JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

JSCDM

Vol. 5 No. 1 (2024) 152-169
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Breast Cancer Diagnosis Using Majority Voting Ensemble Classifier Approach

M. Mohana Dhas[1]*, **N. Suresh Singh**[2]

[1] *Department of Computer Science, Annai Velankanni College, Tholayavattam, Kanyakumari District, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 627012, Tamilnadu, INDIA*
[2] *Department of Computer Applications,*
Malankara Catholic College, Mariagri, Kaliyakkavilai 629153, INDIA

*Corresponding Author: drmohanadhas@gmail.com
DOI: https://doi.org/10.30880/jscdm.2024.05.01.013

Article Info

Received: 1 December 2023
Accepted: 25 April 2024
Available online: 21 June 2024

Keywords

Breast cancer diagnosis, AGBF, MRMR, ABHC, ensemble classifier

Abstract

One of the most common cancer affects women are the Breast Cancer. These are predicted using the diagnosing methodologies. The CT scans are the most common screening mechanisms is the computer aided prediction mechanism. Still, based on the reasons of genetic information, the microarray data is applied as the solution for diagnosing the cancer cells. Dealing with the microarray data it has various consequences, from those consequences one of them is its high dimensionality. The primary intention of this approach is to introduce an ensemble model to diagnose the cancer cells from the dataset. Initially, Adaptive Guided Bilateral Filter (AGBF) is applied to denoise the input images and to filter the images by sharpening them. Afterwards, to segment the images by the conversion of RGB images into grey levels the threshold mechanism is used. Minimum Redundancy Maximum Relevance (MRMR) feature inclusion and Local Search Adaptive Beta Hill Climbing (ABHC) are two hybrid feature selection approaches used in this article where the features and its importance were stored on a dataset. Then the diagnosing of the breast cancer is done by obtaining the majority voting ensemble classifier. The proposed approach is evaluated using a Breast Cancer Histopathological Image Classification (BreakHis) dataset and achieves the classification accuracy of 99.51%.

1. Introduction

The breast cancer is classified from one of the type of tumor cells that takes place in breast. Here both women and men are affected by this but women are widely affected and they suffer a lot. The initial phase of breast cancer involves the unregulated proliferation of cells in the lining of the breast. Initially, during the cancerous growth there should not be any pain or symptoms of cancer growth and where it generates without any symptoms it is limited to the lobe. The change in skin color, change in shape of the breast and a small lump in the breast are the various symptoms of the breast cancer. Some forms of breast cancer may not exhibit symptoms in the early stages. To predict the cancer cells earlier, it is very much needed to undergo early prediction tests. In the breast one of the type of a cancer occurs in the abnormal milk producing cells named Lobular Carcinoma in Situ (LCIS). The lobules in milk producing mammary glands are developed by the Invasive Lobular Carcinoma, moreover the people affected by this experiencing the symptoms such as swelling in the breast, thickening of breast tissue, and change in the texture of the skin. Another one type of cancer is Ductal Carcinoma in Situ in which it does not cause any symptoms. This type of cancer is identified by the mammography and it is the one of the common type of cancer [1].

This is an open access article under the CC BY-NC-SA 4.0 license.

Breast cancer is the most prevalent form of cancer among women globally [2]. Several incidence-based mortality and previous randomized trials studies have shown that participating in breast screening programs leads to a decrease in breast cancer mortality [3]. Many errors are obtained on the dense breasts and in certain human factors: erroneous decision making behavior and the perception of the radiologist. Benign tumors often resemble malignant ones, and malignant tumors sometimes mimic benign ones which are done additionally by the inherent imaging characteristics of tumor contribution [4]. The physician's goal is to examine all Region of Interest (RoI) in the mammogram, including calcifications, distortions, and masses; during cancer diagnosis process. Predicting diseases in the early stages is challenging when it comes to planning new analyses, courses of action, or treatments [5].

For the prediction of breast cancer, the Deep Learning (DL) and the Machine learning (ML) approaches plays a major role and also it turns to be a powerful tool in the diagnosis of cancer. In order to study the complex hierarchical data representations automatically DL uses the artificial neural network. One of the basic concept in DL is transfer learning it reform the field by applying pre-trained models which are trained on large dataset for specific task. From a large amount of data, the pre-trained models have studied about meaningful and intricate features. Here these are applied as a valuable source for feature extraction in cancer diagnosis. Through the utilization of the information included in these pre-trained models, we may accelerate the training process, improve performance of the model, and get around the constraints imposed by inadequate data. Furthermore, by using these pre-trained models, we may concentrate on adjusting the model to the particular task at hand, leading to a more effective and precise diagnosis of breast cancer [6].

A median filter is utilized in the segmentation model alongside a probabilistic clustering model based on multiple kernels for processing noise-free images. For every segment the characteristics were extracted and for identifying the most significant features the Linear Discriminant analysis was used with a DL algorithm for the classification of the brain images. Also various analysis technologies are used beyond the system to overcome the issues [7]. In various fields, threshold-based methods are more suitable for capturing grayscale data that accurately represents the image of the object. The threshold segmentation approach is highly effective in the field of medical image analysis. Here based on the Bi-level and the multi-level the threshold segmentation approaches are classified. In the medical scenario the multiple objects are needed to be described and in the recent times the high amount of attention was received by multi-level threshold. Depending on the basic needs the threshold segmentation level increases which causes complexity that increases exponentially [8].

Feature Selection is a frequently employed approach to minimize computational demands by eliminating unnecessary features and choosing a subset of unique features. For the purpose of classification, the availability of redundant features does not get any importance. In various domains like the travelling salesman problem solving, image enhancement, security, classification, solving of class imbalance problem, and SVM parameter optimization; there are numerous optimization algorithms are used for selecting the features. Here applying the concept of local search was not much older. In recent times researchers successfully developed and applied several approaches from those approach this model outperforms better in the form of classification. It is very difficult to choose the optimal feature subset especially in the wrapper based approaches. The exploitation ability of the feature selection subset was improved by the ensemble of local search and thus it improves the overall learning outcome [9]. The wrapper approaches are obtained as the search algorithms in which various combination of features are generated and analyzed. Although many wrapper techniques produce better results, they come with a high overhead cost and might take a very long time if the feature set is huge [10]. Feature ranking is typically a phase in the feature selection process, which is the ranking of original features according to particular evaluation criteria. It is used in determining the relative significance of several feature sets or features [11].

The contribution of the proposed methodology are as follows:

- In order to improve the diagnosis of breast cancer, the proposed system incorporates a highly effective method for classifying data. Data preprocessing is initially carried out, and the AGBF model assigns the weights.
- The tumor cells are segmented by using a threshold model, in which it applies the process of separating foreground from the background and also it converts the RGB image into grey level image.
- To overcome the high dimensionality on the BreakHis dataset MRMR feature selection algorithm has been applied.
- It applies a local Search ABHC based Feature section model to create an optimal feature subset from features developed by the CNN model.
- The majority voting ensemble mechanism is used for the classification of the breast cancer. It is also applied for boosting the efficiency of the proposed model.

The rest of this article is detailed in the below section. The literature review is explained in section 2. Proposed methodology is detailed in section 3. Section 4 explains about the experimental setup of this article. Section 5 explains the experimental results part. Finally, the article is thoroughly concluded in section 6.

2. Literature Review

In this chapter various articles were reviewed according to the cancer diagnosis with various steps and the method applied for those stages were briefly illustrated.

The identification of breast cancer turns into a challenging issue for researchers and clinicians. So, Iparraguirre-Villanueva et al., (2023) [1] introduced ML models to diagnose the possibility of the patients having breast cancer. For the scaling of the features the normalization approach was used by converting the values of the feature. Thus the accuracy of the method shown high prediction of cancer diagnosis value. To extract the meaningful features automatically and to categorize the various medical conditions to be faster and accurate during the cancer diagnosis. There exist some drawbacks such as the overfitting problem and high consumption in runtime. To overcome these issues Balasubramaniam et al., (2023) [12] developed a classic CNN approach called LeNet to predict the Breast cancer. By this model it helped in the identification of subtle abnormalities in breast image that which shows the presence of cancer in the early stage.

The preprocessing of image in the infrared images were aimed in image denoising and enhancement in helping with small target prediction. The original image and the noise were categorized by the sparse model and low rank. Under high noise the Infrared (IR) image denoising was challenging to perform well. Without proper denoising approach detection and tracking was not possible. Thus to overcome these issues Chen et al., (2023) [13] devised a Weighted Nuclear Norm Minimization (WNNM) integrated with adaptive similar patch. This approaches obtain better denoising outcome not only in common image but also it accomplished higher results even in the infrared images. Prathik et al., (2022) [14] developed a denoising approach that takes the noisy image for processing and for indexing it moves row window. By using Discrete Wavelet Transform (DWT) the noisy image were decomposed the signal to get localization and its coefficients were transformed into independent distributed variables. Depending on the threshold the coefficients were analyzed. This approach had the improved accuracy range and thus it shows the devised approach to be more effective.

The breast cancer area was complex and distributed by shadows or similar tissues. So in that case providing better segmentation image for the cancer diagnosis turns difficult. To overcome this issue a multi-level threshold image segmentation method was developed by Yang et al., (2023) [8] for the partitioning of breast cancer pathological images, yielding improved samples for further analysis in computer-aided diagnosis of breast cancer. In segmentation framework the threshold was the key approach. Also they applied an enhanced differential evolution framework for the cancer diagnosis. This optimization approach ranked better in average ranking of the two statistics. Under a microscope, the complicated structure and hazy cell boundaries in breast cancer histopathology tissue images present difficulties for precise segmentation using conventional thresholding techniques. Here, Wang et al., (2023) [15] a segmentation method that uses the Enhanced Dandelion Optimization (IDO) algorithm to accurately distinguish lesion areas in breast cancer images. In order to find the ideal thresholds, it combines the golden jackal energy judgment process with fallback tactics and a memory matrix. This approach effectively resolves the segmentation problem and provides faster convergence.

To select the effective feature subset from the original features a data processing approach is used in the form of feature selection. Nevertheless, Evolutionary Computation (EC)-based feature selection techniques are unable to successfully eliminate faulty features. Until the termination of the algorithm a small amount of invalid features still exists. To overcome the drawbacks and issues Liu et al., (2022) [11] developed a feature selection approach using a combination of Genetic Programming (GP) and Feature Ranking (FRFS). The FRFS using the Multi-criteria fitness function to lower the number of selected features which was identified as the MFRFS. While maintaining the classification performance MFRFS minimized the number of features. The feature ranking approaches used for the attribute subset evaluation feature selection methods are not very suitable based on the large size of the search space of the possible solution. Most of the univariate feature ranking approaches did not detect the interactions among the factors. To resolve those difficulties, Jiménez et al., (2022) [16] devised two multivariate feature ranking models depending on the pairwise consistency and pairwise correlation. The devised approach was applied using a public dataset for genotype tissue expression and cancer gene expression. This approach statistically outperformed the state-of-art feature ranking approach.

For the improvement of the robustness of feature selection ensemble feature selection approaches were developed. Every feature selection approaches may identify feature subsets that were considered in the space of the local optima. To address the issue ensemble feature selection was applied. To overcome the issues of the problem Hashemi et al., (2023) [17] developed a bi-objective feature selection algorithm depending on the pareto-based ranking called MRMR. The runtime of this approach was low and this approach had the high classification accuracy. The key concept of the Feature selection was ML. Based on the information theory there were various difficulties in the feature selection. Thus for overcoming them Bugata & Drotar (2020) [18] developed a feature selection approach named as MRMR which was the popular application. The MRMR technique offered a theorem regarding the equivalency of MRMR and Max-Dependency for first-order incremental search. We examined this theorem's link to the Max-Dependency criteria and created a discrete counterexample. By employing various other suitable dependency measures the outcome on the high eight dimensional dataset performance of MRMR was originally associated with MI.

A 2D Renyi entropy multi-threshold image segmentation approach was developed by Chen et al., (2023) [19] to assists the pathologists and the evaluation of histopathological images of LN. This approach was depending on the Cukoo Search (CS) algorithm that develops the Diffusion Mechanism (DM) and an ABHC approach named as DMCS approach. This DMCS was based on the Multi threshold image segmentation and it was applied on the renal pathological images. Thus for renal pathological images the devised approach showed the better image segmentation approach. There were various difficulties in the classification, for classification with greatest importance the dimension of the feature vector was applied. For solving the difficulties in the feature selection Ahmed et al., (2021) [20] developed an improved Coral Reefs Optimizer (CRO) by improving its capacity of searching with the assistance of Local Search Algorithm, ABHC. The novel approach devised here was ABCRO, which was developed by integration of CRO and ABHC. Also in finding the optimal solution there were few difficulties and the ACBRO fails. Here the no free lunch theorem was recognized as weakness of the approach. Pneumonia was a severe respiratory disease that affects in our lungs, which affects the patients by oxygen intake and the difficulty in breathing. Children under the age of 5 where the pneumonia was the leading death rate disease. Based on these Shukla et al., (2022) [21] introduced a model in which the features were extracted from the chest radiograph images with the help of DenseNet-201. This extracted features undergo a feature selection by applying a mixture of Binary Sine Cosine Algorithm (BSCA) and ABHC Local Search Algorithm for selecting the features from the extracted set of features. The devised approach had a higher accuracy in the prediction of pneumonia.

3. Proposed Methodology

The primary intention of this approach is to diagnose the breast cancer using the Enemble classifiers. Initially, the input data are acquired from the BreakHis dataset. The obtained data are then forwarded to the preprocessing phase to remove the unwanted noises in the image, which will be done by using AGBF. The AGBF is used for filtering the images once after they are gathered. Afterwards, the preprocessed image will be then forwarded to the segmentation phase which is done by obtaining the threshold value. Then the segmented images are forwarded to the feature selection phase wherein the features are selected by using the MRMR which select the needed feature from the overall feature set. Then the selected features are forwarded to the training and testing set. For the training set the fitness values are calculated. Finally, the base learners are used for the classification by identifying the majority voting classifier for the effective cancer diagnosis. Fig. 1 shows the overall framework of the proposed model.

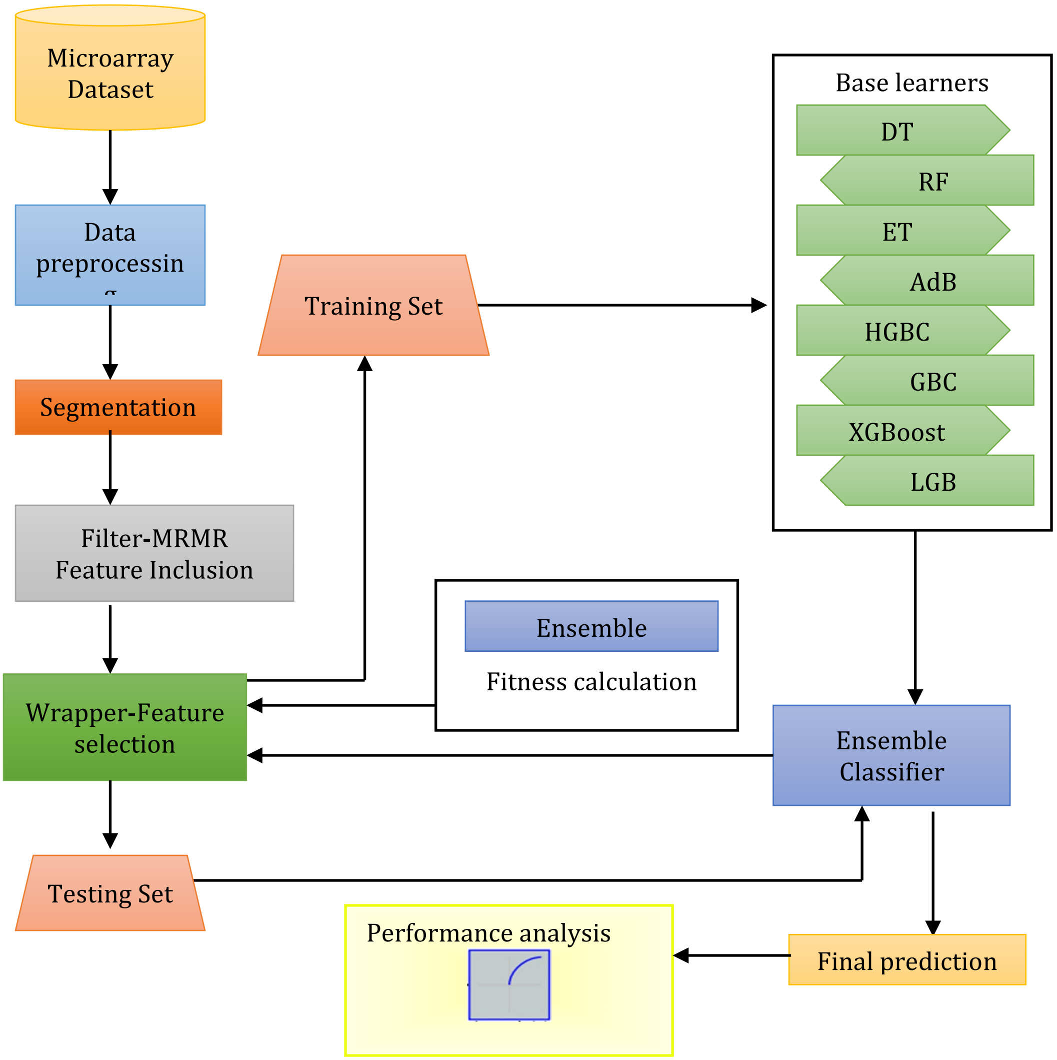


Fig. 1 *Workflow of the proposed model*

3.1 Preprocessing

Considering the Breast cancer prediction approach, for the further analysis image processing plays a major role in the process of preparing input images. Here to extract the meaningful data and to enhance their quality it includes a set of operation used to the raw images. The subsequent analysis shows the improved performance and accuracy as a result of effective image preprocessing. For the detection of the breast cancer the proposed method uses various image preprocessing models to prepare the HPI images [6].

- Resizing of image:
  To decrease the computational complexity and maintain a consistent input image, we resized the images to a standard size of 224×224 pixels. They have specific size requirements while applying pre-trained DL models. The leveraging of the learned features is allowed by image resizing. During the training and interference, the resizing to a standard size makes the process more effective.
- Conversion of BGR-to-RGB:
  The Blue-Green-Red (BGR) are the original HPI color space. From the original color space, it was converted into Red-Green-Blue (RGB) color space which is easily suitable for the subsequent image analysis. Across the dataset this conversion step ensures the color representation.
- Image Sharpening Filter:
  To enhance the fine details and the edges of the images we apply an image sharpening filter. By increasing the contrast between image characteristics, this filter helps with feature extraction by making them easier to differentiate from one another.
- Image Scaling:

  We conducted an image scaling by dividing the pixel values by 255 in order to standardize the pixel values and place them within a consistent range. The convergence of the subsequent ML algorithms is aided by the normalization procedure. The pixel values are in consistent numerical range and it was ensured by the normalization procedure.

- Image Labelling:
  In textual form the original HPI dataset had object labels. By the conversion of object labels into numeric labels, we performed image labelling for facilitating subsequent analysis. This translation makes it possible to employ different ML algorithms for classification problems that require numerical labels.

In the Breast cancer detection approach the preprocessing techniques provides several benefits. The consistent image sizes are ensured by the image resizing and from the dimension variation it minimizes the bias initially. Afterwards, for the reliable feature extraction the conversion of BGR to RGB standardizes color representation. Then the fine details are enhanced by image sharpening and it improves the discrimination of relevant features. Next to that the pixel values are normalized by scaling and it aids the efficient model training. Atlast by the use of the diverse classification algorithms enabled by image labelling. The enhancement of accuracy and reliability of the breast cancer diagnosis system enhance the reliability and accuracy also for the deep feature extraction and ML it prepares the HPI images [6].

3.1.1 Denoising Using Adaptive Guided Bilateral Filter (AGBF)

Obtained from the above preprocessing techniques other than these the denoising have a significant role in the preprocessing which removes unwanted noises in the image. Here the AGBF approach is applied for the denoising of the image. The use of AGBF is to enhance the quality of infrared images by reducing noise and preserving edges.

The noise-removal and sharpening capacities of AGBF are analyzed. Based on the shift variant technique this approach was highly concentrated. The existing enhancements allow AGBF to outperform traditional BLF in terms of noise reduction and image enhancement [7]. The AGBF's filter kernel and weighing function are represented as follows in equation (1) and (2):

$$EPL(K)_B = \frac{1}{\sum_{S\in U_P} u_{EPL_{bs}}(K)} \sum_{S\in U_P} u_{EPL_{bs}}(K)\, K_S \tag{1}$$

$$u_{EPL_{bs}}(K) = exp\left(-\frac{\|B-S\|^2}{2\rho_A^2}\right) exp\left(-\frac{|(K_B+\Xi_B)-K_S|^2}{2\rho_A^2}\right) \tag{2}$$

Where the image sharpening offset is represented by Ξ_B. Equation (3) shows the native choice for this value.

$$\Xi_B = \begin{cases} max(U_P) - K_B \ if \ \Delta_B > 0 \\ min(U_P) - K_B \ if \ \Delta_B < 0 \\ \quad if \ \Delta_B = 0 \end{cases} \tag{3}$$

Where G represents the variation in intensity between the pixels, (U_P) denotes the average of the local window. These two are represented by the given expression which is $\Delta_B = l_B - \vartheta_P$. The $max(U_P)$ represent the maximum values within the local window (U_P) and the $min(U_P)$ represent the minimum values of the local window (U_P).

3.2 Segmentation by Thresholding

Thresholding is a technique of image segmentation that divides an image into two or more regions based on a threshold value. They perform local thresholding for each pixel, depending on the statistical variability of the gradient vector at that pixel. They claim that their method can detect edges more accurately and robustly than conventional methods [7].

In this stage, segmentation involves combining the obtained MRI data from brain tumors. Here the threshold technique is the key strategy in the image segmentation. The total number of distinct pixel values are divided by this approach and by the use of the total number of distinct pixel values it excludes the zeros. The average grey value is identified for the conversion of grey scale image to a binary image [7]. In equation (4) the approach may have characterized.

$$T = \sum_{a=1}^{W} \frac{[\forall(E,G)\in a](E\neq G\neq 0)\rightarrow \emptyset(E,G)}{X} \tag{4}$$

Where, T represents the threshold value. E and G are the pixel values, X indicates the amount of distinct pixel values.

3.3 Filter Based Feature Inclusion Using Minimum Redundancy Maximum Relevance (MRMR)

The key feature of the microarray dataset is high dimensionality, which consists of few dissimilar features. In the model accuracy these irrelevant features create major degradation. One of the most emerging solutions for overcoming the above mentioned issues is feature selection. In order to select the feature with the MRMR which is a well-known feature selection technique. The dimension of the microarray dataset is $M \times N$, as shown in equation (5). Where, M shows the samples, and N shows the features.

$$d_{M\times N} = \begin{bmatrix} F_{11} & F_{12} & \cdots & C_1 \\ F_{21} & F_{22} & \cdots & C_2 \\ \vdots & \vdots & \ddots & \vdots \\ F_{M1} & F_{M2} & \cdots & C_N \end{bmatrix} \tag{5}$$

The J^{th} feature associated with I^{th} sample is represented by F_{IJ}, the class label for the respective sample is C. The equation (6) helps in the calculation of the feature entropy $h(F)$.

$$h(F) = \sum_{C=1}^{n} p_R(C_I)\log\left(p_R(C_I)\right) \tag{6}$$

where, the probability of the class $C_I$is represented as $p_R(C_I)$, $I \in \{1,2,\ldots,M\}$. Equation (7) has the ability to define the conditional entropy of a feature F with respect to the feature vector (f).

$$h(F|f) = \sum_{I=1}^{N} p_R(F_I)\left(\sum_{J=1}^{M} p_R\left(C_J|f\right)\log\left(p_R\left(F_J|f\right)\right)\right) \tag{7}$$

where, the probability of the feature $F_I$is represented by $p_R(F_I)$. The conditional probability of class $C_J$from the feature vector $f$is indicated by the $p_R\left(C_J|f\right)$.
Here the feature $F_I$in terms of feature vector f for the mutual information $mi(F_I|f)$ is expressed in equation (8)

$$mi(F_I|f) = h(F) - h(F|f) \tag{8}$$

Equation (9) expresses the rewritten version of equation (8)

$$mi(F_I|f) = \sum_{I=1,J=1}^{M,N} p_R\left(C_I,F_J\right)\frac{p_R(C_I,F_J)}{p_R(C_I).p_R(F_J)} \tag{9}$$

Based on the redundancy and relevance measure the MRMR algorithm takes place. The classification or prediction accuracy was improved by a high relevance measure(r_L), indicating a wide range of quality genes in the dataset. A large amount of redundant genes available in this dataset was indicated by the high values in the case of redundancy r_D. Here the low value is highly preferred for the redundancy measure [22]. By applying the equation (10) and (11) the $m_L$and m_D can be calculated.

$$Max(r_L) = \frac{1}{|f|}\sum_{F_I.F_J\in f} mi\left(F_I;F_J\right) \tag{10}$$

$$Min(r_D) = \frac{1}{|f|^2}\sum_{F_I.F_J\in f} mi\left(F_I;F_J\right) \tag{11}$$

Considering the class label C for optimizing the $m_L$and m_D the two evaluation criteria are selected in an either or basis they are mutual information difference mi_D and mutual information quotient mi_Q which are indicated in equation (12) and (13).

$$mi_D = Max[r_L - r_D] \tag{12}$$

$$mi_q = Max\left[\frac{r_L}{r_D}\right] \tag{13}$$

From the feature set f consider a feature subset f_X with X selected features. Thus from the remaining feature set the selection of next $(X+1)^{th}$ features are done by applying equation (14) and (15).

$$mi_D = max\left[MI(F_K;C) - \left(\frac{1}{X}\right)\sum_{F_I \in f_X} mi(F_K;F_I)\right] \tag{14}$$

$$mi_Q = max\left[\frac{MI(F_K;C)}{\left(\frac{1}{X}\right)\sum_{F_I \in f_X} mi(F_K;F_I)}\right] \tag{15}$$

Where, $F_K \in f - f_X$

3.3.1 Metric Ranked Feature Inclusion (MRFI)

The MRFI introduced here involves a two-step process: the initial step entails ranking features, followed by selecting the best attribute subset from this ranking. The ensemble classifier is used for the ranking. Based on the training and testing set the entire dataset is categorized. In a ranked order the features are stored in another data frame in the column form. From the whole feature set this model selects a feature and the number of classes is performed by the target and the feature. The number of classes is helpful for predicting the number of classes. The importance of every feature is assigned in the identified score. This procedure is carried out for each feature separately. The entire features and their significance are saved in a data frame. Thus for obtaining the feature ranking data frame is sorted from most importance to least importance [10].

By applying the ranking of the features, the feature set selection stage is performed and for this procedure we develop a novel algorithm. In the ranking the initial feature of RF classifier is obtained and the accuracy of RF classifier is analyzed. The next feature from the ranking is integrated with the other features to form the optimal feature subset, also it recomputed the accuracy of that classifier. We keep the feature in the best selection when accuracy improves. Also, from the optimal feature subset it drops the attribute during the accuracy decreases. This is done iteratively for every attributes to take the final optimal feature subset [10].

3.4 Wrapper Approach Feature Selection Using Local Search Adaptive Beta Hill Climbing (ABHC)

Then the included features are delivered to the feature selection phase. Here the selection of the needed features is obtained by applying the ABHC. The Local search ABHC is used to create an optimal feature subset from the features generated by MRMR model.

Local search serves as a heuristic approach for tackling computationally intensive optimization problems. While solving the issues local search are useful in the way of they are characterized as minimizing or maximizing a criterion across a large domain. Until an optimal solution is discovered or the maximum number of iterations has been reached, the local search explores different solution by applying some local changes. One of the popular local search approach applied here is ABHC. In the local optima the hill climbing usually faces difficulties. However, to overcome these issues ABHC is used [9]. This algorithm takes a T location agent as input in the search space and produces a modified location of the agent as output. This algorithm was depended based on two operators they are $Beta$ operator and the Neighborhood operator (Z). From a solution $T(T_1, T_2, T_3, \dots, T_g)$ given below the Z operator randomly selects a neighbor $T'(T_1', T_2', T_3', \dots, T_g')$:

$$T_n' = T_n \pm rand(0,1) \times Z \;\; where\; n = 1,2,\dots,g \tag{16}$$

Where, the greatest probable distance among the neighbors and the present solution are represented by Z, the function which generates the random number among 0 and 1 are indicated by $rand(0,1)$. The mutation operator in GA motivates the $Beta$ operator. For the new solutions, it randomly assigns values from the similar range with a certain probability $Beta = rand(0,1)$ or as the present solution given below.

$$T_n'' = \begin{cases} T_n \; if\; Beta > rand(0,1) \\ T_n' \qquad\qquad otherwise \end{cases} \tag{17}$$

Where, the j^{th} dimension of updated location of the solution is represented as T_n'', the j^{th} dimension of updated location of the previous solution is indicated by T_n, and the j^{th} dimension of updated location of neighborhood previous solution is indicated by T_n'.

Currently, the values of N and $Beta$ determine the majority of the outcome in this variation of hill climbing. Comprehensive experiments are required to determine the values of these two factors. In order to get around this limitation, ABHC was established. Z and $Beta$ in ABHC are functions of the iteration count.

$$Z(b) = 1 - \frac{b^{\frac{1}{e}}}{MI^{\frac{1}{e}}} \;\; where\; e = constant \tag{18}$$

Where, the maximum number of iterations are represented by MI, b signifies the current iteration number. In the b^{th} iteration the $Beta$ value is represented as

$$Beta(b) = \frac{(max - min) \times b}{MI} + min \tag{19}$$

Where, the maximum values of $Beta$ and current iteration are represented by max, min shows the minimum values of $Beta$ and current iteration. The n compared with T the T'' is better, then the T'' is replaced for T.

3.4.1 Fitness Function

This goal of this section is to outline the process for assessing a proposed solution's quality. For computing the classification accuracy, a candidate solution, it exploited an Ensemble learning classifiers [22]. There are two main components in fitness function they are classification accuracy and number of features, however these components are opposite to one another. It is necessary to decrease the number of features while maximizing classification accuracy. Therefore, the decision was made to utilize classification error. The improved fitness score is indicated by a lower error value. The strength of a given feature set is accessed by a fitness function which is shown in equation (20).

$$FF = c \times \propto (1 - c) \ \times \frac{|f_d|}{|f_s|} \tag{20}$$

Where, the total number of features in the selected feature set is represented as $|f_s|$, $|f_d|$ signifies the total amount of features in the dataset, the error in the classification using feature subset is indicated as $\propto$, and the relative weight value assigned to the number of features and the classification error are signified by $c \in [0,1]$.

3.5 Classification Using Machine Learning Classifiers

The traditional ML algorithms and a mixture of deep feature extraction are used in the study of breast cancer diagnosis. The performing of exploratory data, reduction of dimensionality, enhancing out-of distribution detection, class imbalance addressing are the benefits gained from this approach and it also employs the model compression approaches. The ML algorithms are widely used for achieving the objectives as represented below.

- Decision Tree (DT):
  A supervised learning algorithm that is non-parametric divides the input space into sections and assigns a class label to every section. Data mining is the widely used technique that create classifiers. The classification algorithms have the capacity to handle the vast volume of data in data mining, to classify the newly obtainable data and to classify the knowledge on the basis of class labels and training sets. DT are a sequence of successive models that effectively and cohesively combine a number of fundamental tests in which each test compares a numerical property to a threshold value. Compared to the numerical weights in the neural network, constructing conceptual rules is much simpler. The DT is highly applied for the grouping purposes. However, in data mining the DT is the widely used classification model [23].
- Random Forest (RF):
  To improve the prediction accuracy an ensemble method applies multiple decision trees named as RF. This is a ML model based on tree that analyses the regression magnitude as consensus of number of models. In the training period and outputting mean prediction of the decision trees it operates by constructing a multimedia on a decision tree. The overfitting problems exhibited by tree models are solved by this approach. For combining accuracy, generalization power and ability to capture non-linear relations this model was widely used [24].
- Extra Trees (ET):
  The extension for the RF algorithms are ET which furtherly randomizes the construction of trees. An ensemble of undecorated decision trees is generated by Extra-Trees with the help of conventional top-down methodology. In the selection of attributes and cutting points this process involves considerable randomness in the splitted nodes. From the training sample it builds fully randomized trees with structures independent of output values in most of the extreme cases. This approach clearly has two aspects they are during the node splitting it applies entire random cut point selection and the other is to grow the trees it employs the entire training sample. By aggregating all the trees through majority voting the final prediction is established [25].
- AdaBoost (AdB):

To create a strong classifier an ensemble learning method is introduced which combines multiple weak classifier known as AdB. During the AdaBoost training phase every samples are equally weighted with Wi. By increasing the weights assigned to misclassified data, the weights are subsequently enhanced through repetition. Various weak learners can be integrated with a weighted sum using the AdaBoost technique to produce the ultimate output of the Boosted classifier. However, in the process of implementation the AdaBoost algorithm becomes simple and it improves the classification accuracy [26].

- Histogram Gradient Boosting Classifier (HGBC):
  One of the gradient boosting classifier is HGBC that applies histogram based training for effective prediction and training. To form a high effective strong learner boosting techniques depend on the integration of multiple weak learners. The decision tree is the boosting technique implemented in the gradient boosting. For the problem of classification and regression the newly developed algorithm performs well across wide range of various datasets. The difficulty identified is similar to the nature of Decision Tree approach and a large number of samples and features. For the difficulty of having a large amount of dataset and in the efficiency, the HGB comes up with a solution. The continuous samples into the constant number of bins used by the histogram is mentioned as its name [27].
- Gradient Boosting Classifier:
  An ensemble of weak learners is usually build by a GBC algorithm. After the inspiration of the AdaBoost the GBC is raised. To get the strong learner from a weak learner the boosting method is used. Every new tree in boosting corresponds to the updated original data. Boosting is not done at random; rather, it is customized to the sequence of weak learners. In every stage with in build trees, every iteration uses samples with different weights. AdaBoost merges all of the trees it uses for weight at the end of the procedure [28].
- Extreme Gradient Boosting (XGB):
  An optimized gradient boosting framework integrated with regularization approaches and parallel processing capacities are identified as an XGBoost. The gradient boosting machines that has demonstrated its ability to exceed the computational capacity of boosted tree algorithms thus it becomes the extendible and the cutting edge application. The errors made by the existing models are adjusted by adding the ensemble technique named Boosting. Until the detection of no noticeable improvements the models are added recursively. Using an approach known as gradient boosting, new models are built to predict the residuals of earlier models, which are then combined together to get the final prediction. In order to minimize the loss while adding new models it applies gradient descent algorithms. Both the regression and classification are supported by this technique [29].
- Light Gradient Boosting (LGB):
  LGB is a next level gradient boosting algorithm that applies a leaf-wise algorithm for the growth of trees vertically. To split and grow the tree it selects the leaf that reduces the loss. For the identification of the best splits candidates LGB applies histogram based technique. To emphasize the significance of the data instances and enhance LGBM's training, a sampling algorithm known as Gradient-based One Side Sampling (GOSS) is utilized. Concentrating on the larger gradients in the data samples and disregarding those with smaller gradients is its primary function [30].

3.5.1 Majority Voting Ensemble Mechanism

In order to boost the efficiency of the introduced model one of the widely used fundamental ensemble model is voting. The two major styles of voting are the hard and soft voting. The most basic kind of voting is the plurality model, which allows every voter to select only one option which turns into their favorite choice. The candidates having the high number of vector are represented as the victor. The q becomes the winner if 25% of the voters select alternative p, 40% select alternative q and 35% alternative r. According to Condorcet's jury theorem, if M voters make a simple majority decision and each voter has a chance of selecting the correct option (C), then equation (21) represents the likelihood that the jury as a whole will select the right option.

$$c_X = \sum_{X}^{P=\left[\frac{X}{2}\right]} \frac{X!}{(X-P)!P!} . c^p . (1-c)^{X-P} \tag{21}$$

Therefore, if $c > 5.0, c_X > c$. This shows that the group will probably make better decisions than any one voter. The probability of the ensemble making the correct choice increases as more voters participate. When X tends to be infinite, $c_X \rightarrow 1$ [9] is ideal. By applying the single transferrable vote (STV) process each voter rates every choice in the preferred order. The totaling of all nodes are identified as quota q_T which is expressed in equation (22)

$$q_T = \left\lfloor \left(\frac{v_{TotalVotes}}{v_{DesiredWinners}+1} \right) \right\rfloor + 1 \tag{22}$$

The Droop quota is mostly typical. If a candidate receives the needed amount of vote or more, the candidate is deemed the winner. By applying equation (23) in which the amount of votes to be reallocated to the opposing candidate is v_{Tr}, the number of votes obtained from the second candidate becomes v_{2nd}, the total amount of votes for the current winners is indicated by $v_{TotalWinners}$ and the surplus votes for the winning candidate is indicated by v_{Surp}. Here from the same ballot the surplus vote is transferred to the chosen candidates:

$$v_{Tr} = \left(\frac{v_{2nd}}{v_{TotalWinners}} \right) . v_{Surp} \tag{23}$$

For transferring of votes there are several ways which are random selection, moving of strong candidates and others. This process is repeated until appropriate number of winners is obtained. The candidates having less amount of votes are just got eliminated. If the process become stuck the votes are transferred to others and none of the candidates meet the quota [22].

4. Experimental Setup

To evaluate the introduced cancer diagnosis approach it was trained by using BreakHis datasets [31] and it is executed by applying python tool. This approach is compared with the existing models metrics.

4.1 Dataset Description

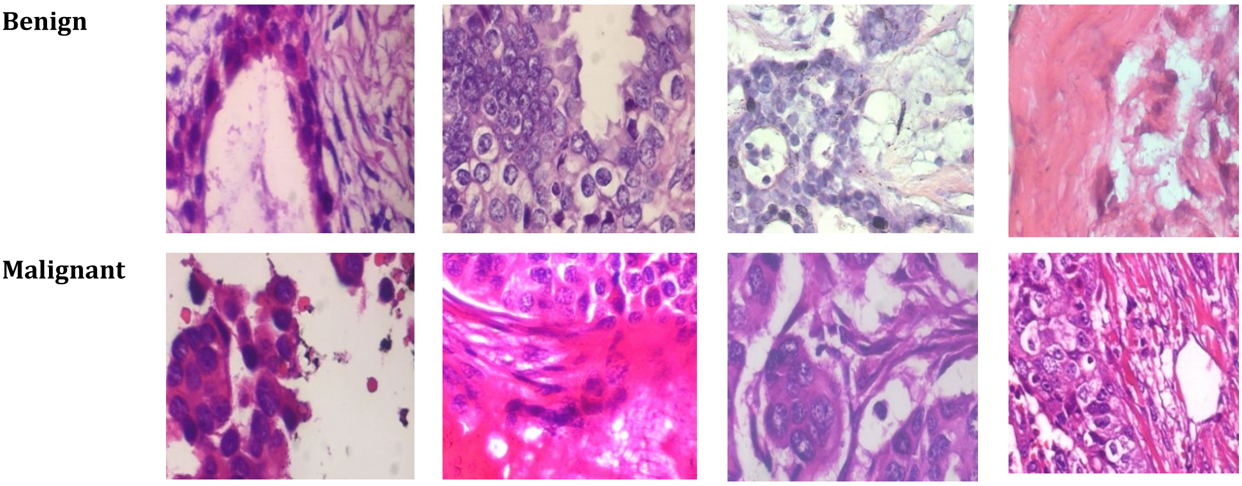


Fig. 2 *Samples Images of BreakHis dataset*

The introduced approach for cancer diagnosis using a cancer dataset to get the better result it applies ensemble classifier and also it compared with the developed model. The datasets used here is Breast Cancer Histopathological Image (BreakHis) dataset [31]. Fig. 2 shows the sample images of the BreakHis dataset. Here, 9,109 microscopic images of breast cancer tissue from 82 patients, magnified by various variables, make up this image by 40x, 100x, 200x, and 400x. This dataset is divided based on two they are the malignant and the benign tumors. From this it carries 5,429 malignant samples and 2,840 benign samples. Generally, the benign tumors grow slowly and remains localized. Malignant tumor can invade and destroy adjacent structures and spread around the areas leads to death. Based on the tumor cells viewed on the microscope the both the malignant and the benign tumors are classified into various types. For the benign breast cancer this dataset carries four histological subtypes they are Tubular Adenona (TA), Phyllodes Tumor (FT), Fibroadenoma (F), and Adenosis (A); also for the malignant breast cancer it has four subtypes they are Papillary Carcinoma (PC), Cucinous Carcinoma (MC), Lobular Carcinoma (LC), and Carcinoma (DC).

4.2 Performance Measures

The various performance metrics used in this approach are Accuracy, Recall, Precision, Sensitivity, Specificy, F score, Misclassification Rate, False Positive Ratio, False Negative Ratio and Mathews Correlation Coefficient.

4.2.1 Accuracy

Accuracy is the measure of the ratio between the sums of correct predictions to the total predictions. The mathematical expression for the accuracy is given in equation (24)

$$Acc = \frac{TPv+TNv}{TPv+TNv+FPv+FNv} \tag{24}$$

Where, False Negative Rate is indicated byFNv, FPv denotes the False Positive Rate, TNv indicates the True Negative rate, and TPv shows the True Positive rate.

4.2.2 Precision

Precision is a measure of ratio between the True positive rate and the sum of all positive. The mathematical expression for the precision is given in equation (25)

$$Pr = \frac{TPv}{TPv+FPv} \tag{25}$$

4.2.3 Sensitivity

Sensitivity is a measure of the ratio between the True Positive and the sum of False Negative and True Positive. The mathematical expression for the Sensitivity is given in equation (26)

$$Sen = \frac{TPv}{TPv+FNv} \tag{26}$$

4.2.4 Specificity

Specificity is a measure of the ratio between the True Negative and the sum of False Positive and True Negative. The mathematical expression for the Specificity is given in equation (27)

$$Spe = \frac{TNv}{TNv+FPv} \tag{27}$$

4.2.5 F Score

F score is a measure of ratio between True Positive to the sum of True positive and half the sum of False Negative and False Positive. The mathematical expression for the F Score is given in equation (28)

$$F\ sc = \frac{TPv}{TPv+\frac{1}{2}(FPv+FNv)} \tag{28}$$

4.2.6 Misclassification Rate

In any classification system the percentage of erroneous observations made are said to be the misclassification rate which is one of the ML metric. It is the ratio between the sum of False Negative and False Positive towards all the total Prediction. The mathematical expression for the Sensitivity is given in equation (29)

$$Mr = \frac{FPv+FNv}{TPv+TNv+FPv+FNv} \tag{29}$$

4.2.7 Mathew’s Correlation Coefficient

The MCC is a quality of binary classifications measure in the ML. The mathematical expression for the MCC is expressed below:

$$MCC = \frac{(TPv\times TNv)-(FPv\times FNv)}{\sqrt{(TPv+FPv)(TPv+FNv)(TNv+FPv)(TNv+FNv)}} \tag{30}$$

4.2.8 False Negative Ratio

The miss rate is said to be as the False Negative Rate in which the possibilities are that the true positives are missed by the Test. The mathematical expression for the False Negative Ratio is given in equation (31)

$$FNR = \frac{FNv}{TPv+FNv} \tag{31}$$

4.2.9 False Positive Ratio

False Positive Rate is calculated between the numbers of negative events and wrongly classified as false positive and total number of negative events. The mathematical expression for the True Negative Ratio is given in equation (32)

$$FPR = \frac{FPv}{TNv+FPv} \tag{32}$$

5. Result and Discussion

The overall outcome of this approach is evaluated between the devised models along with the performance metrics of the existing approaches they are LeNet CNN, ML, Whale Optimization (WO), DNN, and proposed Ensemble classifiers.

Table 1 *Performance of the classifiers*

	Accuracy	Precision	Recall	F1-score	Sensitivity	Specificity
LeNet CNN	86.53	83.68	84.76	87.01	89.64	83.94
ML	84.46	85.16	85.61	84.18	91.45	84.61
WO	88.08	89.97	90.23	90.56	91.19	80.05
DNN	93.86	94.27	95.85	94.37	94.82	82.13
Proposed	97.84	98.56	98.42	97.63	96.60	96.30

The various classifiers performance such as accuracy, F1-score, precision, recall, specificity and sensitivity is shown in Table 1. LeNet CNN, ML, WO, DNN, and Ensemble Learning are the five classifiers are taken in this experiment for classification and its graphical representation is shown in Fig. 4. Then the best feature selection technique Local Search ABHC is taken in this classification process. Then this proposed FS approach is combined with each classifier separately and compared with each other for identifying the best classifier. Therefore, the result of this experiment gives the Ensemble Learning as the better one, because its F1-score, recall, precision, accuracy, sensitivity and specificity giving the better performance compare to the other four classifiers. This proposed classifier gives the greatest classification accuracy that is shown in Fig. 3.

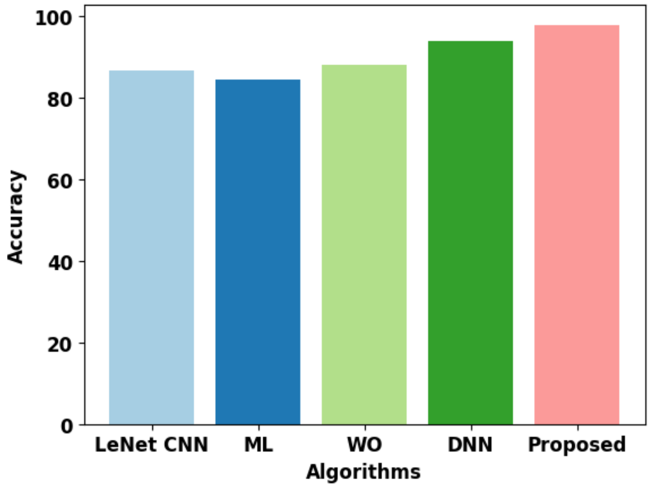


Fig. 3 *Accuracy of various classifiers*

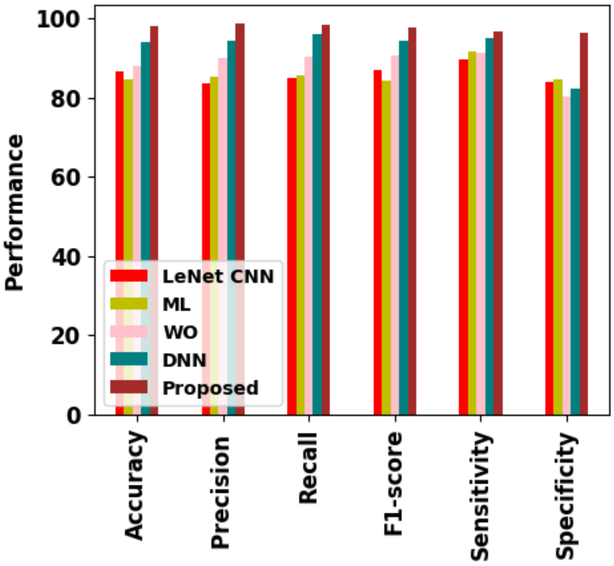


Fig. 4 *Performance of Accuracy, Precision, Recall, F1-score, Sensitivy, and specificity using various classifiers*

Also the remaining performances such as MCR, FNR, FPR, and MCC are shown in table 2 and its graphical representation is given in Fig. 5. The classifiers which are used for analysing the performance are LeNet CNN, ML, WO, DNN, and Ensemble Learning. These classifiers are combined with the better segmentation model of thresholding to get better performance outcomes. Therefore, the result of this experiment gives the Ensemble Learning as the better one, because its MCR, FNR, FPR, and MCC giving the better performance compared to the other four classifiers.

Table 2 *Performance of the classifiers*

	MCR	FNR	FPR	MCC
LeNet CNN	16.37	11.74	18.01	84.36
ML	16.54	9.81	27.44	86.59
WO	14.71	12.95	20.63	89.37
DNN	12.38	5.22	27.59	91.24
Proposed	3.42	4.89	5.02	95.62

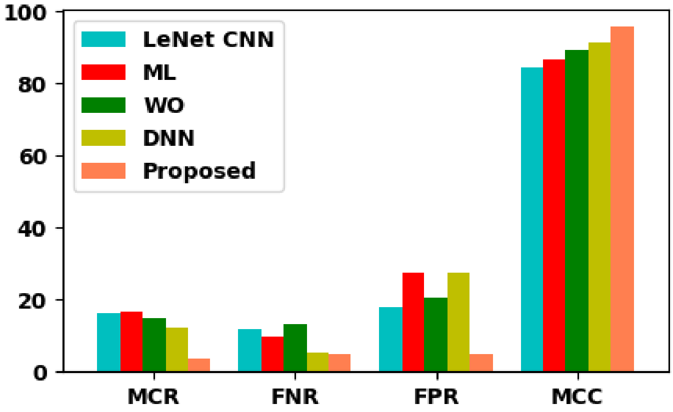


Fig. 5 *Performance of MCR, FNR, FPR, and MCC comparison with proposed and various other models*

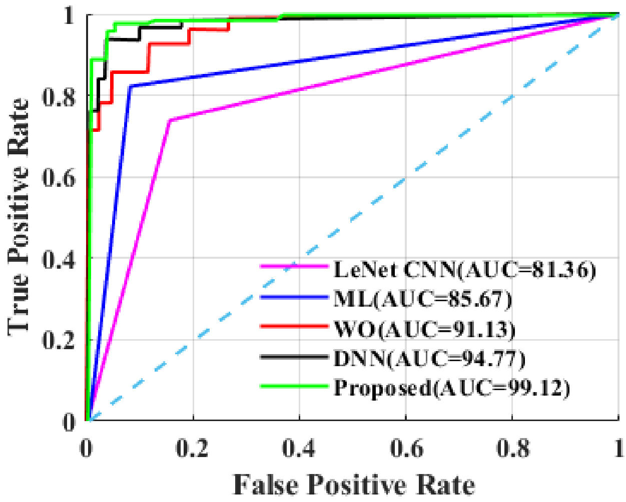
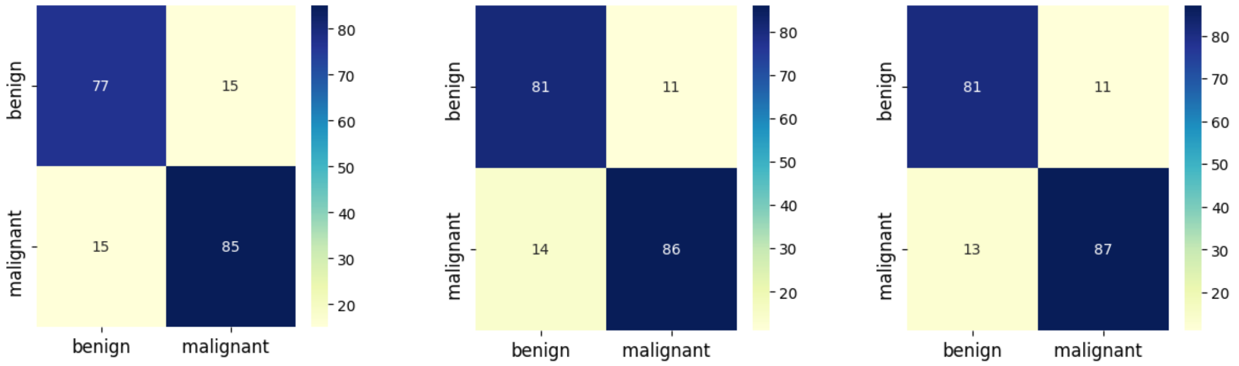
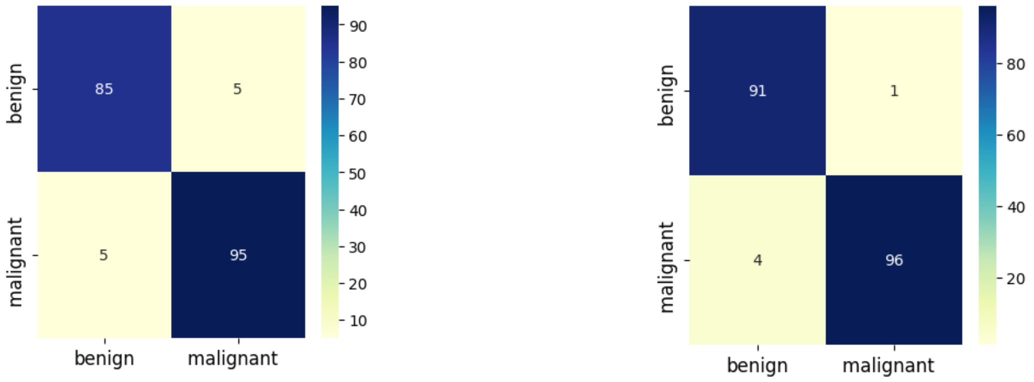


Fig. 6 *ROC curve for evaluating classifiers*

Fig. 6 shows the Receiver Operating Characteristic (ROC) curve of the proposed method, with an AUC value of 99.12 for the model. For the existing approaches the AUC values are 81.36 for LeNet CNN, 85.67 for ML, 91.13 for WO, and 94.77 for DNN. The ROC curves show the balance between a classifier's False Positive Rate and True Positive Rate. Classifiers with curves closer to the top-left corner have better performance. When the curve approaches the 45-degree diagonal in the ROC space, it indicates less accurate classification results.



(a) LeNet CNN **(b) ML** **(c) WO**



(d) DNN **(e) Proposed**

Fig. 7 *Confusion matrix for breast cancer detection*

The confusion matrices in Fig. 7 provide important information about how well different algorithms perform in detecting breast cancer, including their accuracy and rates of misclassification. A table that summarises a classifier model performance on a set of test data is called a confusion matrix. Frequently, it is employed to assess the efficiency of classification models, which attempt to forecast a categorical label for every input occurrence. Then the matrix shows how many false negatives (TN), false positives (FP), false negatives (FN) and true positives (TP) the model generated using the test data. The matrix represents the performance of a classification model that predicts whether a sample is benign or malignant. This figure shows the matrix plotted between the actual and predicted samples. The vertical axis represents the actual labels and the horizontal axis represents the predicted labels. From Fig. 7(a) shows the confusion matrix for LeNet CNN, it has approximately 85 of the samples categorised as malignant and 77 of the samples categorised as benign, whereas approximately 30 of the samples are misclassified overall. Fig. 7 (b) represents confusion matrix depend on the ML approach, where it clearly shows 81% of the samples are declared as the predicted benign cases, 86 of the samples are recognised as the predicted malignant cases and an overall of 25 samples are obtained as the misclassified errors. Then in Fig. 7 (c) demonstrates the confusion matrix according to the WO, in which 81 instances are obtained as benign condition and 87 as the malignant and having the overall misclassification rate of 24. Fig. 7 (d) DNN based confusion matrix in which 85 samples are correctly predicted benign and 95 samples are correctly predicted malignant and the total misclassification values is 10. For the proposed approach the confusion matrix is represented in Fig. 7(e) where 91 samples are the correctly predicted benign and the correctly predicted malignant is 96, thus the overall misclassification rate is 5. Therefore, compared with other models the proposed approach produces less misclassification rates. Overall, the confusion matrices given above helps to assess the model's accuracy and identify areas for improvement.

Table 3 *Comparisons with state of art breakhis dataset*

Reference	Model	Accuracy (%)	Year
[32]	DenseNet-201	98.30	2023
[33]	NASNet-A	99.4	2023
[34]	GARL-Net	97.78	2023
[35]	DenseNet121	99.6%	2023
[36]	Xception	89	2023
[37]	CSVM-H	88.71	2024
proposed		99.51	2024

The performance of the state of art is shown in the below Table 3. Additionally, the table displays the accuracy of different models in various years. Our focus is on studies that used the same datasets to ensure a fair comparison. The new model proposed outperforms other existing models with an accuracy of 99.51%.

6. Conclusion

This paper presents the novel approach depends on ensemble learning to effectively diagnose the breast cancer. The proposed model develops a novel approach which uses AGBF to enhance denoising of the cancer images. Additionally, the proposed approach implies thresholding towards the segmentation phase, also it applies the ranking based MRMR feature selection and the wrapper based Local Search ABHC model used for the effective feature selection model. Thus the proposed model effectively improves the classification accuracy based on identifying the majority voting based ensemble classifier. The performance of the introduced approach is evaluated depend on the evaluation measures namely precision, accuracy, sensitivity, specificity, Misclassification rate, MCC, False Negative Ratio, False Positive Ratio and F1-score. In all the performance measures the introduced model provides high value. It indicates that the devised approach can classify the breast cancer image more efficiently.

Furthermore, for future analysis, we can explore a quicker computation approach and look for a more versatile parameter configuration to achieve an improved optimal solution when addressing the denoising issue. Multi-level thresholds will be utilized for automatic segmentation to achieve the best possible outcomes. It needed to improve a weighted strategy for feature selection algorithms and apply other approaches for the ensemble FS, especially in the form of graph based models and online feature selection.

Acknowledgement

We would like to thank the Department of Computer Science, Annai Velankanni College, Manonmaniam Sundaranar University to support this work.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *M. Mohana Dhas, N. Suresh Singh;* ***data collection:*** *M. Mohana Dhas;* ***analysis and interpretation of results:*** *N. Suresh Singh;* ***draft manuscript preparation:*** *M. Mohana Dhas. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Iparraguirre-Villanueva, O., Epifanía-Huerta, A., Torres-Ceclén, C., Ruiz-Alvarado, J., & Cabanillas-Carbonel, M. (2023). Breast cancer prediction using machine learning models, *International Journal of Advanced Computer Science and Applications, 14(2)*, 610-620.

[2] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: a cancer journal for clinicians, 71(3)*, 209-249.

[3] Duffy, S. W., Tabár, L., Yen, A. M. F., Dean, P. B., Smith, R. A., Jonsson, H., Törnberg, S., Chen, S.L.S., Chiu, S.Y.H., Fann, J.C.Y., & Ku, M.M.S. (2020). Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women, *Cancer, 126(13)*, 2971-2979.

[4] Ekpo, E. U., Alakhras, M., & Brennan, P. (2018). Errors in mammography cannot be solved through technology alone, Asian Pacific journal of cancer prevention: APJCP, 19(2), 291.

[5] Prinzi, F., Insalaco, M., Orlando, A., Gaglio, S., & Vitabile, S. (2023). A YOLO-based model for breast cancer detection in mammograms, *Cognitive Computation, 16(1)*, 107-120.

[6] Sharmin, S., Ahammad, T., Talukder, M. A., & Ghose, P. (2023). A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection, *IEEE Access, 11*, 87694 - 87708.

[7] Ganesh, N., Jayalakshmi, S., Narayanan, R., Mahdal, M., Zawbaa, H. M., & Mohamed, A. W. (2023). Gated Deep Reinforcement Learning with Red Deer Optimization for Medical Image Classification, *IEEE Access, 11*, 58982 - 58993.

[8] Yang, X., Wang, R., Zhao, D., Yu, F., Heidari, A. A., Xu, Z., Chen, H., Algarni, A.D., Elmannai, H., & Xu, S. (2023). Multi-level threshold segmentation framework for breast cancer images using enhanced differential evolution, *Biomedical Signal Processing and Control, 80*, 104373.

[9] Pramanik, P., Mukhopadhyay, S., Mirjalili, S., & Sarkar, R. (2023). Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms, *Neural Computing and Applications, 35(7)*, 5479-5499.

[10] Thejas, G. S., Garg, R., Iyengar, S. S., Sunitha, N. R., Badrinath, P., & Chennupati, S. (2021). Metric and accuracy ranked feature inclusion: Hybrids of filter and wrapper feature selection approaches, *IEEE Access, 9*, 128687-128701.

[11] Liu, G., Ma, J., Hu, T., & Gao, X. (2022). A feature selection method with feature ranking using genetic programming, *Connection Science, 34(1)*, 1146-1168.

[12] Balasubramaniam, S., Velmurugan, Y., Jaganathan, D., & Dhanasekaran, S. (2023). A modified LeNet CNN for breast cancer diagnosis in ultrasound images, *Diagnostics, 13(17)*, 2746.

[13] Chen, J., Zhu, Z., Hu, H., Qiu, L., Zheng, Z., & Dong, L. (2023). A Novel Adaptive Group Sparse Representation Model Based on Infrared Image Denoising for Remote Sensing Application, *Applied Sciences, 13(9)*, 5749.

[14] Prathik, A., Anuradha, J., & Uma, K. (2022). A novel filter for removing image noise and improving the quality of image, *International Journal of Cloud Computing, 11(1)*, 14-26.

[15] Wang, Z., Yu, F., Wang, D., Liu, T., & Hu, R. (2023). Multi-threshold segmentation of breast cancer images based on improved dandelion optimization algorithm, *The Journal of Supercomputing, 80(3)*, 3849-3874.

[16] Jiménez, F., Sánchez, G., Palma, J., Miralles-Pechuán, L., & Botía, J. A. (2022). Multivariate feature ranking with high-dimensional data for classification tasks, *IEEE Access, 10*, 60421-60437.

[17] Hashemi, A., Dowlatshahi, M. B., & Nezamabadi-pour, H. (2023). Minimum redundancy maximum relevance ensemble feature selection: A bi-objective Pareto-based approach, *Journal of Soft Computing and Information Technology, 12(1)*, 20-28.

[18] Bugata, P., & Drotar, P. (2020). On some aspects of minimum redundancy maximum relevance feature selection, *Science China Information Sciences, 63(1)*, 112103.

[19] Chen, J., Cai, Z., Chen, H., Chen, X., Escorcia-Gutierrez, J., Mansour, R. F., & Ragab, M. (2023). Renal Pathology Images Segmentation Based on Improved Cuckoo Search with Diffusion Mechanism and Adaptive Beta-Hill Climbing, *Journal of Bionic Engineering, 20(5)*, 2240-2275.

[20] Ahmed, S., Ghosh, K. K., Garcia-Hernandez, L., Abraham, A., & Sarkar, R. (2021). Improved coral reefs optimization with adaptive β-hill climbing for feature selection, *Neural Computing and Applications, 33(12)*, 6467-6486.

[21] Shukla, S., Arya, K. V., Garg, B., & Lezama, P. (2022, December). Pneumonia Detection Using Gradient Boosting on Selected Features Extracted from DenseNet, *International Conference on Intelligent Technologies*, pp. 181-195.

[22] Panigrahi, A., Pati, A., Sahu, B., Das, M. N., Nayak, D. S. K., Sahoo, G., & Kant, S. (2023). En-MinWhale: An ensemble approach based on MRMR and Whale optimization for Cancer diagnosis, *IEEE Access, 11*, 113526 - 113542.

[23] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning, *Journal of Applied Science and Technology Trends, 2(01)*, 20-28.

[24] Rey-Blanco, D., Zofío, J. L., & González-Arias, J. (2024). Improving hedonic housing price models by integrating optimal accessibility indices into regression and random forest analyses, *Expert Systems with Applications, 235*, 121059.

[25] Razaa, A., Ahmada, A., Iqbala, Z., Yasina, Q., Javeda, H., Shaha, A., & Chaudharya, S. iAFP-ET: A robust approach for accurate identification of antifungal peptides using extra tree classifier and multi-view fusion, *Journal of Xi'an Shiyou University, Natural Science Edition, 19(6)*, 621-637.

[26] Abbas, Q., Albalawi, T. S., Perumal, G., & Celebi, M. E. (2023). Automatic Face Recognition System Using Deep Convolutional Mixer Architecture and AdaBoost Classifier, *Applied Sciences, 13(17)*, 9880.

[27] Ibrahim, M., AbdelRaouf, H., Amin, K. M., & Semary, N. (2023). Keystroke dynamics based user authentication using Histogram Gradient Boosting. IJCI, *International Journal of Computers and Information, 10(1)*, 36-53.

[28] Cahyana, N., Khomsah, S., & Aribowo, A. S. (2019, October). Improving imbalanced dataset classification using oversampling and gradient boosting, *5th International Conference on Science in Information Technology (ICSITech)*, Yogyakarta, Indonesia, pp. 217-222.

[29] Ogunleye, A., & Wang, Q. G. (2019). XGBoost model for chronic kidney disease diagnosis, *IEEE/ACM transactions on computational biology and bioinformatics, 17(6)*, 2131-2140.

[30] Alzamzami, F., Hoda, M., & El Saddik, A. (2020). Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation, *IEEE access, 8*, 101840-101858.

[31] The Breakhis dataset is available at "https://www.kaggle.com/datasets/ambarish/breakhis" accessed on February (2024).

[32] Zheng, Y., Li, C., Zhou, X., Chen, H., Xu, H., Li, Y., et al. (2023). Application of transfer learning and ensemble learning in image-level classification for breast histopathology, *Intelligent Medicine, 3(2)*, 115–128.

[33] Toma, T. A., Biswas, S., Miah, M. S., Alibakhshikenari, M., Virdee, B. S., Fernando, S., Rahman, M.H., Ali, S.M., Arpanaei, F., Hossain, M.A., Rahman, M.M., Niu, M. B., Parchin, N.O., & Livreri, P., (2023). Breast cancer detection based on simplified deep learning technique with histopathological image using BreaKHis database, *Radio Science, 58(11)*, 1-18.

[34] Patel, V., Chaurasia, V., Mahadeva, R., & Patole, S. P. (2023). GARL-Net: Graph Based Adaptive Regularized Learning Deep Network for Breast Cancer Classification, *IEEE Access, 11*, 9095-9112.

[35] Ashurov, A., Chelloug, S. A., Tselykh, A., Muthanna, M. S. A., Muthanna, A., & Al-Gaashani, M. S. (2023). Improved breast cancer classification through combining transfer learning and attention mechanism, *Life, 13(9)*, 1945.

[36] Sayın, İ., Soydaş, M. A., Mert, Y. E., Yarkadaş, A., Ergün, B., Yeh, S. S., & Üvet, H. (2023). Comparative Analysis of Deep Learning Architectures for Breast Cancer Diagnosis Using the BreaKHis Dataset. arXiv preprint arXiv:2309.01007.

[37] Karuppasamy, A., Abdesselam, A., Hedjam, R., & Al-Bahri, M. (2024). Feed-forward networks using logistic regression and support vector machine for whole-slide breast cancer histopathology image classification, *Intelligence-Based Medicine, 9*, 100126.