijila CKS, Anitha J (2010) Performance improved PSO based modified counter propagation neural network for abnormal mr brain image classification. Int J Adv Soft Comput Appl 2(1):65–84
38. Lin CL, Mimori A, Chen YW (2012) Hybrid particle swarm optimization and its application to multimodal 3D medical image registration. Comput Intell Neurosci 2012. doi:10.1155/2012/561406
39. Thangaraj R, Pant M, Abraham A, Bouvry P (2011) Particle swarm optimization: hybridization perspectives and experimental illustrations. Appl Math Comput 217:5208–5226
40. Costa GH, Bermudez J (2009) Registration errors: are they always bad for super-resolution? IEEE Trans Signal Proc 57(10):3815–3826
41. Robinson D, Milanfar P (2006) Statistical performance analysis of super-resolution. IEEE Trans Image Process 15(6):1413–1428
42. Tom BC, Galatsanos NP, Katsaggelos AK (2002) Reconstruction of a high resolution image from multiple low resolution images. In: Chaudhuri S (ed) Super-resolution imaging. Kluwer, Norwell
43. Coelho AM, Estrela VV (2012) EM-based mixture models applied to video event detection. In: Sanguansat P (ed) Principal component analysis—engineering applications, InTech open. doi:10.5772/38129
44. Estrela VV, Magalhaes HA, Saotome O (2016) Total variation in computer vision. In: Kamila NK (ed) Handbook of research on emerging perspectives in intelligent pattern recognition, analysis, and image processing. doi:10.4018/978-1-4666-8654-0.ch002