ITC 4/52
Information Technology and Control
Vol. 52 / No. 4 / 2023
pp. 915-934
DOI 10.5755/j01.itc.52.4.34232

Optimal Trained Deep Learning Model for Breast Cancer Segmentation and Classification

Received 2023/05/28 | Accepted after revision 2023/09/01

HOW TO CITE: Krishnakumar, B., Kousalya, K. (2023). Optimal Trained Deep Learning Model for Breast Cancer Segmentation and Classification. *Information Technology and Control*, 52(4), 915-934. https://doi.org/10.5755/j01.itc.52.4.34232

Optimal Trained Deep Learning Model for Breast Cancer Segmentation and Classification

B. Krishnakumar, K. Kousalya
Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Erode, 638060, India

Corresponding author: bkrishnakumarres1@outlook.com

Breast cancer is the most widespread cancer among women. Based on the International cancer research center analysis, the highest number of deaths among women is due to breast cancer. Hence, detecting breast cancer at the earliest may help the oncologist to make appropriate decisions. Due to variations in breast tissue density, there is still a challenge in precise diagnosis and classification. To overcome this challenge, a novel optimal trained deep learning model (OTDEM)-based breast cancer segmentation and classification are proposed with the following four stages: they are, preprocessing, segmentation, feature extraction, and classification. The input image is passed to the initial stage using the Contrast Limited Adaptive Histogram Equalization (CLAHE) filter to enhance the image. Then the preprocessed image is given to the segmentation stage for the image sub-segments by correlation-based deep joint segmentation. Following that, the features such as statistical features, improved local gradient texture pattern (LGXP), texton features, and shape-based features are derived from the segmented image. Then the derived features are fed to the ensemble model that includes a convolutional neural network (CNN), deep belief network (DBN), and bidirectional graph recurrent unit (Bi-GRU) classifier to finalize the classification outcome. Further, to enhance the performance of the ensemble model, the weight of BI-GRU is optimized via a new algorithm termed Swarm Intelligence - Pelican Optimization Algorithm (SIPOA). This ensures optimal training to make the model more appropriate in its classification process. Finally, the performance of the proposed work is validated over the traditional models concerning different performance measures.

KEYWORDS: Deep Learning, Healthcare, Breast Cancer, classification, segmentation.

1. Introduction

Breast cancer is the world's second most prevalent cancer in females. It is the leading cause of cancer mortality in women [25], [31]. Based on World Health Organization's International Agency for Cancer Research (IARC), breast cancer affects around 2.1 million people yearly and causes 7.6 million deaths. It occurred in females between the ages of 40 and 60 [1], [4], [2]. However, improved medical imaging technology allows for the early detection and treatment of more than 90% of breast cancer patients. Early identification of breast cancer, treatment facilities, and proper diagnosis improve life expectancies.

The methods such as Breast Self-Examination (BSE), Fine-Needle Aspiration Biopsy (FNAB), and mammography are utilized to identify breast cancer. The occurrence of swelling, lumps, or any distortions is detected by the BSE method with the help of self-screening confirmation [22], [3]. Mammography predicts early breast cancer occurrence through symptoms like lump, pain, skin dimpling, or nipple discharge. It uses low-intensity X-rays to create images, enabling classification of breast masses as benign or malignant based on factors like size, texture, gradient, and density. Benign tumors have smooth, round, and well-defined boundaries, while malignant masses have rough, hazy, and uncertain boundaries [15].

Automated segmentation and classification of breast masses in mammograms provide precise diagnosis and treatment, overcoming human error and saving time [16]. However, challenges include variations in masses' texture, location, uncertain boundaries, and low contrast, requiring adaptable segmentation approaches for different phenotypic classes. A high-performance classification method is needed to differentiate between benign and malignant masses [24]. Biopsy-based procedures like FNAB utilize retrieved tissue samples to screen for malignant cells and offer a less painful, faster, and noninvasive alternative [14].

This paper proposes an OTDE-based segmentation and classification model for accurate breast cancer detection, addressing the challenges of variations in mass features and similar morphological appearances among classes.

The contribution of this work is as follows:

- Using CLAHE filtering method, the sample input image is preprocessed to filter the image.
- A correlation-based deep joint segmentation is proposed in this work for the segmentation process.
- Improved LGXP is proposed as the new technique for extracting features from the segmented image along with other statistical features, texton features, and shape-based features.
- Proposing an ensemble model, which involves classifiers like CNN, DBN and Bi-GRU, to train the derived features.
- Also, proposing a novel SI-POA for optimal training. Here, the weight of the Bi-GRU classifier is optimized.

The remaining set of the work is as follows:

Section 2 studied the features and challenges of extant work. The proposed model of breast cancer classification model is defined in Section 3. Section 4 describes the proposed ensemble model and the optimization algorithm. The comparison of suggested model results and discussions are analyzed in Section 4 and finally, the conclusion is summarized in Section 5.

2. Related Works

In 2022, Kosmia Loizidou et al. [17] developed a machine-learning method with neural networks to diagnose breast cancer with the effect on patient outcomes. They developed the systems for the progress of automated computer-aided diagnosis systems. Moreover, they obtained accurate detection of masses of about 99.9%, and by using state-of-the-art temporal analysis, the classification precision of suspicious masses or benign increased from 92.6% to 98%.

In 2021, Wang et al. [30] used a framework called Deep Regional Metastases Segmentation for lymph node status classification in patients. They proposed DSNetutilized for the detection of patch-level regional metastases. Then the metastases from individual slides were predicted by density-based spatial clustering of the system with noise and patient-level pN-stages were determined.

In 2020, Yue et al. [32] proposed a multi-task learning framework in 3D Automated Breast Ultrasound (ABUS) images for the segmentation and classification of breast masses. The proposed framework, which con-

tains two sub-networks such as for segmentation, an encoder-decoder network and for classification, a less-weight multi-scale network. Finally, the probability maps from previous iterations were used to obtain refined feature maps using an iterative training strategy.

In 2021, Zhang et al. [34] concluded multi-task learning (SHA-MTL) model based on hard and soft attention mechanisms for Breast Ultrasound (BUS) images. It has an upsampling decoder and a CNN encoder. These were interrelated by attention gated with soft attention technique. Compared to other methods, this model was effective and it requires less knowledge to obtain better outcomes.

In 2021, Khan et al. [11] showed that benign and malignant cells segmentation is performed by a level-set algorithm. Moreover, support vector machine (SVM) based classification is utilized to classify benign and malignant cells. Gray-Level Co-occurrence Matrix (GLCM) was used to collect texture information of breast masses. In classifying benign and malignant cells, the proposed system achieved high precision of about 96.3%.

In 2021, Singh et al. [27] proposed a list of techniques to perform segmentation, such as CLAHE, FCM, Green channel complement, and morphological operations. They obtain results in the form of numerical readings such as specificity, positive and negative predictive rate, sensitivity, accuracy, false-positive value, and false-negative value, etc. Their research provided enhanced results by indicating and rectifying the presence of microcalcifications in the segmentation process.

In 2019, Shen et al. [26] concluded a mixed-supervision guided method and a Residual-aided Classification U-Net model (ResCU-Net) jointly for segmentation and classification of breast masses. Moreover, an annotation procedure has been used to couple strong supervision as a segmentation process and weak leadership used to label benign and malignant cells. The ResCU-Net included a residual model and SegNet architecture to collect information about cellular tissues of breast thus, the identification of cells could be improved.

Table 1
Features and challenges of extant works

Author [Citation]	Methodology	Features	Challenges
Loizidou et al. [17]	– CLAHE	– Achieved high performance	– Need to develop automated CAD systems to treat breast cancer
Wang et al. [30]	– DSNet – DBSCAN	– Low accuracy for ITC	– Need to explore more advanced segmentation framework networks
Yue et al. [32]	– Multi-task learning VNet model.	– Better outcomes from multi-task models	– It is affected by classification bias.
Zhang et al. [34]	– Multi-task learning SHA-MTL model.	– Increased recognition ability of the model.	– Need to perform in large-scale BUS image datasets.
Khan et al. [11]	– SVM-based classifier.	– High accuracy of 96.3%.	– Prediction with limited datasets is complicated.
Singh et al. [27]	– GCC – CLAHE – FCM	– Effective classification of tumor – Seriousness of caused damage.	– The occurrence of micro calcifications makes the reduction process difficult.
Shen et al. [26]	– MSG method – ResCU-Net	– Well performed in dice and achieved high accuracy.	– There is a challenge to perform well at certain specific thresholds.
Khaoula et al. [12]	– End-to-end UNet model	– It worked as one stage technique for both segmentation and classification processes.	– It must have improvements in data augmentation, using transfer learning, cross-validation, increase of number training epochs.

In 2021, Khaoula et al. [12] proposed a model for detecting, segmenting and classifying breast masses in one stage called as end-to-end UNet model. Moreover, the proposed model utilized public datasets like INbreastand DDSM to estimate the performance of segmentation and classification. Thus, the proposed model has weighted F1 of about 99.19% and 99.65% and dice coefficient of about 99.20% and 99.56% for the DDSM and INbreast datasets, respectively.

2.1. Problem Definition

Breast cancer cases continue to rise by about 0.5% annually, with most cases starting in the ducts and varying in appearance and development between patients. Mammography is widely used for early detection, but it faces challenges in certain situations. Research efforts, including Loizidou's experiment [17], have made advancements in automated CAD systems, but some areas still need improvement. By incorporating advanced methods like CLAHE, the performance of automated CAD systems increased by approximately 5%.

Wang [30] proposed a method with high accuracy which was implemented by DSNet to detect metastases regions in patch level though its performance would reach its peak if it was improved with weakly supervised and semi-supervised learning methods. The model experimentedon by Yue [32] limited the model's performance with the occurrence of classification bias. The multi-task learning (SHA-MTL) model proposed by Zhang [34] will have great performance if it is used by large-scale breast ultrasound image datasets. The research work of Khan [11] has higher accuracy of about 96.3% but limited datasets. The proposed model of Singh [27] would be effective if it reduce the presence of micro calcifications. At specific threshold levels, the work by Shen [26] did not give better results but it was best with area under the curve of 94.57% for segmentation and 96.16% for classification. The proposed work by Khaoula [12] using an end-to-end UNet model worked in one level with both classification and segmentation processes inspite of its performance it still have improvements in cross validation, increase in number training epochs, data augmentation and using transfer learning. Thus, the analysis on reviewed works show that the classification and segmentation in breast cancer needs more advancement to improve its accuracy in the detection of breast cancers.

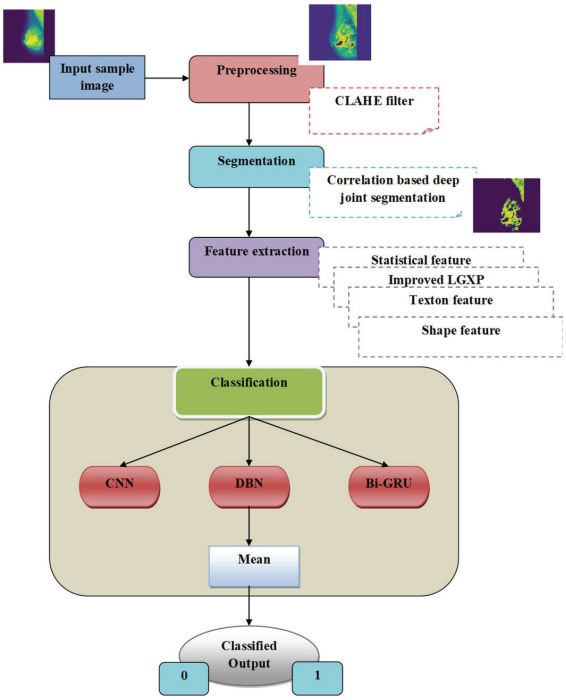
2.2. Major Challenges in Early Stage Prediction

1. In early-stage breast cancer, the tumor size is typically smaller and may have less pronounced visual characteristics compared to advanced-stage cancer. This makes it challenging to detect and extract discriminative features from the images.
2. Early-stage breast cancer may manifest as subtle changes in the breast tissue, such as small microcalcifications or architectural distortions. These features can be harder to detect and analyze accurately compared to the more prominent features seen in advanced-stage cancer.
3. Breast tumors exhibit heterogeneity, which means different regions within the tumor may have varying features. In early-stage cancer, the heterogeneity might be more pronounced, making it challenging to capture and analyze all relevant features.

3. Proposed Segmentation and Classification of Breast Cancer Via Ensemble Model

Breast cancer is the most extensive type of cancer in females. Globally, it affects 2.1 million women each year, and it is the primary cause of cancer-related death among women. If the disease is detected early then the death rate of breast cancer may reduced. The architecture diagram of the proposed OTDE based breast cancer classification model is illustrated in Figure 1. In this paper, an OTDE based breast cancer segmentation and classification model is proposed with the following phases: (1) preprocessing, (2) segmentation, (3) feature extraction, and (4) classification. The input sample image is subjected to the first phase of preprocessing, where CLAHE filter-based image denoising takes place. An improved correlation-based deep joint segmentation method is performed to segment the preprocessed image. Subsequently, the features like statistical features, improved LGXP, Texon based feature and shape-based feature are derived from the segmented image. The models like CNN, DBN, and Bi-GRU are involved in the proposed ensemble model, which utilizes the extracted features as input for classification. For optimal training of Bi-GRU weights, a self-improved pel-

Figure 1
Architecture of the Proposed Method



ican optimization algorithm is presented in this work that ensures an optimal training process.

Preprocessing involves removing unwanted parts from the original image. The input image is preprocessed using the CLAHE filter method to enhance the image with improved contrast without affecting its quality. CLAHE operates on small regions of the image called tiles [9], [13]. The contrast of each tile is adjusted to match the desired pattern for its histogram. Bilinear interpolation is used to connect adjacent tiles, ensuring smooth merging results.

3.2. Segmentation

Image segmentation is the process of dividing a digital image into subsets (image segments) to reduce image complexity and enable further processing. The preprocessed image can be segmented using a proposed correlation-based deep joint segmentation method. This technique measures the distance between the deep points and segmentation points to determine the optimal segments.

3.2.1. Correlation Based Deep Joint Segmentation

The proposed correlation based deep joint segmentation method comprises of three steps: they are joining phase, region fusion phase and segmentation point identification phase [23], [20].

Joining phase:

At first, the preprocessed slide image $I_s^{'}$ is divided into grids then the pixel value is denoted as $I_{si}^{'}$ and the pixels are connected through measuring mean and threshold. Based on the average value of pixels, the mean value is evaluated. This value decides the pixels to be connected when the constant threshold value (i.e. set as 1) processed with the mean value. Then the mean values are evaluated for the grid G that can be formulated as in Equation (1). Here, $I_{si}^{'}$ indicates value of pixels regards to grid, and N refers to pixel count.

$$G = \frac{\sum_{i=1}^{N} I_{si}^{'}}{N}. \tag{1}$$

Then the pixel joining is represented as in Equation (2). Here, t refers to threshold.

$$G = \frac{\sum_{i=1}^{N} I_{si}^{'}}{N} \pm t. \tag{2}$$

Region fusion phase:

Followed by joining phase of the grid, the region fusion phase is processed. According to marked grids, the region fusion matrix is created. Then the region fusion process is subjected to further process by satisfying the following two constraints: (i) Evaluate mean value that should be less than 3and (ii) For individual grid, one grid point is chosen. According to these conditions, the region similarity is first evaluate and fuse the regions to calculate the mapped points. The region similarity can be defined as in Equation (3). Here, $I_{s_j}^{'Y}$ refers to connected pixels in the joining phase and J refers to joined pixels count in grid G.

$$F = \frac{\sum_{j=l}^{n} I_{s_j}^{'Y}}{J}. \tag{3}$$

However, the region are fused together to evaluate the mapped points that can be defined as in Equation (4). Here, z denotes mapped points count in an image.

$$H = \{H_1, H_2, ..., H_z\}. \quad (4)$$

Segmentation point identification phase:

After the region fusion phase, the deep points are analyzed using the missed pixel of joining phase with the mapped points. Moreover, the segment point distance is evaluated inbetween the deep points d_p from the segmented point s_p. Conventionally, the segment point distance is evaluated using Euclidian distance. In this paper, improved correlation segmentation is utilized instead of using Euclidian distance. Pearson correlation is termed as a measure of linear correlation. That is, it measures the direction and strength among the two variables and the value is generated between -1 and 1. The conventional Pearson correlation expression is defined as in Equation (5). Here, md_p refers to mean of d_p values and ms_p refers to mean of s_p values.

$$PC = \frac{\sum (d_p - md_p)(s_p - ms_p)}{\sqrt{\sum (d_p - md_p)^2 \sum (s_p - ms_p)^2}}. \quad (5)$$

The above expressed equation is improved that can be defined as in Equation (6). Here, σ^2 refers to biased weighted sample variance.

$$PC = \frac{\sum (d_p - md_p)(s_p - ms_p)}{\sqrt{\sum (d_p - md_p)^2 \sum (s_p - ms_p)^2}} * \sigma^2. \quad (6)$$

The biased weighted sample variance can be formulated as in Equation (7). Here, I'_{s_i} refers to data value, $\overline{I'_{s_i}}$ refers to mean of data, $\omega_i = \frac{I'_s}{sum(I'_s)}$ and $\omega_t = sum(\omega_i)$.

$$\sigma^2 = \sum_{i=1}^{M} \frac{\omega_i \left(I'_{s_i} - \overline{I'_{s_i}}\right)}{\omega_t}. \quad (7)$$

Thereby, the segmented image can be represented as *Si* after the correlation calculation of Equation (6).

3.3. Feature Extraction: Statistical Feature, Improved LGXP, Texton Feature and Shape Based Feature

Feature extraction is the process of extracting the patterns of an object from the segmented image *Si* The statistical features, improved LGXP, texon feature and shape based feature are extracted from the segmented images *Si*, that are listed below:

3.3.1. Statistical Features

The proposed method involves the statistical features yields considerably non-redundant, understandable, and identifiable characteristics. The statistical features such as mean, min, max, median and standard deviation are evaluated [29].

3.3.1.2. Mean

The mathematical model of mean is formulated as in Equation (8). Here, PC_i indicates noisy segmented image, r denotes row coordinates, c denotes column coordinates, and W refers to window of $m \times n$ size.

$$M_{xy} = \frac{1}{mn} \sum_{(r,c) \in W} PC_i(r,c). \quad (8)$$

3.3.1.3. Min

The min based measure can be formulated as in Equation (9).

$$\tilde{M}_{\min} = \min \{PC_i(r,c) \mid (r,c) \in W\}. \quad (9)$$

3.3.1.4. Max

The max based measure can be formulated as in Equation (10).

$$\tilde{M}_{\max} = \max \{PC_i(r,c) \mid (r,c) \in W\}. \quad (10)$$

3.3.1.5. Median

t is an order-statistic filter that selects the center pixel value from an ordered collection of values within the *mn* neighborhood *W* surrounding the original pixel. Then the median is formulated as in Equation (11).

$$Med = \text{median} \{PC_i(r,c) \mid (r,c) \in W\}. \quad (11)$$

3.3.1.6. Standard Deviation

The mathematical model of standard deviation can be expressed as in Equation (12).

$$\tilde{M}_{SD} = \sqrt{\frac{1}{mn-1} \sum_{(r,c) \in W} \left(PC_i(r,c) - \frac{1}{mn-1} \sum_{(r,c) \in W} PC_i(r,c) \right)^2}. \quad (12)$$

Thereby, the obtained statistical features are represented as $SF = \left[M_{xy}, \tilde{M}_{\min}, \tilde{M}_{\max}, Med, \tilde{M}_{SD} \right]$.

3.3.2. Improved LGXP

Improved LGXP-based features are extracted from the segmented image using a hybrid model of one-dimensional log Gabor and one-dimensional Gaussian smoothing filters. Local phase information is retrieved using band-pass quadrature filters [23], [28]. The selection quadrature filter serves as an improved Gabor filter, allowing simultaneous evaluation of signal amplitude and phase at a given spatial region across various frequencies through filter bank construction and rescaling of the enhanced Gabor filter.

One Dimensional Log Gabor Function: The log gabor function is used to detect the edge of the segmented image. The log gabor filters are implemented in the frequency domain to avoid the complexity. However, there is a limitation in the log function at origin due to varying bandwidth in log gabor. Also, the angular and radial components re combined in polar coordinate system. Thus, the one dimensional log gabor function can be represented as in Equation (13). Here, ω refers to the image ω_0 refers to filters' center frequency and G refers to transfer function.

$$G(\omega,\theta) = \exp\left\{ \frac{\left[\log\left(\omega / \omega_0 \right) \right]^2}{2.\left[\log\left(\sigma_\omega \right) \right]^2} \right\} \cdot \exp\left\{ \frac{\left(\theta / \theta_0 \right)^2}{2\sigma^2} \right\}. \quad (13)$$

One Dimensional Gaussian Smoothing Function: One dimensional Gaussian smoothing function is performed to smooth the edges of the image. Each level is functioned with various degree of smoothing the original image to minimize the noise. The base level of the Gaussian smoothing contains full resolution of segmented image followed by this, the higher level has smoothed model of low level image. The standard deviation Gaussian filter determines the kernel size. The image as per Equation (6) is considered as $f(x,y)$ with the Gaussian kernel can be formulated as in Equation (14).

$$g(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\left(-\frac{x^2+y^2}{2\sigma^2} \right). \quad (14)$$

Then the Gaussian smoothing at level σ can be represented as in Equation (15).

$$G(x,y,\sigma) = f(x,y) * g(x,y,\sigma). \quad (15)$$

Finally, the one dimensional log gabor function and one dimensional Gaussian smoothing function are collectively formulated as in Equation (16). Here, σ refers to standard deviation, x refers to data value, k refers to constant, ω refers to value of data, ω_0 refers to filters' center frequency and σ^2 refers to variance.

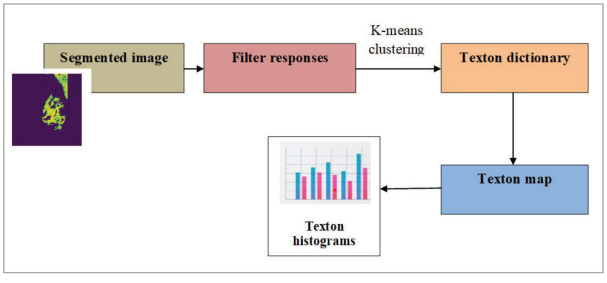
$$G(\omega,\phi,\theta) = \exp\left\{ \frac{\left[\log\left(\omega / \omega_0 \right) \right]^2}{2.\left[\log\left(k / \omega_0 \right) \right]} \right\} \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp \frac{-x^2}{2\sigma^2}. \quad (16)$$

Then the local XOR pattern is applied to the proposed hybrid model of two filter functions. This LXP encodes the filter functions by quantizing the filter functions to the quantized code ranges. Further, the quantized code is applied to the XOR with the related neighbor pixels. These LXP operations are delivered in verilog HDL language using a behavioral approach. Then the features are extracted from the segmented image *Si* and are derived into a vector by the computation of a histogram. Thereby, the features obtained from improved LGXP can be represented as $G(\omega,\phi,\theta)$.

3.3.3. Texton Feature

Textons combine features of existing patterns and textures without comprehensive scrutiny. By utilizing localized filters and clustering, each pixel is mapped to a feature region using k-means clustering, an unsupervised learning method that groups unlabeled inputs. The k-means clustering iteratively determines the optimal value for K centroids and assigns each

Figure 2
Working flow of feature extraction from texton



data point to the closest k-center points, generating a cluster collectively referred to as a texton dictionary (as shown in Figure 2). The resulting features are denoted for further use as T_f.

3.3.4. Shape Feature

Shape features are essential as they characterize an object by capturing its most relevant characteristics while reducing the amount of collected information. The listed shape features include area, orientation, perimeter, eccentricity, and equivalent diameters of the image.

Area: The area of the segmented image region is termed as pixel count in the region. The pixel is plotted within the region is indicated as '1'and the pixel is plotted outside the region is indicated as '0'. Then the mathematical model of evaluating area can be formulated as in Equation (17).

$$A = \sum_{(x,y)\in R} 1. \tag{17}$$

Orientation: It implies the direction of the pixel edge in the image. The orientation-based feature is denoted as O_f.

Perimeter: The perimeter calculation involves three techniques: they are, (1) The outline of the connected region is refined using the thinning technique, (2) the outside of connected region contour length is evaluated by chain code technique, and (3) the outer of connected region contour length is unknown, then its length is evaluated by boundary tracing technique. Then the perimeter based feature is denoted as P_f.

Eccentricity: It is termed as the ratio of the major and minor axis. The two end points of the line segment link to the boundary diameter are known as major axis. The line that is perpendicular to the major axis is called minor axis. Then the eccentricity-based feature is denoted as E_f.

Equivalent diameter: It is termed as the lengthiest distance over the border between the two points. Then the diameter of the region boundary B can be represented as in Equation (18). Here, D indicates any Euclidean distance, b_i and b_j are the boundary points.

$$Dia(B) = \max_{i,j}\left[D\left(b_i, b_j\right)\right] \tag{18}$$

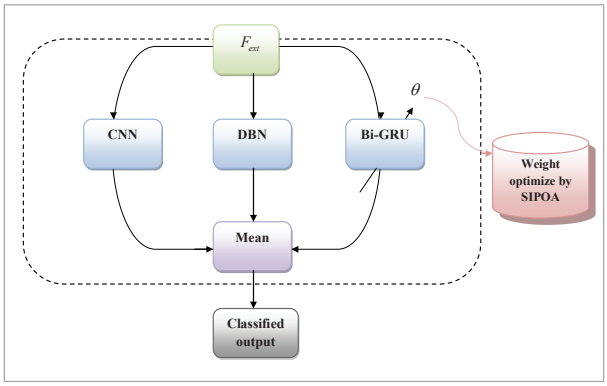
Hence, the obtained features from shape feature are denoted as $s_{bf} = \left[A, O_f, P_f, E_f, Dia(B)\right]$. Thereby, the overall features obtained from statistical feature, improved LGXP, Texton feature and shape based feature are represented as

$$F_{ext} = \left[SF, G(\omega, \phi, \theta), T_f, s_{bf}\right].$$

3.4. Classification Via Ensemble Model

The ensemble model combines multiple classifiers (CNN, DBN, and Bi-GRU) to train extracted features. The classifiers evaluate the mean between actual and predicted values, resulting in high prediction accuracy at low rates. Additionally, a self-improved Pelican optimization algorithm is proposed to optimize the parameters (weight) of Bi-GRU. Figure 3 illustrates the basic structure of the ensemble model, where the outcomes from all three classifiers are averaged to determine the final classification result.

Figure 3
Basic structure of ensemble model



3.4.1. CNN

The feature output is fed into a Convolutional Neural Network (CNN) that reduces the parameter amount. The CNN consists of several layers, including an input layer, alternating convolutional layers, a non-linear layer, a pooling layer, and a fully-connected layer [8]. Each layer contains sub-kernels that gather high-level features, which are then transmitted to the final convolutional layer (fully connected layer). The typical CNN layers include:

Input layer: The input feature result F_{ext}, is often a multidimensional array of data that is exposed to the network.

Convolutional layer: This layer extracts high-level features using learnable kernels applied to the input data.

Non-linear layers: The linear activation is analyzed by non-linear activation functions within this layer that considered input as Con_{lay}, and this layer is also called as the detector stage. The non-linear activation functions such as $\tanh(), sigmoid(), \mathrm{Re}\,LU$ that helps in simulating neuron outcome. As a result, the feature result with non-linear activation rate is represented as NL_{lay}.

The pooling layer: This layer's important goal is to lower the preceding kernel maps' resolution by compressing the features of NL_{lay}, also it reduces the network's computational rate. Furthermore, this layer implies that CNN concentrates on the most well-known patterns since slight alters in before learned features are easier to detect. The convolution of features with the several input maps may combine to each kernel map's output. Thus, the convolutional layer's output map can be stated as in Equation (19). Here, l indicates convolutional layer, $l-1$ indicates pooling layer, K_{ij}^{l} indicates kernel map of l convolutionla layer, b_j^l indicates additive output bias of l layer, $\left(NL_{lay}\right)_j^{l-1}$ indicates input feature of $l-1$ layer, and Q_j indicates input map selection.

$$C_{scr\,j}^{\ l} = f\left(\sum_{i\in Q_j}\left(NL_{lay}\right)_j^{l-1} * K_{ij}^{l} + b_j^{l}\right). \tag{19}$$

Following that, the input feature map, kernel, and convolutional output, and each kernel map in the convolutional layer has the similar weight like filters. By distributing the weights, the parameter amount is reduced, which improves in detecting the same characteristics for all inputs. Finally, the max or average pooling algorithm is utilized by the layers' outcome and this can be serve as input in the next convolutional layer.

Fully connected layer: The convolution output $C_{scr\,j}^{\ l}$ is fed into this layer, which is made up of a generic multi-lateral network. In the preceding layer, some of the final layers are completely connect the activation functions, which produces the characteristics used to train the classifier. As a consequence, the score (intermediate result) from this network is expressed as C_{scr}.

3.4.2. DBN

The feature undergoes generative probabilistic technique called Deep Belief Networks (DBN), designed with multiple layers of RBMs. Each RBM layer involves visible and hidden units. Greedy layer-wise training is used to generate the deep layered representation of . Once an RBM layer is trained, its representation is passed to the next hidden layer. The probability of two hidden layers in DBN is defined as in Equation (20), where represents the hidden layer. The complete equation in Equation (20) is not provided.

d are weight matrices and $q+1$ bias vectors. That is, weight matrix as $w=(w_0,.....w_q)$ and bias vectors as $\left(b_0,.....b_q\right) with\, b_0$ for visible layer.

$$\begin{aligned} &P\left(h_q, h_{q-1}\right) \propto \exp\left(b_q{}^T h_q + b_{q-1}{}^T h_{q-1} + h_{q-1}{}^T w_q h_q\right) \\ &P\left(h_{k_i}=1 \mid h_{k+1}\right) = \sigma\left(b_{q_i} + w_{q+1_{:j}}{}^T h_{q+1}\right)\ \forall i, \forall k \in 1,2,...,q-2 \end{aligned} \tag{20}$$

The DBN network is made up of layered with two RBMs, in which one is visible network v and another is two hidden network h, where $b_{(1)}\, and\, b_{(2)}$ indicates the hidden layers' bias vector, $b_{(0)}$ indicated the visible layers' bias vector, v indicates the visible layers' state vector, $w_{(1)}\, and\, w_{(2)}$ indicates the equivalent weight matrix and $h_{(1)}\, and\, h_{(2)}$ indicates the hidden layers' state vectors. Further, the probability of RBN related distribution can be defined as in Equation (21). Here, the value of

$$\begin{aligned} &\sigma\left(F_{ext}\right) = {}^{1}\!\Big/\!\left(1+\exp\left(-F_{ext}\right)\right)\cdot \\ &P\left(v_{\ i}=1 \mid h_{(1)}\right) = \sigma\left(b_{(0)i} + w_{(1)_{:i}}^{T} h_{(1)}\right) \qquad \forall i. \end{aligned} \tag{21}$$

If the visible unit is real value then utilize the subsequent expression that can be defined as in Equation (22). Here, γ indicates tractability diagonal.

$$v \sim N\left(b_{(0)} + w_{(1)}^{T} h_{(1)}, \gamma^{-1}\right). \tag{22}$$

Thus, the DBN implied hidden layer can be defined as in Equation (23).

$$\begin{aligned} &h_{(1)} = \sigma\left(b_{(1)} + v^T w_{(1)}\right) \\ &h_{(q)} = \sigma\left(b_{(q)_i} + h_{q-1}{}^T w_{(q)}\right), \forall q \in 2,....,m \end{aligned} \tag{23}$$

Thereby, the scores (intermediate result) obtained from the DBN are denoted as D_{scr}.

3.4.3. Bi-GRU

In artificial RNN, gated recurrent units are referred to as a gating mechanism equivalent to that of LSTM [19]. The GRU, on the other hand, decreases the amount of parameters in order to obtain better presentation on slighter to average-sized datasets. The GRU's hidden state is estimated as in Equation (24). Here, u indicates update gate, $\otimes$ indicates element based multiplication and m_t indicates candidate gate.

$$h_t = u \otimes m_t + (1-u) \otimes h_{t-1}. \quad (24)$$

The update gate u is restructured by considering sigmoid function σ that can be formulated as in Equation (25), where θ indicates weight of the update gate, which is optimized by suggested Si-POA as per Equation (29) and b_u indicates bias of the update gate.

$$u_t = \sigma\left(\theta_{uh} h_{t-1} + \theta_{ux} F_{ext} + b_u\right). \quad (25)$$

The candidate gate m_t can be restructured by employing the hyperbolic tangent that is considered as in Equation (26). Here, r denotes reset gate.

$$m_t = \tanh\left(\theta_{hscr} F_{extt} + \theta_{hh}\left(r \otimes h_{t-1}\right) + b_h\right). \quad (26)$$

The reset gate of the GRU can be expressed as in Equation (27).

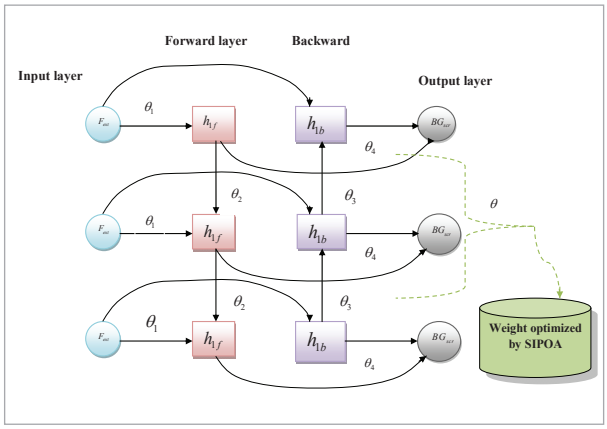
$$r_t = \sigma\left(\theta_{rh} h_{t-1} + \theta_{rF_{ext}} F_{extt} + b_r\right). \quad (27)$$

To address the vanishing gradient problem, a bidirectional GRU (Bi-GRU) is employed, involving only the input gate and a forget gate. This method extracts information from both past and future time steps, enabling precise diagnosis of the present condition. Bi-GRU consists of two cells, one representing the forward input pattern and the other representing the backward input pattern. Figure 4 illustrates the workflow of the Bi-GRU model. The classification result obtained from Bi-GRU is denoted as follows.

Algorithm 1

Initialize the CNN model architecture
Step 1: cnn_model = initialize_cnn_model()
Step 2: cnn_model.compile (loss='categorical_crossentropy', optimizer='Adam', metrics=['accuracy']))

Figure 4
Work flow of the Bi-GRU model



Step 3: cnn_model.fit(X_train, y_train, batch_size=32, epochs=10, validation_data=(X_val, y_val))
Step 4: predictions = cnn_model.predict(X_test)
Step 5: ensemble_predictions.append(predictions)
Initialize the DBN model architecture
Step 6: dbn_model = initialize_dbn_model()
Step 7: pretrain_dbn_model(dbn_model, X_train)
Step 8: fine_tune_dbn_model(dbn_model, X_train, y_train)
Step 9: predictions = dbn_model.predict(X_test)
Step 10: ensemble_predictions.append (predictions)
Initialize the Bi-GRU model architecture
Step 11: bi_gru_model = initialize_bi_gru_model()
Step 12: bi_gru_model.compile (loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
Step 13: bi_gru_model.fit(X_train, y_train, batch_size=32, epochs=10, validation_data=(X_val, y_val))
Step 14: predictions = bi_gru_model.predict(X_test)
Add the Bi-GRU predictions to the ensemble model's list
Step 15: ensemble_predictions.append(predictions)

4. Proposed SIPOA to Optimize the Weight of Bi-GRU

4.1. Input Solution to SIPOA and Objective Function

The weights of Bi-GRU $\theta = \left\{\theta_1, \theta_2, ..., \theta_n\right\}$ are optimally tuned as per the proposed optimization algorithm SI-POA. The weight θ is given as input solu-

tion to the SIPOA. Then the objective function of the proposed model determines minimization of error and that can be formulated as in Equation (28).

$$Fitness = \frac{1}{Accuracy}. \tag{28}$$

4.2. Mathematical Model of Proposed SIPOA

POA is an intelligent process, in which it is designed based on the hunting behavior of pelican [23]. Using this algorithm, the weights are balanced between the exploration and exploitation phase to obtain an optimal solution still it is moderately doing well in speed. Therefore, this algorithm is improved by proposing a novel SIPOA. The flow chart of SIPOA is given in the below Figure 5. This algorithm is termed as population-based algorithm and the pelicans are considered as the population members. The candidate solution means individual population member that provides values for the variables related to the optimization issue based on the position of the candidate solution in the search region. The problem based on lower bound and upper bound are initially randomized the population members is given in following Equation (29). Here, $x_{i,j}$ indicates the i-th candidate solutions' j-th variable value, N indicates the population member count, m indicates population variables count, u_j refers to upper bound of j-th variable, l_j refers to lower bound j-th variable, and *rand* refers to random number between [0, 1].

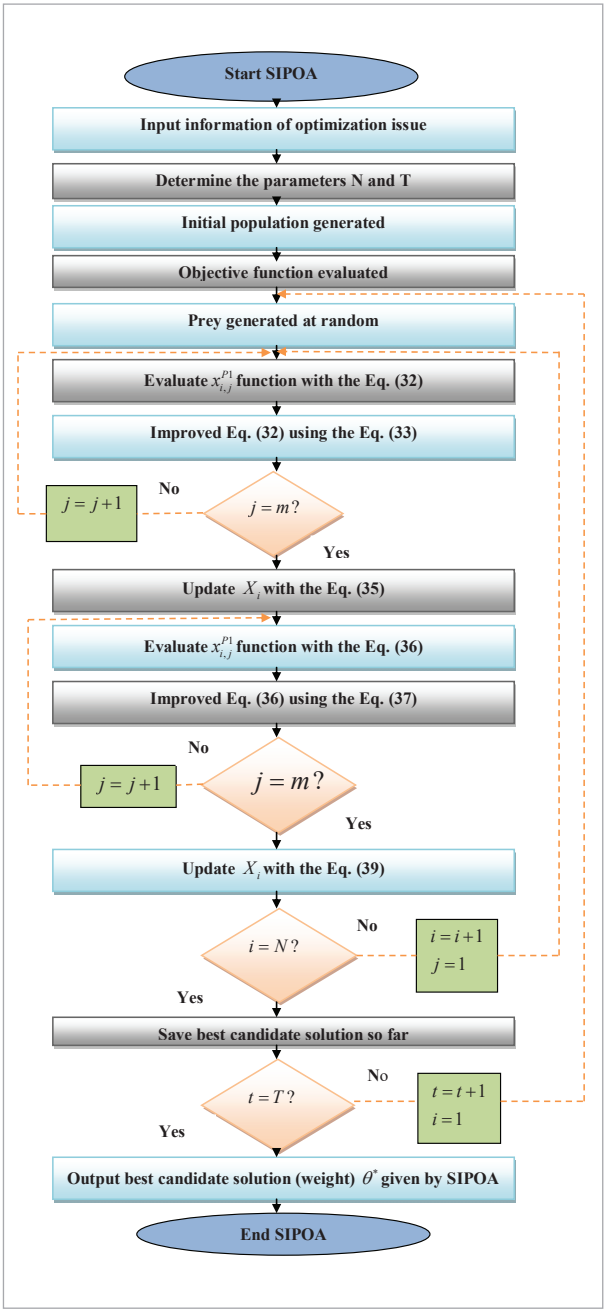
$$x_{i,j} = l_j + rand.\left(u_j - l_j\right), \quad i = 1,2,.....,N; \quad j = 1,2,....,m \tag{29}$$

The weights of the BI-GRU, θ are pelican's population members $X = \left\{\theta_1, \theta_2, ..., \theta_{N,m}\right\}$ that are formulated as population matrix in Equation (30).

$$X = \begin{bmatrix} X_1 \\ : \\ X_i \\ : \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,i} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ : & \ddots & : & \iddots & : \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ : & \iddots & : & \ddots & : \\ c_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N\times m} \tag{30}$$

Each population member is considered as the candidate solution in the proposed model so that the problem given by the objective solution can be calculated according to individual candidate solutions. Therefore, the vectors used in the objective function values are referred as objective function vector that can be formulated as in Equation (31). Here, F indicates the objective function vector.

Figure 5
Flow chart of SIPOA



$$F=\begin{bmatrix}F_1\\:\\F_i\\:\\F_N\end{bmatrix}_{N\times 1}=\begin{bmatrix}F(X_1)\\:\\F(X_i)\\:\\F(X_N)\end{bmatrix}_{N\times 1} \quad (31)$$

The hunting strategy of the pelican is estimated with the following two steps: they are, exploration phase and exploitation phase. The strategy in exploration phase is moving towards prey and the strategy in exploitation phase winging on the water surface. Prey location can be formulated as in Equation (32). Here, $x_{i,j}^{P1}$ indicates pelicans' new position, I indicates the random number that is equal to 1 or 2, p_j indicates the prey location and F_p indicates the value of objective function.

$$x_{i,j}^{P1}=\begin{cases}x_{i,j}+rand.\left(p_j-I.x_{i,j}\right), & F_p<F_i;\\ x_{i,j}+rand.\left(x_{i,j}-p_j\right), & otherwise,\end{cases} \quad (32)$$

Then the Equation (32) is improved using skew tent map [7] that can be defined as in Equation (33). Here, p is the parameter belongs to [0, 1] interval.

$$x_{n+1}=\begin{cases}\dfrac{x_n}{p} & ,0<x_n\\ \dfrac{1-x_n}{1-p}, & else\end{cases} \quad (34)$$

Moreover, the pelicans' new position is recognized based on the objective function value. If the objective function value is updated in that area then the new position is accepted otherwise it is not recognized. This form of updating is termed as effective updating and the process is expressed as in Equation (35). Here, X_i^{P1} indicates pelicans' new status, F_i^{P1} indicates the value of objective function for phase 2.

$$X_i=\begin{cases}X_i^{P1} & ,F_i^{P1}<F_i:\\ X_i & ,else,\end{cases}. \quad (35)$$

Exploitation phase: This phase describes the pelicans winging on the water surface. Once the pelican attain the water surface, the pelicans' wings spread over the water surface to shift the fish upwards and gather the prey in the pelicans' throat pouch. Therefore, more fishes are hunted or caught by the pelicans.

This type of strategy can be modeled as in Equation (36). Here, $x_{i,j}^{P2}$ indicates pelicans' new status based on phase 2, R indicates constant that is equal to 0.2, $R.\left(1-\dfrac{t}{T}\right)$ indicates the $x_{i,j}$'s neighborhood radius, t indicates iteration counter and T denotes the maximum iteration count.

$$x_{i,j}^{P2}=x_{i,j}+R.\left(1-\frac{t}{T}\right).\left(2.rand-1\right).x_{i,j}. \quad (36)$$

The Equation (36) is improved by applying the omega [23] to the sine function that can be defined as in Equation (37).

$$x_{i,j}^{P2}=x_{i,j}+R.\left(\sin\left(\frac{1}{2}\left(\frac{T-t}{T}\right)\right)\right).\left(2.rand-1\right).x_{i,j}. \quad (37)$$

Here, ω (omega) value is evaluated using the following Equation (38), where, t indicates iteration counter and T denotes the maximum iteration count.

$$\omega=3.3*\frac{t}{\left(T-1\right)}. \quad (38)$$

Similarly, the new pelicans' position is recognized or rejected based on effective updating as per phase 1(Exploration phase) that can be defined as in Equation (39).

$$X_i=\begin{cases}X_i^{P2} & ,F_i^{P2}<F_i:\\ X_i & ,else,\end{cases} \quad (39)$$

5. Result Analysis on OTDEM Model for Breast Cancer Segmentation and Classification

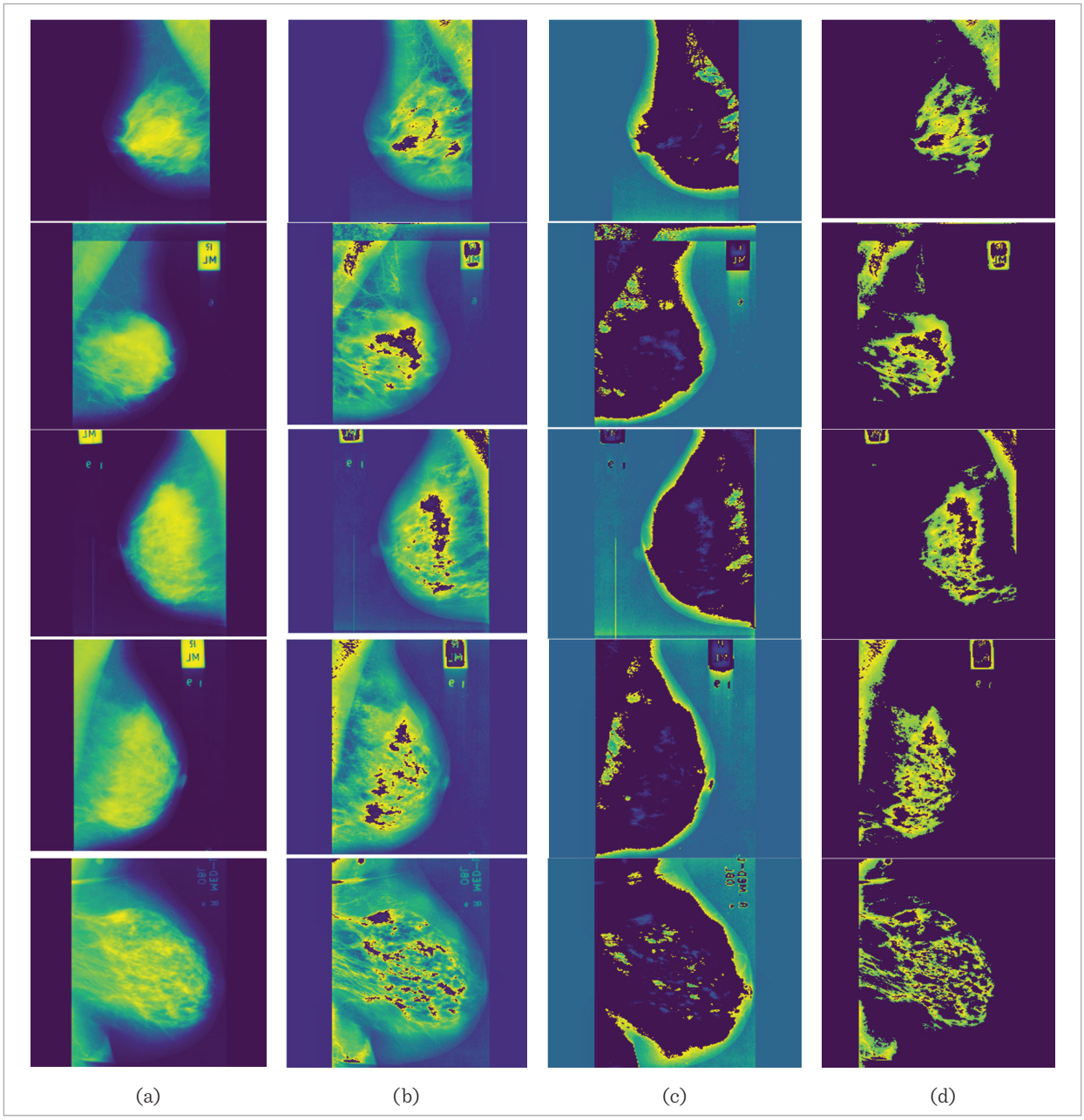
5.1. Dataset and Experimentation

CAMELYON 16 and 17 [30] dataset is used. In this dataset, the total numbers of images used are 322. In addition, the training and testing images for the 60% of learning percentage is 193 and 129, for the 70% of learning percentage is 225 and 96, for the 70% of learning percentage is 258 and 64 and finally for the 90% of learning percentage is 290 and 32, respectively. Moreover, the size of all the images is 64 pixels x

64 pixels. Here, we had used two different classes such as 0 and 1, where 0 represents Non-cancer and 1 represents Cancer cells. The proposed Breast Cancer Segmentation and Classification was implemented in PYTHON. The dataset was accumulated in [30]. The OTDEM was contrasted against the Pelican, COOT, CHIMP based optimization, HGS and BES respectively. In order to evaluate the effectiveness of the OTDEM based on breast cancer segmentation, the examination was carried out in terms of sensitivity, accuracy, FNR, NPV and other measures while adjusting the learning percentage. Additionally, the images for breast cancer segmentation were shown in Figure 6.

Figure 6

Images for Breast Cancer Segmentation a) Original Images b) Preprocessed Images c) Conventional Segmentation Images and d) Improved Segmentation images



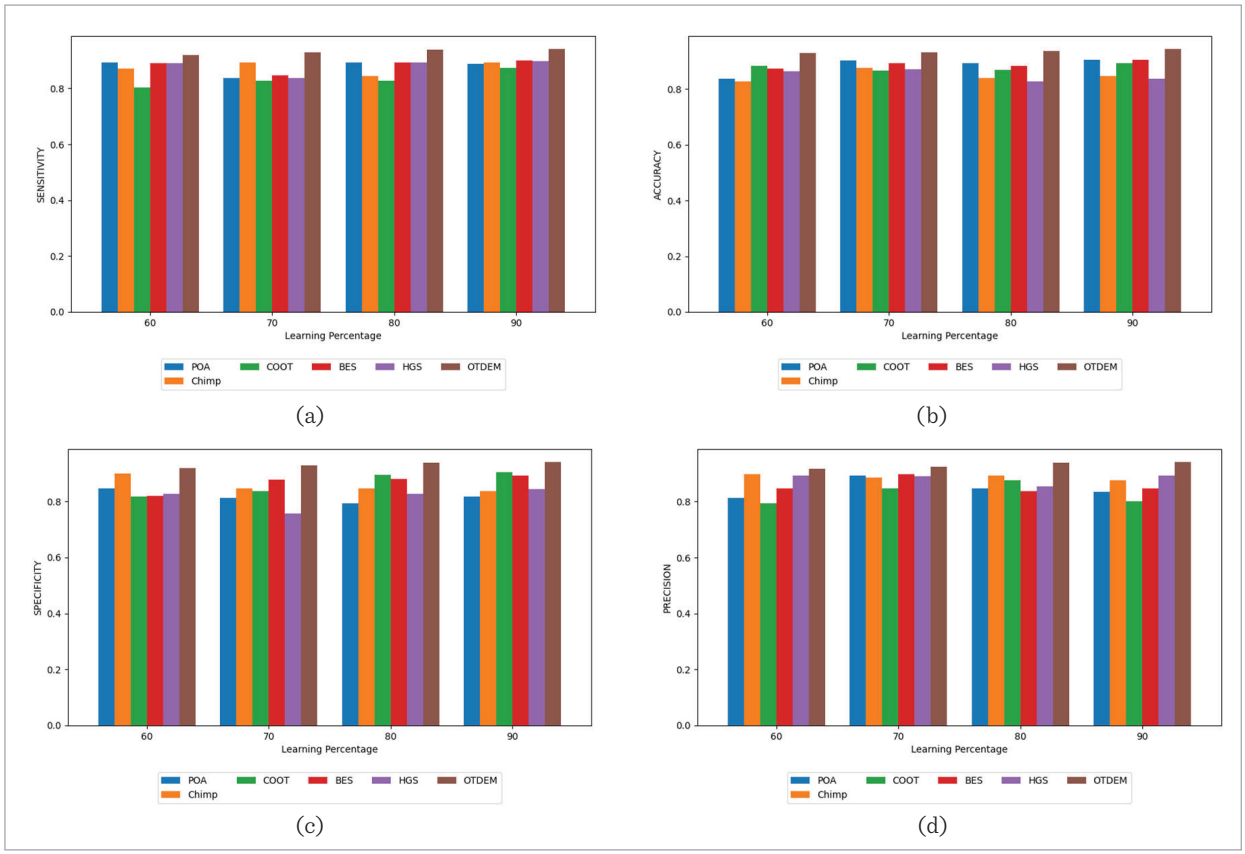
5.2. Assessment on OTDEM and the Existing Approaches Regarding Positive Measure for Breast Cancer Classification

The OTDEM model's positive measure evaluation was done and it was analyzed and examined over the Pelican, BES, CHIMP optimization, HGS and COOT based optimization for breast cancer classification is represented in Figure 7. The assessment was conducted with regard to sensitivity, precision, accuracy and specificity by modifying the learning percentage. In this case, a high positive measure value is required for an accurate classification of breast cancer. In a similar manner, the OTDEM accomplished excellent values for breast cancer classification. More particular, the OTDEM generated the higher accuracy value for the 70th learning percentage is 93.25%, though the CHIMP optimization, POA, BES, COOT Optimization and HGS gained very low accuracy rate of 90.38%, 87.64%, 86.56%, 89.37% and 87.04%, respectively. Moreover, the acceptable sensitivity value is acquired by the BES (90.03%) and the COOT as well as the HGS algorithm (89.78% of sensitivity) at the 90th learning percentage. Nonetheless, the OTDEM maintained highest sensitivity value of 94.01% for breast cancer classification.

In addition, the precision of OTDEM attain increased performance for breast cancer classification in the 60th learning percentage than other existing models. Moreover, the algorithms like POA, HGS, COOT, BESand CHIMP recorded the lowest specificity for the 80th learning percentage is 79.29%, 84.75%, 89.37%, 87.92% and 82.67%, even though the OTDEM scored the highest specificity of 93.75%. Thus, the betterment of the OTDEM is demonstrated for classifying the breast cancer utilizing ensemble classifier (CNN, Bi-GRU and DBN) as well as optimization methodology.

Figure 7

Performance estimation a) Sensitivity b) Accuracy c) Specificity and d) Precision on OTDEM over the conventional methodologies for breast cancer classification

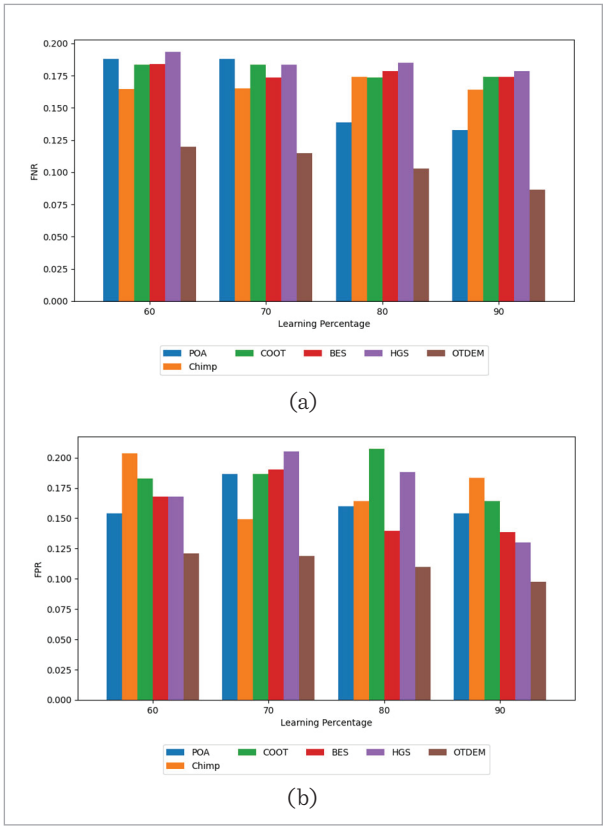


5.3. Assessment on OTDEM and the Existing Approaches Regarding Negative Measure for Breast Cancer Classification

The OTDEM is examined over the conventional approaches such as POA, CHIMP, COOT, BES and HGS, correspondingly based on negative measures. When analyzing Figure 8(a)-(b), the OTDEM resulted with lower negative measure values for breast cancer classification. The OTDEM offered the lowest FPR with superior outcomes for classifying the breast cancer at 90% of learning percentage, than the 60th learning percentage. Mainly, the OTDEM generated 0.102 as FNR for 80th percentage, whereas the standard schemes acquired the minimal FNR, namely, POA=0.138, CHIMP=0.173, COOT=0.174, BES=0.178 and HGS=0.184, respectively. Furthermore, the OTDEM scored the minimal FPR for classification of breast cancer while setting the training percentage as 90% than other models like CHIMP optimization, COOT, HGS, BES and Pelican, respectively.

Figure 8

Performance estimation a) FNR and b) FPR on OTDEM over the conventional methodologies for breast cancer classification



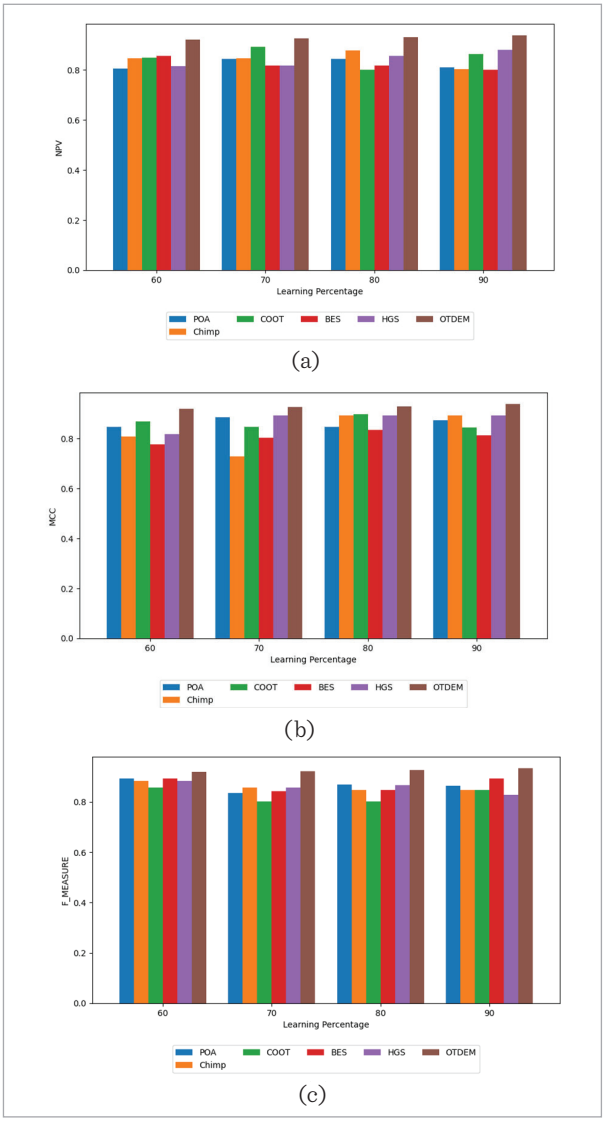
Thus, the OTDEM has acquired lower negative measures with enhanced performance for breast cancer classification with Modified LGXP based feature extraction and optimization concept.

5.4. Assessment on OTDEM and the Existing Approaches Regarding Other Measure for Breast Cancer Classification

The overview of other measure result estimation is shown in Figure 9, to verify the accurate classifica-

Figure 9

Other measure estimation a) NPV b) MCC and c) F-measure on OTDEM over the conventional methodologies for breast cancer classification



tion of breast cancer for the OTDEM. The OTDEM is contradistinction with the previous procedures. Mainly, the NPV of the OTDEM for the classification of breast cancer is 93.76% at the 90th learning percentage, although the POA=81%, CHIMP=86.37%, COOT=80.17%, and HGS=88.03%, respectively. Moreover, the F-measure of OTDEM holds highest value for breast cancer classification when the training percentage is adjusted to 70, which is significantly higher than POA, CHIMP, COOT, BES and HGS, respectively. While the learning percent is set to as 90%, the OTDEM achieved maximal MCC for breast cancer classification than at learning percent 70%; whatever, the established methods holds the lower MCC score. Consequently, the OTDEM significantly improved for all other metrics values for classifying the breast cancer using improved features with the ensemble classifiers (DBN, Bi-GRU and CNN).

5.5. Impact of OTDEM, with and Without Optimization as Well as the Conventional Features for Breast Cancer Classification

The impact on OTDEM, with and without optimization along with the conventional features for breast cancer classification with regard to varied performance measure is illustrated in Table 2. The OTDEM with optimization and Improved LGXP can make the system optimal for breast cancer classification. Regarding the F-measure, the OTDEM has obtained the value of 92.08%, model without optimization is 85.48% and model without improved LGXP is 85.42%, respectively. Similarly, the OTDEM, model without optimization and model without improved LGXP have acquired the accuracy of 93.25%, 84.76% and 85.42%, respectively. In addition, the sensitivity, FPR, MCC and specificity possess the values of 92.89%, 0.1187, 92.77% and 92.94%, respectively. This implies that the OTDEM perform better classification for breast cancer employing ensemble classifiers (Bi-GRU, CNN and DBN) and the optimization strategy.

5.6. Performance Comparison on Ensemble+OTDEM and the Established Classifiers for Breast Cancer Classification

The comparison between the Ensemble+OTDEM is computed to the DNN [10], DCNN [5], SVM, MLP, RF, RNN and ANN for breast cancer classification was depicted in Table 3. While reviewing the results of Table 3, it is worthwhile to observe that the Ensemble+OTDEM accomplished superior findings for classifying the breast cancer than the traditional ways. Especially, FNR of the Ensemble+OTDEM is 0.1149, whereas the DNN is 0.183, DCNN is 0.130, SVM is 0.152, MLP is 0.189, RF is 0.164, RNN is 0.198 and ANN is 0.152, respectively. For the efficacious classification of brain tumor, the Ensemble+OTDEMgained the maximal accuracy than the DNN, DCNN, SVM, MLP, RF, RNN and ANN, respectively. Simultaneously, the precision,

Table 2
Ablation assessment on OTDEM, model without Improved LGXP and model without optimization for breast cancer classification

	Model without optimization	Model without Improved LGXP	OTDEM
Accuracy	0.847673	0.854271	0.932519
MCC	0.846257	0.847657	0.927766
Precision	0.883675	0.897653	0.923868
FPR	0.168793	0.205607	0.118766
F-measure	0.854868	0.854271	0.920782
Specificity	0.817673	0.847657	0.929388
NPV	0.817573	0.893886	0.927575
Sensitivity	0.892763	0.827734	0.928889
FNR	0.198386	0.187374	0.114879

Table 3
Assessment on Ensemble+OTDEM classifier over the traditional classifiers for breast cancer classification

	DNN	DCNN	SVM	MLP	RF	RNN	ANN	Ensemble+OTDEM
Accuracy	0.8241	0.8040	0.7839	0.8643	0.8643	0.8072	0.8595	0.9325
MCC	0.8288	0.8728	0.8927	0.8538	0.8733	0.8477	0.8477	0.9278
Specificity	0.7664	0.8937	0.8176	0.8837	0.8037	0.8271	0.8645	0.9294
F-measure	0.8241	0.8040	0.7839	0.8643	0.8643	0.8298	0.8927	0.9208
NPV	0.8913	0.8696	0.8478	0.8177	0.8348	0.8894	0.9296	0.9276
Precision	0.8928	0.7477	0.8347	0.8037	0.8968	0.8477	0.8277	0.9239
FPR	0.1929	0.1879	0.1999	0.1963	0.1963	0.1729	0.1355	0.1188
Sensitivity	0.8913	0.8696	0.8478	0.8376	0.8176	0.7609	0.8478	0.9289
FNR	0.1838	0.1304	0.1522	0.1887	0.1637	0.1989	0.1522	0.1149

NPV and specificity recorded by the Ensemble+OTDEM is 92.39%, 92.76% and 92.94%, respectively. The improved classification accuracy for breast cancer using the Modified LGXP feature and optimization approach of the Ensemble+OTDEM has demonstrated its superiority.

5.7. Analysis on Statistical Measure of OTDEM Approach and the Conventional Methods with Respect to Error for Breast Cancer Classification

Table 4 illustrates the statistical study on OTDEM scheme is compared to the standard methods like POA, COOT, BWO, BES and HGS using five different statistical measure for breast cancer classification. The outcomes are typically examined using statistical analysis because the optimization process is unpredictable. For accurate classification of breast cancer, the model should accomplish lower error value. Likewise, the OTDEM acquired the minimal error value in most of the statistical measure. More particularly, the OTDEM gained the lowest error value is 1.084 in the minimum statistical measure, this is superior to POA=1.154, COOT=1.137, BWO=1.149, BES=1.179 and HGS=1.184, respectively. Furthermore, analyzing the mean statistical measure, the OTDEM recorded the lowest error value of 1.183, whereas the POA, COOT, BWO, BES and HGS scored the minimum error value of 1.244, 1.258, 1.226, 1.267 and 1.293, respectively. Additionally, in a number of statistical measures, the OTDEM performed better than the earlier methods for breast cancer classification [18, 33].

5.8. Segmentation Evaluation on Correlation Based Deep Joint Segmentation and the Extant Segmentation Approaches for Breast Cancer Segmentation with Regard to Dice and Jaccard Coefficient

Moreover, the Jaccard value of the Correlation based Deep Joint segmentation for the breast cancer segmentation is 0.921, though the Deep Joint Segmentation, FCM and K-means segmentation acquired minimized Jaccard values.

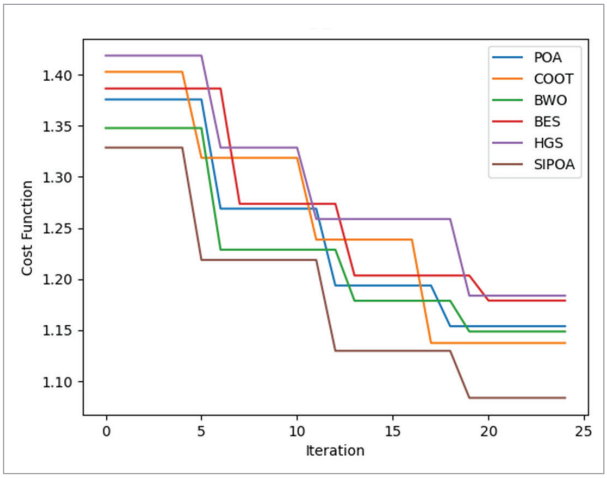
5.9. Convergence Study on SIPOA and the Traditional Methods for Breast Cancer Classification

The Convergence graph analysis on SIPOA is compared with the POA, COOT, BWO, BES and HGS for breast cancer classification is portrayed in Figure 10. Additionally, it is analyzed by altering the iteration (0-25). For improved breast cancer classification, the model must have the lowest error value and the fastest convergence. In the first iteration, the SIPOA and the other approaches obtained greater error rate while as the iteration progressed, the error value get declined for almost all the algorithms. For instance, the SIPOA scored the lowest error value of 1.10, though the POA, COOT, BWO and BES have obtained the highest error rate of 1.18, 1.14, 1.16, 1.19 and 1.20, respectively. Thus,

Table 4
Statistical estimation on OTDEM versus the conventional methods with regard to error for breast cancer classification using the different case scenarios

	POA	COOT	BWO	BES	HGS	OTDEM
Mean	1.24429	1.2583	1.22611	1.26666	1.29313	1.18342
Median	1.19368	1.23865	1.22877	1.27366	1.25878	1.12978
Standard Deviation	0.084837	0.099062	0.074366	0.08177	0.085345	0.08808
Minimum	1.15387	1.13748	1.14869	1.17899	1.18377	1.08377
Maximum	1.37575	1.40277	1.34777	1.38647	1.41868	1.32867

Figure 10
Convergence evaluation on SIPOA versus conventional approaches



the SIPOA acquired the minimal error value in almost all the iterations and it is fast converged over the traditional methodologies as well as the SIPOA has the capacity to classify the breast cancer accurately.

6. Conclusion

The proposed model of OTDE comprises four essential steps, namely preprocessing, segmentation, feature extraction, and classification, for the effective segmentation and classification of breast cancer. The sample image is preprocessed using the CLAHE filtering method to enhance its visibility and quality. The preprocessed image is then segmented using the correlation-based deep joint segmentation method to accurately delineate the regions of interest. Various features, including statistical features, improved LGXP, texton-based features, and shape-based features, are extracted from the segmented image. Subsequently, these extracted features are utilized to train the proposed ensemble model, which combines the strengths of CNN, DBN, and Bi-GRU classifiers. The parameter (weight) of the Bi-GRU classifier is tuned using the proposed SIPOA, further optimizing the performance of the model. The ensemble model, incorporating multiple classifiers, leverages the advantages of each individual classifier to achieve improved accuracy and robustness in breast cancer classification.

In the evaluation of the Ensemble+OTDEM model, notable performance metrics were recorded. The precision, negative predictive value (NPV), and specificity of the model were found to be 92.39%, 92.76%, and 92.94%, respectively. These results demonstrate the effectiveness of the proposed model in accurately identifying and classifying breast cancer. For future directions, further research can be conducted to explore the integration of additional advanced image processing techniques or deep learning architectures to enhance the performance of the segmentation and classification stages. Additionally, expanding the dataset and conducting extensive clinical validations can strengthen the generalizability and reliability of the proposed model. Furthermore, exploring the potential of incorporating multimodal information, such as clinical data or genetic markers, could provide a more comprehensive and accurate assessment of breast cancer.

The proposed advantage of the ensemble model is its ability to leverage the complementary strengths of multiple classifiers. By combining CNN, DBN, and Bi-GRU, the ensemble model benefits from their diverse approaches to feature extraction and classification. This synergistic integration enhances the overall performance, increasing the accuracy, precision, and specificity of the breast cancer diagnosis. Moreover, by employing the SIPOA for parameter tuning, the model achieves further optimization, ensuring optimal training and improved results.

References

1. Agaba Ameh, J., Mohammed, A., Sahalu, B. J., Hayatu, H. I., Haruna, C. Improved Multi-Classification of Breast Cancer Histopathological Images using Handcrafted Features and Deep Neural Network Dense Layer. Intelligent Systems with Applications, 2022, 14. https://doi.org/10.1016/j.iswa.2022.200066
2. Ahmed, I., Muhammad, S. MDA-Net: Multiscale Dual Attention-based Network for Breast Lesion Segmentation Using Ultrasound Images. Journal of King Saud University Computer and Information Sciences, 2022, 34, 7783-7299. https://doi.org/10.1016/j.jksuci.2021.10.002
3. Ashraf, A. B., Gavenonis, S. C., Daye, D., Mies, C., Rosen, M. A., Kontos, D. A. Multichannel Markov Random Field Framework for Tumor Segmentation with an Application to Classification of Gene Expression-Based Breast Cancer Recurrence Risk. IEEE Transactions on Medical Imaging, 2013, 32(4), 637-648. https://doi.org/10.1109/TMI.2012.2219589
4. Chu, H., Jiatai, L., Jinhai, M., Yi, W., Qingling, Z., Bingchao, Z., Xin, C., Xipeng, P., Zhenwei, S., Zeyan, X., Su, Y., Lixu, Y., Huan, L., Xiaomei, H., Changhong, L., Guoqian, H., Zaiyi, L. Multi-Layer Pseudo-Supervision for Histopathology Tissue Semantic Segmentation using Patch-Level Classification Labels. Medical Image Analysis, 2022, 80. https://doi.org/10.1016/j.media.2022.102487
5. Dina, A., Ragab, O. A., Maha, S., Jinchang, R., Stephen, M. A Framework for Breast Cancer Classification Using Multi-DCNNs. Computers in Biology and Medicine, 2021, 131. https://doi.org/10.1016/j.compbiomed.2021.104245
6. Gao, Y., Zhang, L., Zhou, J., Yuan, W., Qiu, Y. Improved Extreme Learning Machine based on Adaptive Dual-Strategy Optimization Algorithm and Its Application. Research Square, 2022. https://doi.org/10.21203/rs.3.rs-2293384/v1
7. Hao, L., Lianbing, D., Zhaoaq, G. A Robust Image Encryption Algorithm Based on a 32-bit Chaotic System. IEEE Access, 2020, 10. https://doi.org/10.1109/ACCESS.2020.2972296
8. Ivars N. Deep Convolutional Neural Networks: Structure, Feature Extraction and Training. Information Technology and Management Science, 2017, 20. https://doi.org/10.1515/itms-2017-0007
9. Jagadeesh Kumar, S. J. K., Parthasarathi, P., Mofreh A. H., Mehedi M., Jehad F. A., Mohamed A. Breast Cancer Detection Using Breastnet-18 Augmentation with Fine Tuned Vgg-16. Intelligent Automation & Soft Computing, 2023, 36(2), 2363-2378. https://doi.org/10.32604/iasc.2023.033800
10. Karthik, S., Srinivasa, P. R., Chandra M. Breast Cancer Classification using Deep Neural Networks. Knowledge Computing and Its Applications: Knowledge Manipulation and Processing Techniques, 2018, 1. https://doi.org/10.1007/978-981-10-6680-1_12
11. Khan, S. U., Islam, N., Jan, Z., Khalid, H., Syed, I. A. S., Muhammad, H. A Machine Learning-based Approach for the Segmentation and Classification of Malignant Cells in Breast Cytology Images Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM). Neural Comput & Applic, 2022, 34, 8365-8372. https://doi.org/10.1007/s00521-021-05697-1
12. Khaoula, B. S., Naima, K., Mohamed, N. S., Ahmed, T. Breast Cancer: One-Stage Automated Detection, Segmentation, and Classification of Digital Mammograms using UNet Model based-Semantic Segmentation. Biomedical Signal Processing and Control, 2021, 66. https://doi.org/10.1016/j.bspc.2021.102481
13. Kharel, N., Alsadoon, A., Prasad, P. W. C., Elchouemi, A. Early Diagnosis of Breast Cancer Using Contrast Limited Adaptive Histogram Equalization (CLAHE) and Morphology Methods. 8th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2017, 120-124. https://doi.org/10.1109/IACS.2017.7921957
14. Kim, C. M., Park, R. C., Hong, E. J. Breast Mass Classification Using eLFA Algorithm Based on CRNN Deep Learning Model. IEEE Access, 2020, 8, 197312-197323. https://doi.org/10.1109/ACCESS.2020.3034914
15. Lee, H., Park, J., Hwang, J. Y. Channel Attention Module with Multiscale Grid Average Pooling for Breast Cancer Segmentation in an Ultrasound Image. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2020, 67(7), 1344-1353. https://doi.org/10.1109/TUFFC.2020.2972573
16. Liao, W. X. Automatic Identification of Breast Ultrasound Image Based on Supervised Block-Based Region Segmentation Algorithm and Features Combination Migration Deep Learning Model. IEEE Journal of Biomedical and Health Informatics, 2020, 24(4), 984-993. https://doi.org/10.1109/JBHI.2019.2960821
17. Loizidou, K., Skouroumouni, G., Nikolaou, C., Pitris, C. Automatic Breast Mass Segmentation and Classification Using Subtraction of Temporally Sequential Digi-

tal Mammograms. IEEE Journal of Translational Engineering in Health and Medicine, 2022, 10, 1-11, 1801111. https://doi.org/10.1109/JTEHM.2022.3219891

18. Maqsood, S., Damasevicius, R., Maskeliunas, R. TTCNN: A Breast Cancer Detection and Classification Towards Computer-aided Diagnosis Using Digital Mammography in Early Stages. Applied Sciences, 2022, 12(7). https://doi.org/10.3390/app12073273
19. Medari, T. Bidirectional Grated Recurrent Unit for Shallow Parsing. Indian Journal of Computer Science and Engineering, 2020, 11. https://doi.org/10.21817/indjcse/2020/v11i5/201105167
20. Michael, M. K., Arokia, R. Deep Joint Segmentation for the Classification of Severity-Levels of Glioma Tumour Using Multimodal MRI Images. IET Image Processing, 2020. https://doi.org/10.1049/iet-ipr.2018.6682
21. Pavel, T., Mohammad, D. Pelican Optimization Algorithm: A Novel Nature-Inspired Algorithm for Engineering Applications. Sensors, 2022, 22. https://doi.org/10.3390/s22030855
22. Prabhpreet, K., Gurvinder, S., Parminder, K. Intellectual Detection and Validation of Automated Mammogram Breast Cancer Images by Multi-Class SVM Using Deep Learning Classification. Informatics in Medicine Unlocked, 2019, 16. https://doi.org/10.1016/j.imu.2019.01.001
23. Praveen, K., Shantha, R., Selva, K. VLSI Architecture for LGXP Texture for Face Recognition. Journal of Intelligent and Fuzzy Systems, 2014, 27. https://doi.org/10.3233/IFS-1412366
24. Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., Chen, H. A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. IEEE Access, 2021, 9, 71194-71209. https://doi.org/10.1109/ACCESS.2021.3079204
25. Se, W. C., Na, R. B., Kang, R. P. Deep Learning-based Multi-Stage Segmentation Method Using Ultrasound Images for Breast Cancer Diagnosis. Journal of King Saud University - Computer and Information Sciences, 2022. doi:10.1016/j.jksuci.2022.10.20
26. Shen, T., Gou, C., Wang, J., Wang, F. Y. Simultaneous Segmentation and Classification of Mass Region from Mammograms Using a Mixed-Supervision Guided Deep Model. IEEE Signal Processing Letters, 2020, 27, 196-200. https://doi.org/10.1109/LSP.2019.2963151
27. Singh, B., Kaur, R., Kaur, A. Discoursing Novel Procedure for Segmentation and Classification of Mammograms. SN Computer Science, 2021, 2, 61. https://doi.org/10.1007/s42979-021-00454-6
28. Tamilselvi, P. R., Thangaraj, P. Improved Gabor Filter for Extracting Texture Edge Features in Ultrasound Kidney Images. Modern Applied Science, 2010, 4. https://doi.org/10.5539/mas.v4n4p62
29. Vijay, K., Priyanka, G. Importance of Statistical Measures in Digital Image Processing. International Journal of Emerging Technology and Advanced Engineering, 2012, 2(8).
30. Wang, L., Song, T., Katayama, T., Jiang, X., Shimamoto, T., Leu, J. S. Deep Regional Metastases Segmentation for Patient-Level Lymph Node Status Classification. IEEE Access, 2021, 9, 129293-129302. https://doi.org/10.1109/ACCESS.2021.3113036
31. Wessam, M. S., Moustafa, H. A. Deep Learning in Mammography Images Segmentation and Classification: Automated CNN Approach. Alexandria Engineering Journal, 2021, 60, 4701-4709. https://doi.org/10.1016/j.aej.2021.03.048
32. Yue, Z., Houjin, C., Yanfeng, L., Qin, L., Xuanang, X., Shu, W., Pew, T. Y., Dinggang S. Multi-Task Learning for Segmentation and Classification of Tumors in 3D Automated Breast Ultrasound Images. Medical Image Analysis, 2021, 70. https://doi.org/10.1016/j.media.2020.101918
33. Zebari, D. A., Ibrahim, D., Zeebaree, D. A. Systematic Review of Computing Approaches for Breast Cancer Detection based Computer-Aided Diagnosis using Mammogram Images. Applied Artificial Intelligence, 2021, 35(15), 2157-2203. https://doi.org/10.1080/08839514.2021.2001177
34. Zhang, G., Zhao, K., Hong, Y. SHA-MTL: Soft and Hard Attention Multi-Task Learning for Automated Breast Cancer Ultrasound Image Segmentation and Classification. International Journal of Computer Assisted Radiology and Surgery, 2021, 1719-1725. https://doi.org/10.1007/s11548-021-02445-7