



Hybrid Wolf Pack Algorithm and Particle Swarm Optimization Algorithm for Breast Cancer Diagnosis

Badriya Al Maqbali

*Department of Process Engineering,
International Maritime College Oman
Suhar Port Falaj Al Qabail, Suhar OM, Oman
badriyaalm15@gmail.com*

Abstract: In women, breast cancer is deadly disease which is increased the death rate of women. By exploiting the mammogram images, a precise and early recognition of breast cancer is a complex task. Therefore, a new breast cancer recognition technique was proposed that considered five important stages: segmentation, preprocessing, feature extraction, feature selection as well as classification. Initially, by exploiting the median filtering as well as Contrast Limited Adaptive Histogram Equalization (CLAHE), input mammogram images are preprocessed. Subsequently, through the region growing method, the preprocessed images are fed to segmentation. Then, from the segmented image, texture, geometric and gradient features are extracted. The feature vector length is higher, it is important to choose optimal features. Moreover, the optimal features chosen are performed using the proposed optimization method. After completing the selection of the optimal features, they are fed to the classification procedure including the Neural Network (NN) classifier. As an innovation, to improve the precision of diagnosis (benign as well as malignant), the NN weight is chosen optimally. The NN weight optimization and the optimal feature selection are attained using the Hybrid Wolf Pack Algorithm (WPA) and Particle Swarm Optimization (PSO) Algorithm called the Hybrid WPA-PSO algorithm. At last, the performance analysis is performed between the proposed and conventional techniques.

Keywords: Breast Cancer, Feature Extraction, Mammogram Image, Neural Network, Segmentation,

Nomenclature

Abbreviations	Descriptions
SNR	Signal to Noise Ratio
EM	Electromagnetic
NN	Neural Network
COIS	Custom Optical Imaging System
SR	Spatial Resolution
MEMS	Micro Electro Mechanical Systems
WPA	Wolf Pack Algorithm
CLAHE	Contrast Limited Adaptive Histogram Equalization
PSO	Particle Swarm Optimization
CNR	Contrast to Noise Ratio
IDC	Invasive Ductal Carcinoma
MRI	Magnetic Resonance Image
AI	Artificial Intelligence
CDF	Cumulative Distribution Function

1. Introduction

In worldwide, due to breast cancer, each year over a half-million women die. In numerous countries, to minimize breast cancer-associated screening, morality for breast cancer is developed around the globe over the last three decades. In treatment, screening together with enhancements has ensued in a minimization in the mortality rate of breast cancer, however, this disease is still considered as the number one reason for female cancer death [1].

Many of the clinical examinations are based upon the acoustic wave interaction or EM by means of fluids as well as human body tissues however on a similar frequency range, not all the tests work. Xrays

are commonly exploited as it presents more reliable diagnostic consequences in spite of the diverse kinds of clinical imaging examinations. It discharges electromagnetic radiations that are maximum energy and even though for several applications ionizing radiations were shown as advantages in humans. It has the ability to generate detrimental effects yet when performing diagnosis examination by exploiting X-rays. Hence, the probability to generate enduring alters in the cells its exploits in biological tests rises that will trigger cell mutation till its death [2]. Additionally, there subsists pseudo alarm probability when using these systems for the tumor diagnosis. For example, breast tumor detection have analyzed raised rate of false positives as well as false negatives. Even though in early diagnosis, other modalities namely the MRI might help, the technique undergoes from few limitations for instance the raised costs, detachment for a huge amount of diseased people, and defective target detection for lesser diameter tumors because of minimized distance among the MRI slices and target [12] and [13].

For breast cancer diagnoses, several imaging techniques could be exploited Breast Ultrasound, Mammograms, SNR, and Breast MRI. However, it is constricted by numerous modules like CNR and the SR undergoes a difficult errand. During breast rationing entire tumor evacuation medical process stays challenging due to the nonattendance of the ideal intra-operative tumor edge appraisal system [3].

The main contribution of this work is to propose a highly developed breast cancer recognition model by preprocessing, feature extraction, Segmentation, feature selection, and classification. By exploiting median filtering as well as the CLAHE input mammogram image is preprocessed and that image is segmented through region growing algorithm. The features such as geometric, texture, and gradient features are extracted from the segmented image at the time of the feature extraction phase, and the optimal features are fed to the classification in the NN classifier. Using the developed optimization model, the optimal feature selection and NN weight update are attained, so that the precise of recognition is high.

2. Literature Survey

In 2020, Mohamed Hisham Aref et al [1], examined breast tissues presents several reactions to light absorption, transmission, as well as especially the reflection against the spectral range. To examine variable responses a COIS was modeled to monochromatic LEDs to spotlight the diversity in the reflectance properties of normal/malignant tissue. In 2020, D. Ghieh et al [2], worked on the method to treatment for breast lesions. Here, they have presented detailed guidelines and breast image modalities for breast imaging. Moreover, novel upcoming methods in this study present a betterment for diagnose method in dense breast tissue. In 2021, Meha Desai and Manan Shah [3], worked on the network functioning and its modeling and subsequently breast malignancy diagnosis and its accuracy. This was performed using the network and finally, it determines which network performs better than the other network. Here, Convolutional neural network performance was superior to the Multilayer perceptron in terms of detection and diagnosis the breast cancer. In 2018, Kihan Park et al [4], designed a convenient diagnostic tool and a MEMS-based biochip, that was combined with a microheater for temperature control and a piezoresistive sensing layer to measure the reaction force. Finally, for breast cancer, examine the electromechanical coupling factor was exploited as a biomarker. For breast tissue, the piezoelectric model has explained with preliminary experimental outcomes on five sets of normal as well as IDC examples in 25-45°C temperature range. In 2020, Ioannis Sechopoulos et al [5], worked on breast cancer detection in digital breast tomosynthesis as well as digital mammography. Mainly, this research was examined the present abilities of AI.

3. Proposed Method for Diagnosis of Breast Cancer

The proposed breast cancer detection model is shown in Fig 1. The adopted model tried to model standard and highly developed detection technique by considering five main phases as Pre-processing, feature extraction, segmentation, feature selection, and finally classification. Initially, the CLAHE and median filtering model is exploited for preprocessing the input image.

The median filtering and CLAHE results are indicated Img_{MM} and Img_{CLAHE} correspondingly. By exploiting the Region Growing algorithm the segmentation process is performed for the pre-processed image Img_{MM} . Img_{SEG} is indicated as the ensuing result of the segmented image.

Subsequently, features such as Texture ($f^{Texture}$), Geometric feature ($f^{Geometric}$), and Gradient ($f^{Gradient}$) are extracted from Img_{SEG} . By considering $f^{Geometric}$, $f^{Texture}$ and $f^{Gradient}$ the integrated feature set attained and it is indicated as f_1^{feat} . The extracted feature vector length is high and therefore,

they are required to optimize. Through the NN classifier the optimally chosen features $f_i^{\text{feat}^*}$ are fed to the classification procedure. From NN, the classification output presents the diagnostic attainment if the equivalent image belongs to malignant else benign. Using the hybrid WPA and PSO method, the optimal feature selection is performed. In addition, the NN weight at the time of training is updated using the developed technique formula. Actually, in NN, the optimal feature selection with optimized weight is performed so that the detection or diagnosis accuracy is obtained utmost.

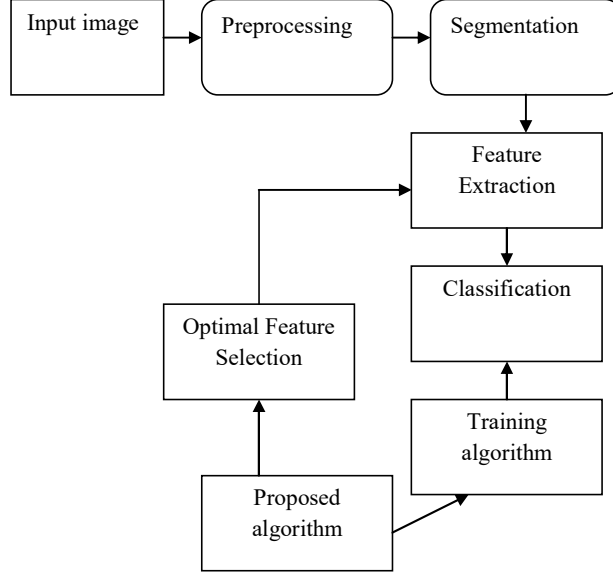


Fig. 1. Architecture model of the proposed method

3.1 Preprocessing

The initial process is the Pre-processing it is carried out using the CLAHE and the median filtering. Img_{in} indicates as the input mammogram.

3.1.1 CLAHE: On the basis of the five main processes, the CLAHE pipeline operation is performed [6]. At the starting phase, the mammogram image is decomposed into the equivalent-sized rectangular blocks with the histogram alterations that occurred. Through via CDF the mapping function is carried out in the clipped histogram. In eq. (1), the clip point mathematical formulation is indicated, in that the number of the pixels, pixels dynamic range in every block are indicated as S and N_i , correspondingly. In addition, Max_{slope} denotes the maximum slope, β denotes the clip factor, correspondingly. In eq. (2), for CDF (f_{CDF}), the mapping function is to remap the image blocks in the greyscale image, in that the grey level of a pixel is shown by exploiting the variable f .

In eq. (3), the mathematical formulation of the remapping function (f_{remap}) is indicated.

Moreover, the utmost number of pixels in every block is indicated as max_1 . From the CLAHE operation, the image attained is indicated as Img_{CLAHE} .

$$\alpha = \frac{S}{N_i} \left(1 + \frac{\beta}{100} Max_{slope} \right) \quad (1)$$

$$f_{CDF} = \sum_{z=0}^a I_{pdf} \quad (2)$$

$$f_{remap} = f_{CDF} \times max_1 \quad (3)$$

3.1.2 Median Filtering Model

From CLAHE, the results are fed to the median filtering [7] to remove arbitrary noise from the image. Eq. (4) states median filtering formulation, in that $Img_{MM}(x_i, y_i)$ as well as $Img_{CLAHE}(x_i, y_i)$ indicates output image, as well as input CLAHE, operated image. The 2-D mask with the size $n \times n$ is indicated as T , in that \hat{j} and \hat{k} is the pixels in the image.

$$Img_{MM}(x_i, y_i) = \text{med} \left\{ Img_{CLAHE}(x_i - \hat{j}, y_i - \hat{k}) \mid \hat{j}, \hat{k} \in T \right\} \quad (4)$$

3.2 Segmentation using Region Growing

For tumor segmentation, pre-processed image Im_{gMM} is fed to a region growing algorithm. One of the easiest and elegant region growing techniques is the seeded region growing technique. The image segmentation with this method is attained by means of considering the neighbor pixel that is adjacent to a set of points which are indicated as seed points. In [8], a clear description of the seeded region algorithmic procedure is described.

3.3 Feature Extraction

In this work, features such as Gradient features Geometric features, as well as Texture, are extracted [9] to diagnose breast cancer.

3.3.1 Geometric Features

The properties of geometric associated with the ROI are shown using the geometric features and these features in the image indicating the pixels group. The fundamental features of the geometric features comprise such as the perimeter, area as well as compactness in the scalar form. Amid all other properties, the area measure is represented as the simple regarding size as well as ROI extraction indicates the number of pixels. Here, the perimeter is stated as the perimeter of the region. The region compactness is represented on the basis of both the perimeter and area. Using compactness, the object's roundness is calculated. The mathematical formulation for perimeter (P_i), area (A_i), and compactness (U) are indicated in Eq. (5), Eq. (6), and Eq. (7), correspondingly. In addition, the pixels are indicated as w, x within a shape. In Eq. (1), the ROI vector content in w location and the ROI vector content in x location are represented as $X_i - ROI []$ and $Y_i - ROI []$, correspondingly.

The w, x coordinates of the pixel creating the curve are indicated in vector outline as $W - edge [perimeter]$ and $X - edge [perimeter]$, correspondingly in Eq. (5). The extracted geometric feature is indicated as $f^{Geometric}$.

$$A_i = \sum_w \sum_x (A_{i_{w,x}}, W-ROI[A_i] = w, X-ROI[A_i] = x) \quad (5)$$

$$P_i = \sum_w \sum_x P_{i_{w,x}}, (W-edge [P_i] = w, X-edge [P_i] = x) \quad (6)$$

$$U = \frac{P_i^2}{A_i} \quad (7)$$

3.3.2 Texture Features

The coarseness, smoothness, as well as image regularity, are measured by the texture features. The value of the mean pixel is indicated as the mean within ROI and it is important to determine ROI brightness. Eq. (8), indicates the numerical formulation for mean. As in Eq. (9), using the mean global area, the contrast ratio among the complete image as well as the specific region is decided. By exploiting eq. (10), regarding the uniformity in the intensity is calculated in the histogram associating with the ROI. Using eq. (11), the image standard deviation (SD) is calculated. As per Eq. (12), the relative smoothness of intensity in ROI is calculated.

$$Mean = \sum_{w=0} [I_i W-ROI[w], X-ROI[x]] \quad (8)$$

$$Mean\ Global = I_i [W-ROI[w], X-ROI[x]] \quad (9)$$

$$Uniformity = \sum_{w=0}^{A_i} HN^2 [I_i [W-ROI[w], X-ROI[w]]] \quad (10)$$

$$SD = \sum_{w=0}^{A_i} (I_i [W-ROI[w], X-ROI[w]] - Mean)^2 \quad (11)$$

$$Smoothness = \frac{1}{1 + SD^2} \quad (12)$$

By exploiting eq. (13), by means of real-valued probability distribution asymmetry related to the arbitrary variable is measured by exploiting eq. (13). In eq. (14), by exploiting the entropy the disorders are calculated, and also it represents the mathematical formulation of entropy in the greyscale domain. By exploiting eq. (15), the image correlation is calculated, the normalized histogram vector is indicated as HN as well as the original mammographic image is indicated as $I_i [] \leftarrow Im_{gSEG}$. By exploiting the variable $f^{Texture}$ the extracted texture feature is indicated.

$$Sk = \sum_{w=0}^{Ai} I_i [W-ROI[w], X-ROI[w]] - \text{Mean} \quad (13)$$

$$\text{Entropy} = \frac{HN(I_i [W-ROI[w], X-ROI[w]])}{\log HN [W-ROI[w], X-ROI[w]]} \quad (14)$$

$$\text{Correlation} = \sum_{w=0}^{Ai} \frac{HN(I_i [W-ROI[w], X-ROI[w]])}{I_i [W-ROI[w], X-ROI[w]]} \quad (15)$$

3.3.3 Gradient Features

The gradient image is indicated as derivation associating with the local image. The gradient features are represented as sobal mean (SM), sobal uniformity (SU), sobal entropy (SE), sobal mean global area (SMG), sobal standard deviation (SSD), sobal skewness (SK), sobal smoothness (SS), as well as sobal correlation (SC). The gradient image vector and the gradient normalized histogram and are shown as $GR[]$ GHN. The variable $f^{\text{Geometric}}$ signifies extracted gradient features.

$$SM = \sum_{w=0}^{Ai} GR [W-ROI[w], X-ROI[w]] \quad (16)$$

$$SMG = \sum_{w=0}^{Ai} GR [W-ROI[w], X-ROI[w]] \quad (17)$$

$$SU = \sum_{w=0}^{Ai} GHN^2 GR [[W-ROI[w], X-ROI[w]]] \quad (18)$$

$$SSD = \sum_{w=0}^{Ai} (GHN^2 GR [[W-ROI[w], X-ROI[w]]] - SM)^2 \quad (19)$$

$$SS = \frac{1}{1 + SSD^2} \quad (20)$$

$$Sk = \sum_{w=0}^{Ai} (GR [[W-ROI[w], X-ROI[w]]] - SM) \quad (21)$$

$$SE = \frac{\sum_{w=0}^{Ai} (GHN [GR [[W-ROI[w], X-ROI[w]]]]^*)}{\log GP [GR [[W-ROI[w], X-ROI[w]]]]} \quad (22)$$

$$SC = \sum_{w=0}^{Ai} (GHN [GR [[W-ROI[w], X-ROI[w]]]]^* GR [W-ROI[w], X-ROI[w]]) \quad (23)$$

For optimal selection, the feature extraction $f_i^{\text{feat}} = f^{\text{Geometric}} + f^{\text{Texture}} + f^{\text{Gradient}}$ is subsequently fed to the adopted model. $f_i^{\text{feat}^*}$ indicated as the extracted optimal features and it is input to NN.

4. Optimal Feature Selection and Classification

4.1 Objective Model

As stated before, to choose the optimal feature, the adopted method is exploited, which is used to update the weight of NN. To select the optimal features, the integrated feature, f^{feat} (geometric feature, texture, and gradient) is subjected as a solution to the adopted model. f^{feat} indicates the features, which are chosen from the segmented region that $f_i^{\text{feat}} = \{f_1^{\text{feat}}, f_2^{\text{feat}}, \dots, f_n^{\text{feat}}\}$ are optimally chosen using the

adopted model. Subsequently, the optimally chosen features are indicated as $f_i^{\text{feat}^*} = \{f_1^{\text{feat}^*}, f_2^{\text{feat}^*}, \dots, f_n^{\text{feat}^*}\}$.

Moreover, ten features need to be optimally selected from the total of 19 features that are extracted. This paper mainly concentrated on the diagnostic accuracy maximization and it is shown in eq. (24).

$$\text{Obj} = \text{Min} \left(\frac{1}{\text{Accuracy}} \right) \quad (24)$$

4.2 Classification

Using the adopted technique the optimal features are created by exploiting NN, whereas the training approach is replaced by the hybrid WPA-PSO method. At first, the three layers such as input, hidden as well as output layers possess complete NN structure; hidden layer output is decided. In eq. (25), the NN hidden output is exhibited, in NN the hidden neurons are indicated as m , whereas, $m = 1, 2, \dots, N_m$ as well as bias weight of hidden neurons is represented as wh_{bm}^H . The input neuron is signified as y , whereas, $y = 1, 2, \dots, N_y$ and the output neuron is indicated as h , whereas, $h = 1, 2, \dots, N_h$. The wh_{bh}^c represents the output neuron bias weight h^{th} . The number of hidden neurons, output neurons as well as input neurons is indicated as N_m , N_h and N_y , correspondingly. From the y^{th} input neuron to m^{th} hidden neuron the weight is stated as wh_{ym}^H . Moreover, from m^{th} hidden neuron, wh_{mh}^c indicates the weight to h^{th} output neuron. The error function Err indicates a difference among the actual output (Out_{act}) as well as (Out_{pre}) predicted output as well as equivalent numerical formulation as exhibited in Eq. (26).

To distinguish the malignant as well as benign images the NN consequences is exploited. In Eq. (25) and Eq. (26), the activation function is indicated as F .

$$Hi = F \left(wh_{bm}^H + \sum_{y=1}^{N_y} wh_{ym}^H f^{feat*} \right) \quad (25)$$

$$Out_{pre} = F \left(wh_{bh}^c + \sum_{m=1}^{N_m} wh_{mh}^c Hi \right) \quad (26)$$

$$Err = \left\{ \underset{wh_{bm}^H, wh_{ym}^H, wh_{bh}^c, wh_{mh}^c}{\text{argmin}} \sum_{y=1}^{N_y} |Out_{pre} - Out_{act}| \right\} \quad (27)$$

In Eq. (27), $W = wh_{bm}^H, wh_{ym}^H, wh_{bh}^c, wh_{mh}^c$. Here, the developed method is used to tune these weighting parameters optimally.

4.3 Proposed WPA-PSO model

The hybrid WPA-PSO algorithm can attain a better complementary effect. The method used in the solution space is: subsequent to PSO particles [11] search for particles, in the WPA [10] the search process of wolves is exploited to search using rules to decide the last novel location.

In this work, the hybrid technique uses utmost likelihood algorithm to obtain the solution of a and b . The computation formulations of a and b are exhibited as below

$$\begin{cases} a = \frac{n}{1 - e^{-bt}} \\ \frac{n}{b} = at_n e^{-bt_n} \end{cases} \quad (28)$$

In eq. (28), n indicates the identified failure number;

t_i indicates the time of failure (i); The particular model is to substitute the first term in eq. (28) into the second term and perform a mathematical transformation to model a new equation only associated with parameter b , as stated as follows

$$f = b - \frac{n(1 - e^{-bt_n})}{nt_n e^{-bt_n} + (1 - e^{-bt_n}) \sum_{i=1}^n t_i} \quad (29)$$

f indicates new fitness function, as well as parameters in the formulations are recognized except b .

5. Result and Discussion

The breast cancer diagnosis was on basis of a developed model and the outcomes of each analysis were performed. The number of iterations, as well as population size used in this work, is 100 as well as 10, correspondingly. The performance validation of the proposed method was evaluated with the existing techniques regarding the specificity, Accuracy, Precision, False Negative Rate (FNR), False Discovery Rate (FDR), False Positive Rate (FPR), NPV, F1-score as well as Matthews correlation coefficient (MCC).

Table 1 Performance analysis of the proposed model over the conventional models

Measure	GA	PSO	Proposed model
Accuracy	0.38385	0.98988	0.83838
FDR	0.53333	0.21329	0.13285
Specificity	0.11111	0.83333	0.88889
Sensitivity	0.93333	0.93333	0.8
Precision	0.35559	0.98591	0.85913
FPR	0.88889	0.15559	0.11111
FNR	0.055559	0.25559	0.2
NPV	0.11111	0.83333	0.88889
MCC	0.09598	0.59091	0.59305
F1_score	0.52222	0.95852	0.82959

Table 1 summarizes the analysis of the proposed model regarding the feature selection and classification with conventional techniques, by analyzing the performance metrics such as accuracy, precision, specificity, FPR, FNR, NPV, F1-score, FDR, as well as MCC. Here, the proposed model is 22% better than the GA, 25% better than the PSO for accuracy, the proposed model is 18% better than the GA, 22% better than the PSO for precision. Therefore, it is obvious that the adopted technique is better than the conventional techniques, regarding all the performance metrics which are used for breast cancer diagnosis from the analysis.

6. Conclusion

This paper proposes an intellectual breast cancer diagnosis technique by considering preprocessing, segmentation, and feature extraction, and the selection and classification model. Moreover, by exploiting the median filtering as well as the CLAHE model the preprocessing was performed in the presented image. Then, by exploiting the Region growing algorithm, the segmentation process was attained. From the segmented tumor, the geometric, gradient, and texture features were extracted in the feature extraction stage. The feature vector length was large, hence this paper was intended to choose the optimal features using a novel optimization method named the hybrid WPA-PSO method. At last, the optimally chosen features were fed the classification procedure through an NN classifier. Additionally, a novel training approach was exploited whereas the NN weight was updated using the adopted technique therefore utmost accuracy was attained.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

- [1] Mohamed Hisham ArefIbrahim H. AboughalebYasser H. El-Sharkawy,"Custom optical imaging system for ex-vivo breast cancer detection based on spectral signature",Surgical Oncology4 November 2020.
- [2] D. GhiehC. SaadeG. Berjawi,"Staying abreast of imaging – Current status of breast cancer detection in high density breast", Radiography28 June 2020.
- [3] Meha DesaiManan Shah,"An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)", Clinical eHealth24 November 2020.
- [4] K. Park, W. Chen, M. A. Chekmareva, D. J. Foran and J. P. Desai, "Electromechanical Coupling Factor of Breast Tissue as a Biomarker for Breast Cancer," in IEEE Transactions on Biomedical Engineering, volume. 65, number. 1, page no. 96-103, Jan. 2018,

- [5] Ioannis Sechopoulos, Jonas Teuwen, Ritse Mann, "Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art", *Seminars in Cancer Biology* Available online 9 June 2020.
- [6] Iztok Fister, Iztok Fister Jr, Xin-She Yang and Janez Brest, "A comprehensive review of firefly algorithms", *Swarm and Evolutionary Computation*, volume. 13, page no. 34-46, 2013.
- [7] Y. Chang, C. Jung, P. Ke, H. Song and J. Hwang, "Automatic Contrast-Limited Adaptive Histogram Equalization With Dual Gamma Correction," *IEEE Access*, volume. 6, page no. 11782-11792, 2018.
- [8] Youlian Zhu, Cheng Huang, "An Improved Median Filtering Algorithm for Image Noise Reduction", 2012 International Conference on Solid State Devices and Materials Science, *Physics Procedia*, page no. 609-616, 2012.
- [9] J. Sanchez Hernandez, E. Martinez Izquierdo and A. Arquero Hidalgo, "Improving Parameters Selection of a Seeded Region Growing Method for Multiband Image Segmentation," *IEEE Latin America Transactions*, volume. 13, no. 3, page no. 843-849, March 2015.
- [10] X. Gao and P. Wan, "Target recognition of BP network based on Optimized wolf pack algorithm," 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, page no. 2123-2127, 2018.
- [11] N. Rokbani, A. Abraham and A. M. Alimi, "Fuzzy Ant Supervised by PSO and simplified ant supervised PSO applied to TSP," 13th International Conference on Hybrid Intelligent Systems (HIS 2013), Gammarth, Tunisia, page no. 251-255, 2013.
- [12] K. Srinivas, "Prediction of E-Learning Efficiency by Deep Learning in EKhool Online Portal Networks", *Multimedia Research*, Volume 3, Issue 4, October 2020.
- [13] Arvind Madhukar Jagtap, "Developing Deep Neural Network for Learner Performance Prediction in EKhool Online Learning Platform", *Multimedia Research*, Volume 3, Issue 4, October 2020.

