

A Novel Ensemble Bagging Classification Method for Breast Cancer Classification Using Machine Learning Techniques



Naga Deepti Ponnaganti*, Raju Anitha

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India

Corresponding Author Email: nagadeepthiponnaganti@sircrrengg.ac.in

<https://doi.org/10.18280/ts.390123>

ABSTRACT

Received: 22 December 2021

Accepted: 12 February 2022

Keywords:

breast cancer, ensemble bagging, weighted voting, ensemble bagging weighted voting classification

Breast cancer is observed as a dangerous disease type for women in the world. The clinical experts stated that early detection of cancer helps in saving lives. To detect cancer in the early stage, medical image processing is observed as an effective field. Medical Image processing with an appropriate classification mechanism improves accuracy and image resource with minimal processing time. To detect breast cancer several machine learning techniques are evolved for cancer classification. However, those machine learning techniques are subjected to increased time consumption and limitation in the accuracy of classification. This paper proposed an Ensemble Bagging Weighted Voting Classification (EBWvc) for the classification of breast cancer. Initially, to resolve to overfit in machine learning bagging is applied for collected data. The ensemble bagging classification provides effective training to machine learning for reduced computational time and improved performance characteristics. The weighted voting is adopted for the classification of cancer in the breast. The performance of proposed EBWvc is analyzed comparatively with consideration accuracy, precision, recall, and F1 -Score. The comparative analysis of results exhibited that proposed EBWvc exhibits improved performance than existing classification techniques.

1. INTRODUCTION

In recent years, cancer is stated as abnormal cell growth in which body cells are classified into a continuous manner and invade nearby tissues. The terms cancer is names based on the human body it originates [1]. Generally, cancer is classified as malignant and benign. In those, benign is stated as simple cysts which does not exhibit significant impacts on nearer tissues. Even benign is not a class of cancerous. Whereas malignant spread to other human bodies and grow such as bones and organs. Worldwide, breast cancer is the second leading cause of death among women [2]. To reduce the mortality rate caused by breast cancer, earlier detection and treatment are considered an important factor. To diagnosis breast cancer Mammography and Ultrasound (US) is utilized for screening and diagnosis of a breast cancer diagnosis. Recent studies stated that breast cancer diagnosis uses an imaging model using mammography [3].

Breast cancer starts in the lump of the breast with a diagnosis of cancer in women [4]. In western countries, analysis of the survey expressed that for every 11 women 1 are affected by breast cancer. The prolonged growth of breast cancer leads to change in the shape of the breast, skin dimpling, fluid in nipples, or red patch scaly skin. To increase the lifetime of breast cancer patients it is necessary to detect cancer at an early stage and provide appropriate treatment for increasing survival rate [1]. In breast cancer detection, X-ray mammography is considered a golden standard but it provides false negative and positive rates. However, this method is not appropriate for women with denser tissues in the breast [5].

The statistical model of mammographic texture provides spatial variation in the structure. However, this technique is not an effective method for the analysis of very low-speed analysis [6].

For high-risk analysis, MRI is considered an efficient method for cancer diagnosis. But with an effective determination of sensitivity and specificity increases the overall accuracy [7]. In the microwave imaging process, backscatter leads to image falsification [8]. On other hand, electrical impedence tomography leads to the increased cost of processing [9]. However, UltraSound imaging is also a complex factor due to the decomposition of data, which describes the speckle information. To explore the inner parts of the human body Ultrasound imaging is adopted through high-frequency waves, which is destructive and non-invasive technology. This implies that it does not impact on tested target without any discomfort and pain. Ultrasound waves transmitted from the sound waves return if any crack or impurity in the objects. The analysis of resultant echoes exhibited an analysis of various parameters [10]. The high-resolution detection process leads to reduced cost and increased flexibility with significant advantages of ultrasound imaging. Ultrasound Imaging (UT) provides an effective characterization of soft tissues. It is involved in the classification of Computer-Aided Design. The lesion segmentation act as a crucial role in the CAD system for feature computation for the analysis of lesion shape for estimation of classification accuracy [11].

Radial Gradient Index (RGI) filtering is involved in lesion detection of ultrasound image automatic classification. The

classification of mammogram images is sub-sampled with consideration of different factor 4. Initially, lesions of images are detected and segmented using Watershed segmentation. However, the Region of Interest is obtained is not correct and leads to the falsification of segmentation [12]. The particular method incorporates Edge detection with the identification of Region of Interest. The efficiency of the method is based on the utilization of edge detection. The tumour detection is based on the consideration of Automated Breast Ultrasound with an adaptive threshold approach for the detection of the tumour. The system overall efficiency is improved through the selection of threshold [13]. To diagnosis breast cancer several methods and algorithms are designed for classification such as Convolution Neural Network (CNN), Naive Bayes, Support Vector Machine (SVM), and so on. Those CNN and deep learning are utilized for object detection and classification [14].

This paper, proposed an EBWvc for breast cancer classification to eliminate overfitting and time consumption for cancer detection. Initially, bagging is adopted for reducing overfitting in collected mammogram images of the breast. The collected image data were applied through ensemble bagging for the training of images. Once, the training is completed weighted voting is incorporated for accurate classification of cancer. Simulation results stated that the proposed EBWvc exhibits improved classification accuracy, precision, recall, and F1-Score compared with the existing technique. The analysis of results expressed that proposed EBWvc performance is approximately improved than the existing technique value with approximate value of 5%, 10%, 6% and 10% respectively.

This paper is organized as follows: In section, II presented existing literature related to cancer classification. In section, III stated the overall methodology adopted for the proposed EBWvc scheme along with algorithm description. In section, IV provides simulation setup, results, and comparative analysis. Finally in section V provides the overall conclusion of the research.

2. RELATED WORK

This section presented literature related to the existing classification technique developed for breast cancer classification. To detect breast cancer for prediction and classification several machine learning algorithms were developed. The existing machine learning algorithm utilized for prediction and diagnosis of breast cancer are Artificial Neural Network (ANN), Naive Bayes, Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Convolutional Neural Network, etc. Several researchers concentrated on the diagnosis of breast cancer with consideration of mammogram datasets collected from several datasets such as the Wisconsin dataset, SEER dataset, etc. from several hospitals. With consideration of different mammogram datasets existing literature extract features and selects different features for detection and classification of breast cancer. These machine learning mechanisms are subjected to the limitation of limited accuracy and computational complexity. Evaluated ensemble machine learning tool for medical image filtering [15]. Also, the proposed ensemble machine learning mechanism incorporates BPNN for medical image processing for breast cancer classification. Through simulation analysis, the ensemble classification achieves an accuracy of 92%. Through the

utilization of the Wisconsin Breast cancer data set [16], the author developed a voting ensemble classification algorithm. The ensemble voting classification scheme achieves a classification accuracy of 93%. To improve classification accuracy [17], Qasem developed the Adaboost classifier. The evaluation of the Adaboost classifier in the classification of cancer achieves an accuracy of 88%.

To evaluate the breast cancer classification and classification, Alzubier and Chitrea [18] evaluated bagging SVM for classification of breast cancer. The bagging SVM achieves an accuracy of 90%. Aruna and Rajagopalan used Naive Bayes and SVM with the UCI database and achieved 68-79% accuracy [19]. Tsehay et al. developed a weakly supervised computer aided detection system that was used biopsy to learn data [20]. Nayak and Gope used Naive Bayes and SVM and got 88% accuracy [21].

Pena-Reyes and Sipper [22] developed a fuzzy-genetic approach for the classification of malignant and benign samples of cancer. Similarly, Chen and Hsu [23] presented a genetic algorithm (GA) for extraction based on a decision-making model with decision rules within the neural network (NN). Übeyli [24] comparatively examined the performance of different Neural network classifiers with Support Vector Machine (SVM). Also, with SVM feature selection better classification accuracy is achieved in the study of Akay [25]. Also, the resultant features of the neural network are classified with a reduction of data dimensions with the integration of association rules [26]. Zheng et al. [27], and Nilashi et al. [28] comparatively examined different classifiers with linear programming for radial basis function with the optimization approach. Further, weights are updated for recursive least square for substitute back propagation for classification.

Neural Network classification is conducted by Zhang et al. [29] using rough set feature estimation and integration of k-nearest neighbor classifier (KNN) for breast cancer diagnosis. Shahnaz et al. [30] developed a decision tree classifier for improving cancer prediction. The performance of the proposed classifier is comparatively examined with the existing approach for performance analysis. Also, SVM features [31] are integrated for the weighting factor for improving classification with normalization of min-max approach, and classification weights are evaluated using SVM. Similarly, Liu et al. [11] proposed an integrated approach with a combination of rough set and genetic algorithm utilization in Back propagation Neural Network (BPNN). The analysis is based on the consideration of the WDBC dataset for analysis. In the generation of features, Li et al. [32] developed a genetic programming algorithm for modified Fisher linear discriminant analysis for extraction of features with an estimation of minimal distances. The different classifier is comparatively examined those are SVM, Bayesian, and ANN. The analysis expressed that four classifier exhibits effective performance in the classifier.

The breast cancer classification is estimated for benign and malignant within the Artificial Immune System is presented for analysis stated in the study of Singh [33]. Mohammed [34] developed a k-means clustering and SVM classifier for estimation of tumor membership function estimation with recognition of hidden patterns separately. The SVM model is evaluated based on the calculation of the new membership feature estimation.

Ontiveros-Robles and Melin [35] developed a knowledge-based system for maximization of performance in terms of Regression Tree (CART), classification, clustering, and

principal component analysis for resolving multicollinearity. The classification performance is increased through subset feature selection with the integration of brainstorm optimization with the evaluation of the KNN classifier. Similarly, Patrício et al. [36] comparatively examined the different machine learning algorithms for effective classification. The comparative examination is performed for Logistics Regression, NN, Deep Convolutional NN, Random Forest (RF), and Naive Bayes (NB). The deep neural network is evaluated based on the classification function of the softmax approach in the study of Naik et al. [37] with the rectification of linear units. The experimental analysis expressed that data pre-processing with the z-score method is normalized. Also, SVM ensemble learning-based classification [38] is performed for analysis, The algorithm for the ensemble approach incorporates different kernel functions with an evaluation of SVM machines. The analysis leads to a classifier count of 12 in the count. Ting et al. [5] developed a hybrid feature selection subset to gain information with the integration of simulated annealing with GA for relevant feature identification. The performance is measured with consideration of features measured using KNN, SVM, and BPNN. The highest performance is observed in the BPNN algorithm. Also, Virmani and Agarwal [39] utilizes smooth group regularization of feedforward NN considered as the input layer. Additionally, the input nodes are involved in the reduction of input data dimensions for processing. Lu et al. [40] proposed a novel recurrent neural network (RNN) to handle the perturbed time-varying underdetermined linear system with double bound limits on residual errors and state variables. Beyond that, the bound-limited underdetermined linear system is converted into a time-varying system that consists of linear and nonlinear formulas through constructing a non-negative time-varying variable. Then, theoretical analyses are conducted to verify the superior convergence performance of the proposed RNN model.

3. OVERVIEW OF EBWVC FOR CANCER CLASSIFICATION

Datasets. The dataset for analysis is evaluated based on the UCI Wisconsin Diagnosis Breast Cancer dataset. The dataset is collected from the University of Wisconsin Hospitals, Madison. The collected dataset consists of 699 instances with the inclusion of fine-needle aspirates for the identification of breast tissues. To detect attributes are benign or malignant nine attributes are recorded for identification of data features. The missing values in the instances are discarded within the dataset with the inclusion of 683 cases contains 444 instances for identification of the benign type and 239 features are malignant.

Image acquisition. The analysis of mammogram images is collected with an analysis of mammographic image analysis of cancer database society. The mammogram images are classified as normal, benign, and malignant with the identification of breast fatty-glandular. Usually, images collected through mammograms are digitized with edge factors. Then, it has been reduced with a clipped size value of consideration of every image pixel size as $1024 * 1024$ pixels. The database incorporates 322 mammogram images in those 95 images that are considered for analysis. However, detection of cancer in the dense region of the breast is a challenging task in the mammogram process which affects the performance of

classification.

Pre-processing. The database consists of mammogram images that information about unwanted information and estimates the noise in the background. Initially, the pre-processing stage mammogram images are eliminated and make them suitable for processing. In the breast cancer region, noisy and unwanted noises are eliminated. The mammogram images provide information about the yellowish region in mammogram images and provide label and background about the mammogram images. Obtain the gradient image by image smoothing and differentiation, as other gradient-threshold related methods, then Split the gradient image into overlapped blocks and compute local thresholds for each local area and finally Perform edge labeling using obtained local thresholds. With gradient-based threshold estimation labels of images are eliminated from those input images. To generate the mask for input images several morphological feature operations are conducted for specified database classes for identification of benign, malignant, and normal. Through the achievement of mammogram images values proposed EBWvc classifier involved in the identification of those values. The proposed classifier is involved in the estimation of efficient classification efficiency with the reduction of generalized errors and bounds rather than mean square error estimation within the dataset.

This paper presented a cancer classification scheme for the identification of cancer in an earlier stage for diagnosis. The proposed EBWvc focused on the classification of cancer in the breast region with minimal processing time. To proposes a novel ensemble bagging classification method for breast tumour classification using a machine learning algorithm. The proposed system is introducing a simple and efficient approach to detect the cancer region in mammogram images. Our approach also segments the cancer region on the input mammogram image. Breast density is well known to be a risk factor in patients with suspected/known breast neoplasia. Extensive research in the field of qualitative and quantitative analysis on different tissue characteristics of the breast has rapidly become the chief focus of breast imaging. In Figure 1 overall architecture of the proposed EBWvc is presented.

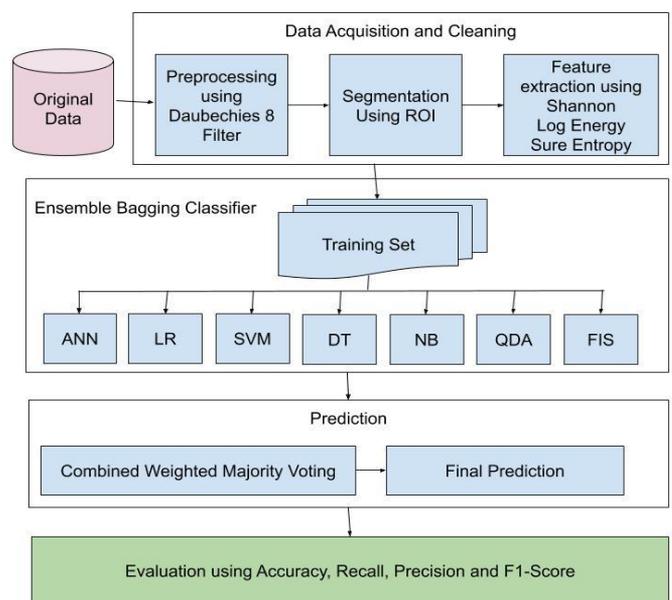


Figure 1. Overview of proposed EBWvc method

The proposed Ensemble Bagging, Weighted Voting, Ensemble Bagging Weighted Voting Classification (EBWvc) method will implement a series of steps on the input original data. Initially the data is subjected to pre processing using Daubechies 8 filter which enhances the quality of the image using wavelet transformation. The mammogram will be classified as a number of different regions using segmentation. The proposed method uses the Region of Interest (ROI) segmentation. The region of interest (ROI) is generally defined by using the pixel density values. With the segmentation using ROI, the ROI is selected as a rectangle in the region. After this the features are selected using Shanon long entropy and sure entropy methods.

3.1 Ensemble bagging classification method

In the segmentation, process features are extracted based on the segmentation region of interest. The ensemble bagging classification is based on the combination of different classifiers such as Bagging SVM, AdaBoost Classifier, Ensemble Machine learning, Random Forest (RF) in which the weighted voting ensemble mechanism is adopted for high-risk cancer prediction. Bagging approach involved in the prediction of multiple data version for acquiring aggregated cancer prediction.

The main steps in bagging concept are presented as follows:

1. Building learning algorithm origin and estimation of training data set.
2. The single origin learning algorithm is not higher for dataset.
3. The learning algorithm repeated based on numerous times for predictive function sequence and voting.
4. Through multiple models output results are increased with respect to accuracy.

The bagging advantage is based on the minimization of overfitting for increasing the classification accuracy rate. With classification involved in multiple fusion of sensor data, the classification relies on SVM does not imply an incremental learning sample for the data stream. It utilizes a new sample for adjusting the single classifier parameters to the internal structure. In the case of larger datasets, consumers focused on the consumption of a large amount of time and resources based on internal structure parameter adjustment. However, this is prone to the problem of ease of adaptability. To solve these issues learner accuracy diversity is improved for learners and improves the classifier generalization performance through ensemble bagging approach.

Ensemble learning scheme utilization of multiple training datasets with consideration of different subsets. The data subset for each base classifier is defined through a combination of the base classifier with the implementation of a new ensemble classifier. In the case of incremental data, the ensemble bagging algorithm is involved in the extraction of the training set. In involved in the construction of ensemble classifier for analysis of classifier reflect emerging information for the creation of different dataset from the sample set. Based on the set of subsample learning is classifier is evaluated through majority vote method learning for implementation of an ensemble learning algorithm on an incremental basis. In the ensemble learning process, solution is achieved based on the consideration of the following factors those are presented as follows:

The aggregator is represented as S and self-sampling is obtained through Tth round aggregator sample S. Within the

aggregator Tth subset S_t is extracted for aggregator value S within the range (t=1, 2,...,T). Every sample subset S_k incorporates Nth sample. In base classifier new training sample for learning is stated as S_t using bagging algorithm. The weak classifier each subset S_t denoted as φ(x, S_k) along with estimation of weak classifier error rate φ(x, S_k) and presented in Eq. (1) as follows:

$$\varepsilon_t = \sum_{(x, y_t)} [\phi_t(x, S_k) \neq y_t] / |S_t| \quad (1)$$

The training set distribution S_t is independently extracted with estimation of tth weak classifier combined with φ(x, S) represented as strong classifier. Based on this, final decision is obtained when processing with strong classifier φ(x, S) for achieving tth weak classifier value of φ(x, S_k). In majority of voting scheme majority of voting approaches are categorized based on test sample.

The steps of ensemble bagging with weighted Voting algorithm 1 are as follows:

Algorithm 1: Ensemble Bagging

Input:

Training Set $S = \{(x, y)\}; j = 1, 2, \dots, m$

Learning rate L

Number of Classifiers T

for I=1, 2, ..., T

Extract m-th sample from S

Lear L from S_k: $N = L(S_k)$

Combining classifier $N(x) = \arg \max_{y \neq x} \sum_{y \in x} 1$

end for

Output: Ensemble N(x)

The comparative analysis of predictive result performance is based on the consideration of classifier. The prediction classification is based on the sample set for training process.

3.2 Weighted majority voting

In ensemble classification performance majority-based voting scheme is broadly utilized. The proposed method automatically generates the hyperplane to classify the expected results of either benign or malignant. General representation for weighted voting is given in the equation. In majority voting, the class label y is predicted via majority (plurality) voting of each classifier C is given in Eq. (2).

$$y = \text{mode}\{C_1(x), C_2(x), \dots, C_n(x)\} \quad (2)$$

(q: w1, w2, w3)

where, w1, w2, w3 are weightages for classifiers.

The first step is to generate training dataset D_i (i=1, 2, ..., k) from D. Next process step is to develop an individual classifier C_i from each training dataset D_i. At last, the labeled class can be received by considering the results of the weighted majority voting mechanism through a vote of the individual predictions. For instance, x image data is classified with the integration of individual classifier prediction. Each instance of the class is performed based on assigned labels in weighted Majority voting (WMV) with an ensemble mechanism for sorting of classes. The WMV ensemble mechanism is generally denoted as Plurality Vote (PV) approach. Most often, the WMV

mechanism is applied for equating the performance of various models. Mathematically weighted voting is represented in Eq. (3).

$$class(x) = \arg \max_{c_i \in dom(y)} \left(\sum_k g(y_k(x), C_i) \right) \quad (3)$$

where, the classification of the K^{th} classifier is denoted as $y_k(x)$ and $g(y, c)$ represents about the index function which can demonstrated as in Eq. (4).

$$g(y, c) = \begin{cases} 1 & y = c \\ 0 & y \neq c \end{cases} \quad (4)$$

If the probabilistic classifier is utilized, the crisp classification $y_k(x)$ is got from the following Eq. (5).

$$class(x) = \arg \max_{c_i \in dom(y)} \hat{P}M_k(y = c_i | x) \quad (5)$$

where, M_k is applied to demonstrate the classifier k and $\hat{P}M_k(y = c_i | x)$ represents about the probability of class c for an instance x .

4. RESULTS AND DISCUSSION

The entire implementation of the proposed EBWvc is done in the MATLAB tool and the configurations considered for the experimentation are: PC with Windows 10 OS, 4GB RAM, and Intel I3 processor.

4.1 Database description

The breast cancer image samples for experimental analysis of the proposed EBWvc are taken from the Wisconsin Diagnosis Breast Cancer dataset. The Wisconsin Diagnosis Breast Cancer dataset has a large collection of a cropped section of normal and cancer cells.

4.2 Performance metrics

The performance of the proposed EBWvc is evaluated based on metrics, such as accuracy, precision, Recall and F1-score.

Accuracy: Accuracy is stated as identification of closeness detection made for classification represented in Eq. (6).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

where, TP represents true positive value, TN denoted true negative, FN stated false negative, and FP denoted false positive.

Recall: It provides the identification of correct positive samples through the proposed EBWvc as stated in Eq. (7) as follows:

$$Recall = \frac{TP}{TN + FN} \quad (7)$$

Precision: It measures EBWvc correctly identified instances from the retrieved instance and it is presented in Eq. (8).

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

F1-Score: It combines precision and recall value of proposed EBWvc to determine efficiency of proposed EBWvc and it is defined in Eq. (9) as follows.

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (9)$$

4.3 Comparative methods

For the comparative analysis, this work considers the following works: Support Vector Machine (SVM) [21], Random Forest (RF) [20], Bagging - SVM [19], Adaboost Classifier [18] and Ensemble Machine learning [16]. The above mentioned techniques are comparatively examined with proposed EBWvc.

4.4 Simulation results

In this research to evaluate the performance of the proposed EBWvc approach histogram processing of images is evaluated. The images collected from a mammogram are evaluated through simulation in MATLAB. Input images are processed using an enhanced approach and intensity values are evaluated using EBWvc. By the use of weighted voting, the image edge values are identified. The image processing adopted in this research for processing is presented and elaborated in this section. In Figure 2 image pre-processed in EBWvc is presented.

Figure 2 illustrates the preprocessed image with EBWvc limit conditions. For the processed image the image gets smoothed and processed accordingly. A comparison of the original image with the processed image illustrates that our preprocessing mechanism effectively smoothens mammogram images. This pre-processed image is applied as input for the EBWvc mechanism with varying pixel intensity values. Based on image pixel intensity different points of the image are extracted and segmented for a clear dataset identification process. EBWvc approach evaluates the input pixel based on the intensity of the image pixels. The comparative analysis of segmentation automatic classification is developed for the proposed EBWvc. The automatic classification is based on a peer review of mammogram image oncology. Through analysis cancer, 2D slices are estimated based on clinical practices of the gold standard. In Figure 3 cancer segmented region for proposed EBWvc is presented.

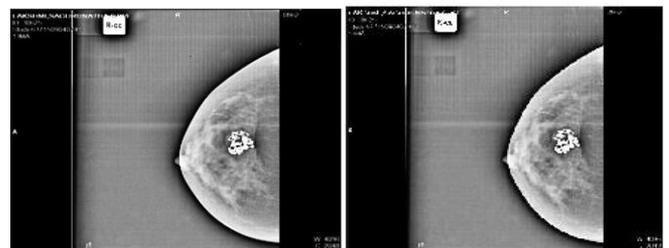


Figure 2. Before pre processed image and after pre processed image

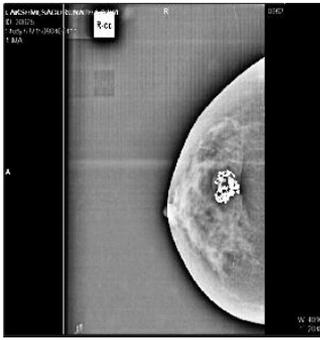


Figure 3. Image segmented with EBWvc

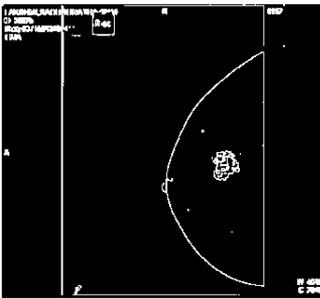


Figure 4. Tumor extracted with EBWvc

Usually, segmentation of breast cancer is considered a critical factor due to the existence of different numbers of artifacts leads to increased non-cancer regions such as glands and blood vessels. Also, geometric variability is evaluated for a varying number of factors involved in the processing of segmentation. The proposed EBWvc proposed in this paper does not require any manual intervention for data processing. Even the proposed scheme does not consider any consumption size or shape for cancer regions. The significant segmentation of cancer is achieved through geometric viability with the determination of spline distance involved in the estimation of complex bid field (IHO) utilized for slicing mammogram images. In Figure 4 extracted cancer image using EBWvc is illustrated.

The identification of tumor in the breast region with Weighted voting based technique involved in an approximation of cancer position for reducing processing steps with the elimination of cancer tissues. The comparative analysis of results expressed that the classification of proposed EBWvc provides improved classification performance. The continued and spatial smoothness of the breast cancer region is involved in the estimation of the level set for the detection of the cancer region. Quantitative analysis with Gold standard involved in the estimation of axial slices without any significant performance variation.

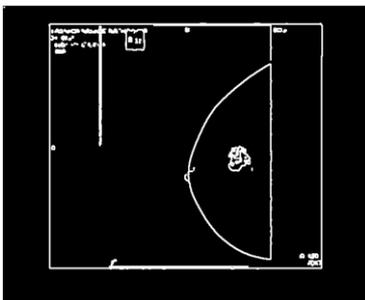


Figure 5. Image with straighten image condition

Figure 5 presented about cancer region extracted and classified for mammogram images for reproducibility for segmentation of classification with Intra and inter variability segmentation of results. In the defined framework the parameter values are unaltered and results are obtained with repeated times with indication of reproducibility. The mammogram images are incorporated with modality increases the computational efficiency with a reduced processing time of slices. In this scenario, RTP is assisted with the estimation of cancer boundaries with the estimation of unbiased and time-effective factors. The analysis is based on the estimation of the automation process for segmentation of tumour regions. The system increases the robust performance with the provision of different protocols and consideration of different models.

In Table 1 proposed EBWvc approach classification parameters are presented in terms of benign and normal regions of breast. The parameters considered for analysis are accuracy, precision, Recall and F1-Score. The analysis is based on consideration of tumour with three categories such as Benign, Normal and Average. In benign stage accuracy is observed as 96%, precision as 92%, recall value of 88% and F1-Score as 91%. In case of normal stage accuracy is observed as 94%, precision as 88%, recall value of 92% and F1-Score as 93% and average accuracy is observed as 95%, precision as 90%, recall value of 90% and F1-Score as 92% for EBWvc.

In Figure 6 proposed EBWvc performance measured for cancer classification is presented.

From analysis of results obtained for proposed EBWvc it is observed that performance is significantly improved. The analysis of results expressed that proposed EBWvc exhibits significant performance in classification with consideration of three categories of tumour such as benign, normal and average. In Table 2 presented about overall parameters measured for proposed EBWvc is presented along with exiting techniques.

Table 1. Performance measurement of EBWvc

Methods	Accuracy	Precision	Recall	F1-Score
Benign	96	92	88	91
Normal	94	88	92	93
Average	95	90	90	92

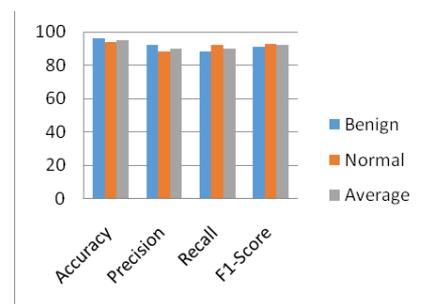


Figure 6. Performance measurement of EBWvc

In Figure 6 proposed EBWvc performance measured for cancer classification is presented.

From analysis of results obtained for proposed EBWvc it is observed that performance is significantly improved. The analysis of results expressed that proposed EBWvc exhibits significant performance in classification with consideration of three categories of tumour such as benign, normal and average. In Table 2 presented about overall parameters measured for proposed EBWvc is presented along with exiting techniques.

In Figure 7 the overall comparison of proposed EBWvc

with existing classifiers is presented. The accuracy measured for proposed EBWvc is measured as 95% which is significantly higher than existing techniques. The accuracy measured for an existing technology such as Ensemble Machine Learning, Bagging – SVM, SVM, AdaBoost Classifier and RF is measured as 92%, 90%, 88%, 88%, and 87% respectively. The measurement of precision provides 87%, 86%, 84%, 83%, 81%, and 90% for Ensemble Machine Learning, Bagging – SVM, SVM, AdaBoost Classifier, RF and EBWvc respectively. The analysis expressed that the proposed EBWvc provides improved performance rather than existing techniques. The recall measurement expressed that 87%, 84%, 86%, 86%, 82%, and 90% for Ensemble Machine Learning, Bagging – SVM, SVM, AdaBoost Classifier, RF and proposed EBWvc respectively. The comparative analysis expressed that the proposed EBWvc provides improved performance than the existing classification technique. Similarly, F1-Score is measured as 88%, 82%, 84%, 90%, 86% and 92% for Ensemble Machine Learning, Bagging – SVM, SVM, AdaBoost Classifier, RF and proposed EBWvc respectively. Through analysis, it is concluded that the proposed EBWvc exhibits improved performance rather than the existing classification technique.

The comparative analysis of proposed EBWvc with existing classifiers expressed that EBWvc presents higher performance. The performance measurement of EBWvc is higher than the existing classifier in terms of accuracy, precision, recall, and F1-score. The performance measurement stated that the proposed EBWvc exhibits approximately 10% improved performance in terms of accuracy than Ensemble Machine Learning, Bagging – SVM, SVM, AdaBoost Classifier, RF. The precision value measured is approximately 5% higher than the Ensemble Machine Learning, Bagging – SVM, SVM, AdaBoost Classifier. Also, the analysis of Recall and F1-Score 7% and 10% higher than Ensemble Machine Learning, Bagging – SVM, SVM, AdaBoost Classifier, RF and Ensemble Machine Learning, SVM, AdaBoost Classifier, RF respectively.

Table 2. Overall comparative analysis

Methods	Accuracy	Precision	Recall	F1-Score
Ensemble Machine Learning	92	87	87	88
Bagging – SVM	90	86	84	82
SVM	88	84	86	84
AdaBoost Classifier	88	83	86	90
RF	87	81	82	86
Proposed EBWvc	95	90	90	92

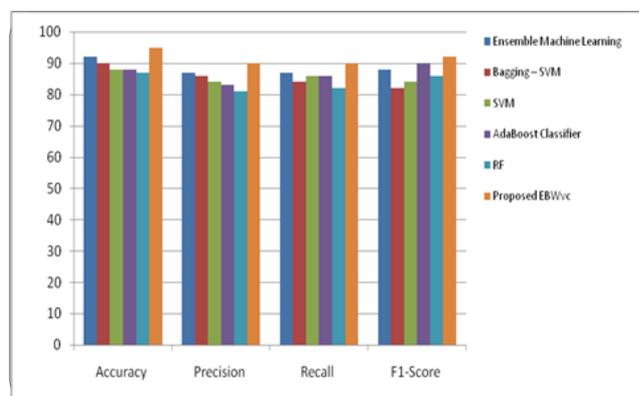


Figure 7. Overall comparison of classifiers

5. CONCLUSION

An image improvement technique is developing for earlier cancer detection and diagnosis. This paper attempts to solve the problem of the classification of breast cancer using a machine learning algorithm. This paper presented an EBWvc for the classification of breast cancer based on benign and normal. The proposed EBWvc was involved in the identification of cancer parts for diagnosis. Initially, ensemble bagging is applied for the optimization of breast dataset for improving classification accuracy. In the second stage, weighted voting is adopted for the classification of cancer. The analysis of results demonstrated that the proposed EBWvc exhibits an improved classification rate rather than the existing classification technique. In the future, the proposed classification technique can be improved through medical image fusion for the diagnosis of several diseases in human.

REFERENCES

- [1] Wang, P., Song, Q., Li, Y., Lv, S., Wang, J., Li, L., Zhang, H. (2020). Cross-task extreme learning machine for breast cancer image classification with deep convolutional features. *Biomedical Signal Processing and Control*, 57: 101789. <https://doi.org/10.1016/j.bspc.2019.101789>
- [2] Boumaraf, S., Liu, X., Zheng, Z., Ma, X., Ferkous, C. (2021). A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. *Biomedical Signal Processing and Control*, 63: 102192. <https://doi.org/10.1016/j.bspc.2020.102192>
- [3] Xia, L., Yao, Y., Dong, Y., Wang, M., Ma, H., Ma, L. (2020). Mueller polarimetric microscopic images analysis based classification of breast cancer cells. *Optics Communications*, 475: 126194. <https://doi.org/10.1016/j.optcom.2020.126194>
- [4] El-Bendary, N., Belal, N.A. (2020). A feature-fusion framework of clinical, genomics, and histopathological data for METABRIC breast cancer subtype classification. *Applied Soft Computing*, 91: 106238. <https://doi.org/10.1016/j.asoc.2020.106238>
- [5] Ting, F.F., Tan, Y.J., Sim, K.S. (2019). Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications*, 120: 103-115. <https://doi.org/10.1016/j.eswa.2018.11.008>
- [6] Fang, Y., Zhao, J., Hu, L., Ying, X., Pan, Y., Wang, X. (2019). Image classification toward breast cancer using deeply-learned quality features. *Journal of Visual Communication and Image Representation*, 64: 102609. <https://doi.org/10.1016/j.jvcir.2019.102609>
- [7] Conti, A., Duggento, A., Indovina, I., Guerrisi, M., Toschi, N. (2021). Radiomics in breast cancer classification and prediction. *Seminars in Cancer Biology*, 2: 238-250. <https://doi.org/10.1016/j.semcancer.2020.04.002>
- [8] Khan, S., Islam, N., Jan, Z., Din, I.U., Rodrigues, J.J.C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, 125: 1-6. <https://doi.org/10.1016/j.patrec.2019.03.022>
- [9] Bouchal, P., Schubert, O.T., Faktor, J., et al. (2019). Breast cancer classification based on proteotypes

- obtained by SWATH mass spectrometry. *Cell Reports*, 28(3): 832-843. <https://doi.org/10.1016/j.celrep.2019.06.046>
- [10] Agrawal, U., Soria, D., Wagner, C., et al. (2019). Combining clustering and classification ensembles: A novel pipeline to identify breast cancer profiles. *Artificial Intelligence in Medicine*, 97: 27-37. <https://doi.org/10.1016/j.artmed.2019.05.002>
- [11] Liu, N., Qi, E.S., Xu, M., Gao, B., Liu, G.Q. (2019). A novel intelligent classification model for breast cancer diagnosis. *Information Processing & Management*, 56(3): 609-623. <https://doi.org/10.1016/j.ipm.2018.10.014>
- [12] Kumar, A., Singh, S.K., Saxena, S., et al. (2020). Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Information Sciences*, 508: 405-421. <https://doi.org/10.1016/j.ins.2019.08.072>
- [13] Dora, L., Agrawal, S., Panda, R., Abraham, A. (2017). Optimal breast cancer classification using Gauss - Newton representation based algorithm. *Expert Systems with Applications*, 85: 134-145. <https://doi.org/10.1016/j.eswa.2017.05.035>
- [14] Vijayarajeswari, R., Parthasarathy, P., Vivekanandan, S., Basha, A.A. (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement*, 146: 800-805. <https://doi.org/10.1016/j.measurement.2019.05.083>
- [15] Khuriwal, N., Mishra, N. (2018). Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In 2018 IEEMA Engineer Infinite Conference (eTechNxT), New Delhi, India, pp. 1-5. <https://doi.org/10.1109/ETECHNXT.2018.8385355>
- [16] Deng, C., Perkowski, M. (2015). A novel weighted hierarchical adaptive voting ensemble machine learning method for breast cancer detection. In 2015 IEEE International Symposium on Multiple-Valued Logic, Waterloo, ON, Canada, pp. 115-120. <https://doi.org/10.1109/ISMVL.2015.27>
- [17] Qasem, A., Abdullah, S.N.H.S., Sahran, S., Wook, T.S. M.T., Hussain, R.I., Abdullah, N., Ismail, F. (2014). Breast cancer mass localization based on machine learning. In 2014 IEEE 10th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, Malaysia, pp. 31-36. <https://doi.org/10.1109/CSPA.2014.6805715>
- [18] Alzubier, M.H., Chittraa, V. (2019). Classification of breast cancer using machine learning techniques. *International Journal of Research and Analytical*, 58.
- [19] Aruna, S., Rajagopalan, S.P. (2011). A novel SVM based CSSFFS feature selection algorithm for detecting breast cancer. *International Journal of Computer Applications*, 31(8): 14-20.
- [20] Tsehay, Y.K., Lay, N.S., Roth, H.R., Wang, X., Kwak, J.T., Turkbey, B.I., Pinto, P.A., Wood, B.J., Summers, R.M. (2017). Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images. *Medical Imaging 2017: Computer-Aided Diagnosis*, 10134: 20-30. <https://doi.org/10.1117/12.2254423>
- [21] Nayak, S., Gope, D. (2017). Comparison of supervised learning algorithms for RF-based breast cancer detection. In 2017 Computing and Electromagnetics International Workshop (CEM), Barcelona, Spain, pp. 13-14. <https://doi.org/10.1109/CEM.2017.7991863>
- [22] Pena-Reyes, C.A., Sipper, M. (1999). A fuzzy-genetic approach to breast cancer diagnosis. *Artificial Intelligence in Medicine*, 17(2): 131-155. [https://doi.org/10.1016/S0933-3657\(99\)00019-6](https://doi.org/10.1016/S0933-3657(99)00019-6)
- [23] Chen, T.C., Hsu, T.C. (2016). A GAs based approach for mining breast cancer pattern. *Expert Systems with Applications*, 30(4): 674-681. [https://doi.org/10.1016/S0933-3657\(99\)00019-6](https://doi.org/10.1016/S0933-3657(99)00019-6)
- [24] Übeyli, E.D. (2007). Implementing automated diagnostic systems for breast cancer detection. *Expert Systems with Applications*, 33(4): 1054-1062. <https://doi.org/10.1016/j.eswa.2006.08.005>
- [25] Akay, M.F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, 36(2): 3240-3247. <https://doi.org/10.1016/j.eswa.2008.01.009>
- [26] Karabatak, M., Ince, M.C. (2009). An expert system for detection of breast cancer based on association rules and neural network. *Expert systems with Applications*, 36(2): 3465-3469. <https://doi.org/10.1016/j.eswa.2008.02.064>
- [27] Zheng, B., Yoon, S.W., Lam, S.S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications*, 41(4): 1476-1482. <https://doi.org/10.1016/j.eswa.2013.08.044>
- [28] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics*, 34(4): 133-144. <https://doi.org/10.1016/j.tele.2017.01.007>
- [29] Zhang, X.T., Zhang, Y., Gao, H.R., He, C.L. (2018). A wrapper feature selection algorithm based on brain storm optimization. In *International Conference on Bio-Inspired Computing: Theories and Applications*, Beijing, China, pp. 308-315. https://doi.org/10.1007/978-981-13-2829-9_28
- [30] Shahnaz, C., Hossain, J., Fattah, S.A., Ghosh, S., Khan, A.I. (201). Efficient approaches for accuracy improvement of breast cancer classification using wisconsin database. In 2017 IEEE Region 10 humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, pp. 792-797. <https://doi.org/10.1109/R10-HTC.2017.8289075>
- [31] Wang, H., Zheng, B., Yoon, S.W., Ko, H.S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2): 687-699. <https://doi.org/10.1016/j.ejor.2017.12.001>
- [32] Li, F., Zurada, J.M., Wu, W. (2018). Smooth group L1/2 regularization for input layer of feedforward neural networks. *Neurocomputing*, 314: 109-119. <https://doi.org/10.1016/j.neucom.2018.06.046>
- [33] Singh, B.K. (2019). Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: a comparative investigation in machine learning paradigm. *Biocybernetics and Biomedical Engineering*, 39(2): 393-409. <https://doi.org/10.1016/j.bbe.2019.03.001>
- [34] Mohammed, D.E.S.B. (2019). A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection. *Expert Systems with Applications*, 139: 112824. <https://doi.org/10.1016/j.eswa.2019.112824>
- [35] Ontiveros-Robles, E., Melin, P. (2019). A hybrid design of shadowed type-2 fuzzy inference systems applied in

- diagnosis problems. *Engineering Applications of Artificial Intelligence*, 86: 43-55. <https://doi.org/10.1016/j.engappai.2019.08.017>
- [36] Patrício, M., Pereira, J., Crisóstomo, J., Matafome, P., Gomes, M., Seica, R., Caramelo, F. (2018). Using Resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer*, 18(1): 29. <https://doi.org/10.1186/s12885-017-3877-1>
- [37] Naik, A.K., Kuppili, V., Edla, D.R. (2020). Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function. *Soft Computing*, 24(6): 4575-4587. <https://doi.org/10.1007/s00500-019-04218-6>
- [38] Rao, H., Shi, X., Rodrigue, A.K., et al. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing*, 74: 634-642. <https://doi.org/10.1016/j.asoc.2018.10.036>
- [39] Virmani, J., Agarwal, R. (2019). Effect of despeckle filtering on classification of breast tumors using ultrasound images. *Biocybernetics and Biomedical Engineering*, 39(2): 536-560. <https://doi.org/10.1016/j.bbe.2019.02.004>
- [40] Lu, H., Jin, L., Luo, X., Liao, B., Guo, D., Xiao, L. (2019). RNN for solving perturbed time-varying underdetermined linear system with double bound limits on residual errors and state variables. *IEEE Transactions on Industrial Informatics*, 15(11): 5931-5942. <https://doi.org/10.1109/TII.2019.2909142>