An Effective of Ensemble Boosting Learning Method for Breast Cancer Virtual Screening Using Neural Network Model

AHMED HAMZA OSMAN AND HANI MOETQUE ABDULLAH ALJAHDALI

Department of Information System, Faculty of Computing and Information Technology, King Abdulaziz University–Rabigh, Rabigh 21589, Saudi Arabia

Corresponding author: Ahmed Hamza Osman (ahoahmad@kau.edu.sa)

ABSTRACT Radial Based Function Neural Network models (RBFNN) are currently used deep-rooted methods for assessing the stages of diagnosis of chronic diseases. The goals of this research are to suggest a model for the diagnosis of breast cancer, and to be able to estimate the stages of development of pre-malignant breast tumors. The significance of the study is to develop an integrated RBF neural network with ensemble features using the boosting method. The importance of the ensemble boosting method is to generate a sequence of models to achieve more precise predictions. One of the ensemble boosting advantages is that it can take longer to build and to score than a RBF NN standard model. By using ensemble boosting, the accuracy of breast tumor diagnosis increased and thus it became easier to know the stage of the tumor, and whether it was malignant or benign. This will help doctors to select appropriate treatment for each tumor stage, consequently leading to the salvage of cancer patients with this type of tumor. The suggested RBFNN method was examined on the different type of UCI breast cancer datasets. The general diagnosis accuracy based on 10-fold cross validation using RBFNN method obtained 97.4%, 98.4%, 97.7% and 97.0% for the WBC, BCD, BCP, and WBCD UCI datasets respectively. The effectiveness of the proposed method was confirmed by comparing accuracy improvement both before and after using ensemble boosting, and it was found to be more accurate compared with other breast cancer diagnosis methods such as Logistic Regression (91.5%), KNN (96%), SVM (89%), Decision tree (95.13), CNN (97.66%), and Naive Bayes (91%).

INDEX TERMS Cancer disease, ensemble boosting, prediction, radial based function, neural network.

I. INTRODUCTION

Breast cancer is considered one of the common categories of cancer in women, and it is the second most common cancer causing cancer among lung tumors. Breast cancer is a rapid and unstable division of breast cells, causing a lump in the breast. It then passes through the lymph nodes to other parts of the body. This type of disorder may occur in milk ducts, glandular tissue or other breast tissue. All women are advised to check regularly for signs of breast cancer and to consult their doctor if there is a lump in the breast or if there is any change noticed in the tissue. It should also be noted that most breast lumps are benign. Many diagnostic procedures that can be used to diagnose breast cancer, but the most effective ones that help to diagnose the least possible damage on the physical and physical [1]. The early diagnosis

of breast cancer can significantly improve prediction and survival chances, because it can facilitate early medical care for patients. More specific identification of benign cancer can eliminate unnecessary treatment in patients. Therefore, several studies are now underway on accurate diagnosis of breast cancer and on whether tumors are classified as benign or malignant. Data mining and its unique benefits in serious features identification from multifaceted breast cancer corpus is extensively considered to be the procedure of choice in breast cancer pattern prediction and classification modelling [2]. This research aims to observe which features are most helpful in diagnosing malignant or benign cancer and to identify general trends that may aid us in model selection and hyper parameter selection. The goal is to classify whether the breast cancer is benign or malignant. To achieve this, the neural network classification method was used with ensemble boosting techniques to fit a function that can predict

The associate editor coordinating the review of this manuscript and approving it for publication was Ruqiang Yan.

the discrete class of new input [3]. The main objectives of this research are: to investigate the current breast cancer prediction model in order to improve its prediction diagnosis results; to improve the accuracy of breast cancer diagnosis; and to reduce the misdiagnosis errors based on the ensemble boosting learning method with RBF neural network algorithm. In the diagnosis of tumors and follow-up investigations, identifying and treating early breast cancer can save lives. A tumor that has been diagnosed early on is more likely to be treated successfully if it is not too large and has not spread. When cancer has spread, efficient diagnosis is tougher, and chances of survival are usually significantly lower [4]. The study of conventional machine learning indicates that the techniques are only used in a limited fashion, although data mining methods have great potential for medical research and enhance the diagnosis capacity of current computer-aided diagnosis (CAD) systems such as ALzubi *et al.*[5] and Kumar *et al.* [6].

Emphasize and investigation of the proposed method on premalignant breast tumors is one of the main tasks for this study. This study aims to predict the real suspected cancer cases and classify the tumors situations to (begin and malignant). The contributions of this study has been investigated by training the proposed method with real different cancer cases, and then examine them with whole parts of the real data samples to highlight the discrimination pattern of the premalignant breast tumors.

This research is organized into six sections. Section I has described the introduction of this research. Section II discusses the related studies. Section III describes the material and method that was used to build the proposed model and its necessary approach. The deep discussion of integrated model is presents in Section IV. Experiments results discussion of the proposed model will be illustrated in Section V. Finally, the summary and future work will be concluded in Section VI.

II. RELATED WORKS

Data extraction is the process of finding useful patterns in order to represent information from large data sets in medical filed such as Breast cancer [7]–[11], brain tumor diagnosis [6], [12]–[14], and diabetes diagnosis [15], [16]. This experience is useful for improving decision-making support, prevention, diagnosis and disease treatment [17]. The data extraction process with multiple attributes which may handle large volumes of data is known as a logical process for detecting interesting patterns in big data [15]. Data mining is intended to find patterns, reduce complexity and save time on processing. The prediction model for benign and malignant breast cancer has been developed by Chaurasia *et al.* [18]. The data set for Wisconsin's breast cancer consisted of sixty-nine cases, two categories (malignant and benign) and nine integer clinical attributes, for example cell size uniformity. To develop the data set of 683 examples, the researchers deleted 16 cases with the missing values from the data set. Results were, benign: 458 (65.5%) and malignant: 241 (34.5%) [19]. Yue *et al.* examined machine learning (ML)

techniques used for the diagnosis and prognostics of breast cancer [20]. The researchers based their study on studies with the aid of SVM[21], NN[22], KNN[23], [24], and DT algorithms [25]. In addition, the breast cancer list from Wisconsin was used. Classification and prediction accuracy of master learning systems have demonstrated their significant capability. A concise and informative knowledge collection was provided by the researchers. This data has been provided in a table of sources, algorithms, sampling methods and reliability of classification. Banu and Thirumalaikolundusubramanian [26] introduced a breast cancer prediction classifying technique. The researcher reported that Bayes classifications such as Boosted Augmented Naive Bayes, Tree Augmented Naive Bayes, and Bayes Belief Network can be used for the best classification performance and accuracy. To improve the accuracy, the gradient boosting techniques (GB) are combined with all classifiers. The three classifiers achieved approximately a comparable 90.1% accuracy before using the GB system. All classificatory tests were, however, enhanced by GB [24]. Akinsola *et al.* [27] proposed a prediction method for breast cancer that can help doctors identify a breast cancer diagnosis based on patients' clinical data (classes were a benign tumor). The home dataset was used in Lagos, which included more than 1700 cases. Eleven attributes were picked, such as cell size, cell shape and rank. To identify data on breast cancer, three supervised learning algorithms were used. C4.5, WEKA's toolkit was used to study Multilayer Perceptron (MLP) and Naïve Bayes. The Naïve Bayes classification data mining algorithm was predicted by Mirajkar and Lakshmi [22]. The proposed approach was used for estimating the risk of certain cancer types. The Naïve Bayes algorithm was used to classify cancer symptoms as the risk of breast and ovarian cancer. In the medical sector, the diagnosis was always a major issue on the basis of various tests carried out in different patients. Tests are designed to help a doctor diagnose correctly and reliably mass informative tumor information and cancer research feature data can now be accessed using information systems. The study reported that the breast cancer rate among women around the world, with about 1.7 million cases diagnosed in 2012, is counted as the second most common cancer among humans. The goal of the study is to collect data for the work of a program of learning machine to facilitate accurate diagnosis of breast cancer in the early stages of the disease. It is the main common category of cancer in America, with an average of eight out of every eight American women with breast cancer (12%). Most of these cases were recorded in the US state of Wisconsin and began to study and collect data [28]. Potdar and Kinnerkar[29] presented a comparison of the problem of learning the problem on the predecessors of the manner that has been the most important information on the basis of the disease that can be used in diagnosis of diseases. The search for the nature of the manner in the manner cannot be used to compare the identification of the manner in the predictor of the manner in the predictor of the manner in the breast cancer. The algorithms that were adopted were compared to the

Kit-Nearest Neighbors (KNN), and Bayesian Network Classifiers Research[29]. The alternative nitrate networks is the best so that the accuracy of the prediction by 97.4% [30]. Osman [21] proposed automatic breast cancer diagnostic approach using two-step clustering method and support vector machine algorithm. The hybrid approach aims to improve diagnosis precision and the medical misdiagnosis identification, thus solving breast-tumor-related screening problems. The two-step method and SVM algorithm was combined and used to distinguish the incoming tumors in order to determine the hidden characteristics of benign and malignant cancer. When analyzed in the UCI-WDBC, the evolved hybrid method improves its precision by 99.1%. [21]. To properly understand hidden breast cancer patterns, the analysis of the sequence of symptoms using machine learning (ML) is an important process. ML algorithms require the necessary extracting pattern from the huge corpus [31]. Cruz-Roa *et al.* [32] also divided the slide in breast histology into 100×100 patches that could be used to identify invasive carcinoma regions using his deep learning network. To classify eight groups of breast cancer, Han *et al.* [33] has used the deep learning framework. The Euclidean distance was minimized with the implementation of deep learning frameworks in the same class, while the Euclidean distance was maximized between two random class samples. Correspondingly, Wimmer *et al.* [34] merged the CNN with a layer of Fisher functions to encode the regional characteristics of tumors of breast cancer in a more selective region which effectively distinguishes breast cancer forms. Chen *et al.* [35] suggested a deep cascaded networks architecture, which would easily recover mitosis candidates while maintaining high-sensitivity. The candidates obtained were then identified by the next deep CNN, which can more specifically distinguish between mitosis and hard imitations. This method has therefore provided high performance in the identification of images of breast cancer. Future work on CAD systems will focus on improving quality for breast cancer diagnosis. The selection of features for the creation of a CAD program is a major step. In recent years a wide range of features have been created to describe breast cancer. Further work is therefore required to calculate robustness characteristics that can deliver a high level of classification precision. Choosing the optimal subset of features for guided learning problems involves a thorough search. The potential for bias in CAD systems varies. The above discussions indicate that the cancer research community uses both supervised and unregulated data mining approaches, although the majority of the work uses the semi-controlled approach. The provided literature is intended to assist clinicians in early diagnosis of breast cancer and in developing a strong and computationally efficient CAD program. Because data mining needs enough annotated training data, many researchers use a combination of public and private data and data enhancement techniques to overcome the problem of data scarcity.

This paper introduces an automated system for classifying breast cancer based on ensemble boosting learning and Radial Based Function Neural Network for the improvement of the classification of breast tumors (RBF-EBL). It is intended to assist medical professionals in the treatment of breast cancer.

The important of this research in the diagnosis and control of tumors of breast cancer is to diagnose and treat breast cancer at an early stage. When not so large and has not spread at an early stage, this is more likely to be successfully treated When cancer spreads, effective treatment is difficult and chances of survival are usually much lower.

III. MATERIALS AND METHODS

The technique used to diagnose breast cancer is discussed in this section. A reliable approach must be developed in order to enhance the identification of breast cancer measures before this work is conducted. The procedure for developing this approach is provided based on different stages. These stages are demonstrated in Figure 1, as follows:
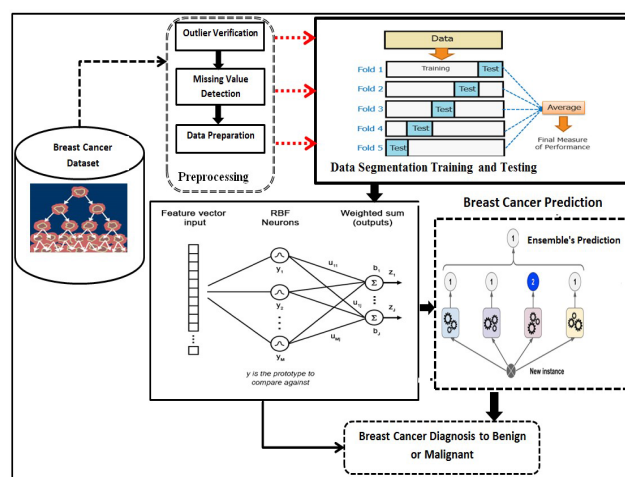

FIGURE 1. Proposed model.

A. DATASET

To accurately detect cancer as being benign or malignant, a collection of specimens the proposed method tested based on four types of the breast cancer UCI repository. There are: Wisconsin Data set for Diagnostic Breast Cancer (WDBC) [36]; Breast Cancer Original (WBC) [19]; Breast Cancer Prognostic (BCP) [37]; and Breast Cancer Diagnosis (BCD) [19]. Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

The characteristics of the fine needle aspirator (FNA) in the breast mass are calculated from a digitalized image. Some of the images identified by Bennett[38]. Applicable characteristics were nominated using comprehensive search in the space of (1-4) characteristics and (1-3) separating planes. The real linear program applied to achieve the separating plane in the 3-dimensional space as defined by Bennett[39]. Some samples reach regularly as Dr. Wolberg reports his medical cases. The dataset consequently reflects the chronological data groups from (1989 to 1991).

TABLE 1. Features of the (WDBC) samples.

Feature	Feature Name	Description
F1	ID	ID number
F2	Diagnosis	The diagnosis of breast tissues (M = malignant, B = benign)
F3	Radius	Distances from center to points on the perimeter
F4	Texture	Standard deviation of gray-scale values
F5	Perimeter	Size of the core tumor
F6	Area	-
F7	Smoothness	Local variation in radius lengths
F8	Compactness	Perimeter^2 / area - 1.0
F9	Concavity	Severity of concave portions of the contour
F10	Concave Points	Number of concave portions of the contour
F11	Symmetry	-
F12	Fractal dimension	coastline approximation - 1

This section provides information details of the WDBC as examples of the used breast cancer dataset. The WDBC is a selection of the machine learning library in the University of California, Irvine [36]. Table 1 demonstrates the WDBC dataset.

There are 569 instances and different features of actual value. Table 1 provides the summary of the attributes. The class feature of the dataset is benign or malignant. 357 cases are benign and 212 are malignant in this dataset. No sample has a missing value of the attribute.

B. PREPROCESSING STAGE

Preprocessing is a major data mining stage in which the dataset is cooked prior to the learning process. Outliers within the dataset are removed and missing values are updated, and data cleaning processes are carried out. In addition, the dataset features are analyzed as a data preparation step to specify the standard error, maximum and mean value for each attribute.

C. DATA SEGMENTATION BASED ON CROSS-VALIDATION STAGE

The proposed model is often used in practical applications to deal with problems still unresolved. In addition to matching the target function in the training sets, the highly practical of the suggested model has outstanding generalization quality on the testing set. A part of the data set is used for verifying the proposed model's performance, which the algorithm's weights have been determined for, and a method for determining the number of data features is used to improve the performance of generalization. The following procedure is used to optimize the use of the data set:

1. *Divide the dataset into sub-sets of (K), where (i)th is validation and the rest of the sub-sets (K−1) are the training sets.*
2. *Switch I from (1−K), practice, and validate the proposed method, using the development method for achieving*

an optimum number of features within the dataset in training and testing stage.
3. *The average number (K) obtained is estimated and then the optimum number of features is obtained for the entire data set.*

D. RBF NEURAL NETWORK ARCHITECTURE

RBF has benefits of good generalization, faster training, easy design, and robust tolerance to feature noise. In 1988, Broomhead and Lowe developed the RBF network framework [40]. The RBF network structure is considered an important learning strategy for data classification in a neural network. The traditional estimation hypothesis is based on the RBF network structure. It has the capacity to calculate extensively. The RBF organization, as well as its smoother structure and a much speedier method of planning, is a contrasting and well-known approach for the excellent Multilayer Perceptron. The RBF scheme is responsible for the precise implementation of a multidimensional data arrangement[41]. It can be regarded as a kind of useful link network [42]. It is considered as one of the most simple and powerful network architectures[43]. Broomhead and Lowe introduced radial-base functions in the development of neural networks and explored how RBF models non-linear relations and generalized or interpolated data points [43].

The RBF neural network architecture consists of three layers:

1) INPUT LAYER

The input layer is represent dataset feature includes the main features of breast cancer dataset that discussed and reported in table 1. These features are illustrated in Figure 1 and Figure 2 as (x1, x2,.. x12) except x2 that represent the output of the RBF neural network.
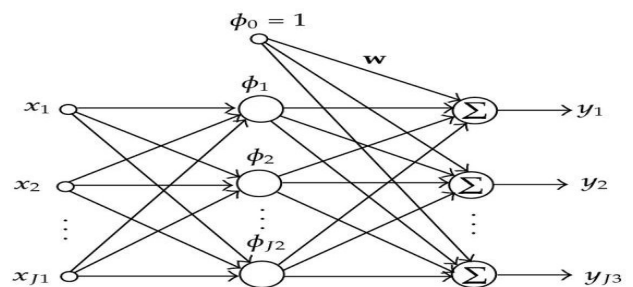

FIGURE 2. Architecture of the RBF neural network.

2) HIDDEN LAYER

In the hidden layer are put the activation features which are the exponential functions. The RBF feature is an inverter neural network with three layers (*J1-J2-J3*), as demonstrated in Figure 1. In the hidden-layer, each node uses a RBF defined by $\phi(r)$ as its non-linear activation function. The hidden layer transforms the input non-linear and the output-layer represents a linear integrator that maps non-linearity to a new

space. Commonly, the RBF is utilized on all nodes; that is, the radial base function nodes have the non-linearity:

$$\phi_i(\vec{x}) = \phi(\vec{x} - \vec{ci}), \quad i = 1 \ldots J2 \tag{1}$$

where:

$\vec{ci}$ is define the prototype of the sample or middle of the ith node

$\phi(\vec{x})$ is a radial base function

In the hidden layer, addition of a neuron with a permanent $\phi_0(r) = 1$ activation function that model the biases of the output layer neurons. With the linear optimization strategy, the RBF network provides an overall optimum solution for weighting variable in the mean-square (MS) error as discussed by Osman and Alzahrani [44]. Figure 2 shown the RBF neural network architecture:

3) OUTPUT LAYER

The $y_i(\vec{x})$ is define the ith output values, and it equal the joining weight from the hidden layers to the output layers. The RBF is characteristically nominated as the Gaussian function, and such an RBF is typically labeled the Gaussian RBF network.

The reason of using the RBF neural network is that the RBFN can define any arbitrarily complex decision boundary. In other words, you can always improve its accuracy by using more RBF neurons. In addition, the significance function of each node in RBNN's hidden layer can be easily interpreted. However, this reason is complicated in many of data classifiers.

The input layers, hidden layers, and output layers have define as $J1$, $J2$, and $J3$ neurons respectively. In Figure 2, an activation function $\phi_0(\vec{x}) = 1$ corresponds to the bias in the output layers, whereas $\phi_i(\vec{x})$'s define the non-linearity at the hidden nodes. Only non-linear activation functions allow such networks to compute non-trivial problems with just a few nodes. It can be shown that the activation functions as a "(1) digital network, depending on the input. The values of hidden layer determined by converts the input into a non-linear one and a linear combination maps the non-linearity into a new space is the output layer. The RBF network is a receptive area or a localized network with localized RBFs such as the Gaussian RBF network. When the input is close to the node prototype, the localized approach method provides the highest output. Input vectors close together generate mostly identical outputs for a well-trained and localized RBF network, whereas distant input-vectors generate quite independent outputs. It is the intrinsic property of local generalization. A receptive field networks is an associated neural network, since the input to the network determines only a small subspace. This property is highly attractive as the receptive field function changes.

E. ENSEMBLE BOOSTING LEARNING METHOD

Boosting means a family of algorithms that can transform weak learners into powerful learners. The key principle of improvements is the adaptation of a set of soft learners to weighted data versions-models that are only marginally better than random guessing, such as small decision trees. Cases that were misclassified as 'error' by earlier rounds were given more weight [29]. The classifications are then combined to generate the final classification by weighted majority voting (classification) or a weighted total (regression). The main difference between the methods of boosting and bagging, for example, is that base learners are sequenced in a weighted data version [30], the error rate of which is slightly below a random selection. AdaBoost originally developed the first soft C1(x), which is a decision stump in most cases. When the weight of that observation is increased – when at least a misclassification is made by one of the classification methods – subsequently, the second classifier builds with new weight. The predictions of all classification systems are combined with weighted majority voting techniques, thus generating a final prediction which can be calculated as:

$$C(x) = \text{sign}(M = 1amCm(x)) \tag{2}$$

where *a1, a2,..* and *am* are numbers that are generated by the boosting algorithm for the improvement of the sequence [32].

The most popular approach for ensemble learning is bagging and boosting, with both focused on adding variety by adjusting the training set, so that the learning algorithm is applied repeatedly over different training settings. Each boosting phase adds weight to each of the training data points. Bagging conducts a random sampling of the data by substitution, uses bootstrap sampling to obtain the subsets of data for basic apprenticeship learning, uses voting for classifying and averaging for regression to combine the performance of basic apprentices and uses category voting and average regression to measure the results of specific students. When boosting, the weights of these weights is modified by means of random sampling and is substituted for the weighted data to give greater importance to the previously unclassified samples. Ensemble classification is a general term for the combination of several classification algorithms to improve the system's predictive efficiency [20], [37]. Previously, various scientists [38]–[41] used classification ensembles and debated their supremacy over individual classification [42]. The Ensemble Boosting Learning algorithm is shown as:

Ensemble boosting learning algorithm
   Adaboost(*sample, L, T*)
1) $D1(i) \leftarrow 1/m$ for each of m examples
2) **For** $t \leftarrow 1$ to T
3) $h_t \leftarrow$ apply L to *sample*s using D_t distribution
4) $\epsilon_t \leftarrow$ sum of $D_t(i)$ for examples miss-classified by h_t
5) $\beta_t \leftarrow (1 - \epsilon_t)/\epsilon_t$
6) $D_{t+1}(i) \leftarrow \begin{cases} D_t(i)\beta t & \text{if } h_t \text{ miss-classifies example } i \\ D_t(i) & \text{otherwise} \end{cases}$
7)   normalize D_{t+1} so it sums to 1
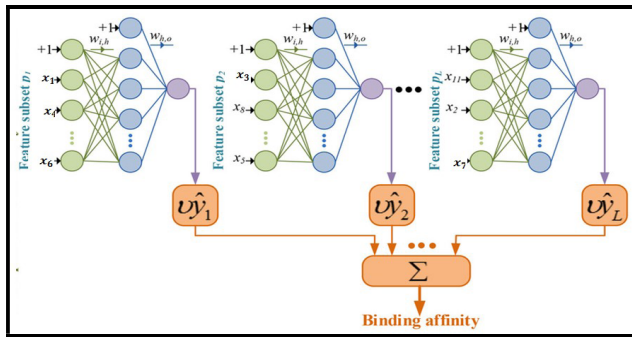8) **return** $h_t, \ldots, h_T$ and $\beta_1, \ldots, \beta_T$

FIGURE 3. Structure of RBF- EBL method.

IV. RBF-EBL METHOD

In this section, the proposed method integrated the ensemble boosting classifier with the RBF neural network algorithm for breast cancer diagnostic and indicated that the ensemble approach imitates human reasoning as multiple perspectives are taken into account before making a final decision. The RBF have benefits of flexible design and solid tolerance to breast cancer features that content noise samples. The main reason for using the RBF neural network in this paper is that the proposed model not needs to have multiple hidden layers because of combining the ensemble boosting learning approach which focuses to increase the learning process during the RBF prediction procedure.

The unanimous ballot was taken, because the real world application of this study needs great trust, since the division of patients with breast cancer in bio-classes, among a variety of decision-making fusion methods, present in the report. The main objectives of using the ensemble boosting learning in this study is to improve the learning level by collecting the breast cancer samples with similar patterns together thus, the complexity will be reduced and the diagnosis interpretation will be accurate. The integration structure between the ensemble boosting learning method and RBF neural network (RBF-EBL) is demonstrated in Figure 3.

The hybrid module uses the integration between the ensemble boosting learning classifier with RBF neural network. The features input are divided into different subsets ($x1,x2, ..x12$) and then used as input features of the RBF neural network. The Parameters ($x1,x2, ..x12$) are describing the characteristics of the cell nuclei present in the Wisconsin breast cancer dataset and *can be defined as:*

- Parameter x1 represents the ID of the patient.
- Parameter x2 represents the diagnosis results that can be malignant or benign.
- Parameter x3 represents the radius distances from center to points on the perimeter.
- Parameter x4 represents the texture standard deviation of gray-scale values.
- Parameter x5 represents the perimeter (size of the core tumor).
- Parameter x6 represents the area of the breast cancer.

- Parameter x7 represents the smoothness (local variation in radius lengths).
- Parameter x8 represents the compactness (perimeter^2 / area - 1.0) g).
- Parameter x9 represents the concavity (severity of concave portions of the contour).
- Parameter x10 represents the concave points (number of concave portions of the contour).
- Parameter x11 represents the symmetry.
- Parameter x12 represents the fractal dimension and the coastline approximation normally represented by (1).

The υy_i is defining the ith output values, and it equals the joining weight from the hidden layers to the output layers. The outputs of the subsets with each RBF are then binding together as on output. This module uses a hybrid combination method in order to combine predictions from individual classifiers on the basis of their importance to classify various types of breast cancer datasets. The hybrid approach employs 10-fold cross-validation to separate data sets as training and testing samples into different groups with and without ensemble boosting learning. The RBF classifier employed with the output of the ensemble boosting learning algorithm for possible obtaining high accuracy. Each training and testing experiment uses breast cancer datasets features as an input variable to the RBFNN. Then the target output is a class feature (benign or malignant). The results of the RBFNN classifier with ensemble boosting learning proved an improvement performance when the RBFNN technique classified the dataset with ensemble boosting learning output.

In contrast, a combination of winner-takes-all and weighted-most majority strategies refers to the combination of the ensemble boosting learning algorithm. This combination method aims to both reduce the misdiagnosis which existed and increase the accuracy of healthy breast cancer diagnosis. The majority model with small value predictions would benefit from minority models with much higher value predictions in the weighted majority rule. The prediction coefficients of the majority of models with appropriate large values are lost in the winning process in contrast with the only model with the greatest value. The integrated module is built in the learning stage of the ensemble boosting process. Each RBF network extracts a classification result for each cross-validation fold from the data set. This classification must then be integrated to scheme the final classification model.

V. EXPERIMENTATION AND RESULTS DISCUSSION

In the experiments, the breast cancer datasets were utilized to specify the breast tumor level as either malignant or benign. The data had each case described as either a malignant or a benign sample. The integrated method employed by learning and testing the data based on an integration ensemble boosting learning method and RBF NN technique. The main objective of ensemble boosting learning in this study is to improve the learning level by gathering the breast tumor cases composed of similar samples thus decreasing the misdiagnosis
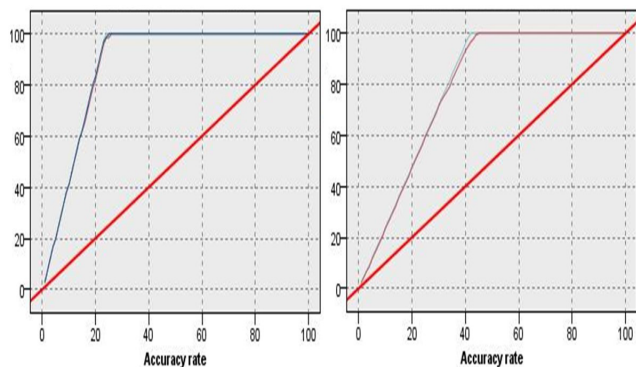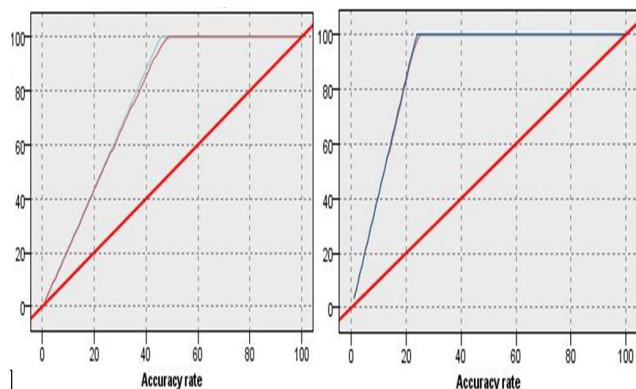
FIGURE 4. Best diagnosis result using RBF.

FIGURE 5. Best diagnosis result using EBL-RBF.

FIGURE 6. Practical experimental results between the proposed EBL-RBFNN with other classification methods.

TABLE 2. The average accuracy results of the proposed method.

Dataset	Average Accuracy based on RBF N.N		Average Accuracy based on EBL-RBF N.N	
	Training	Testing	Training	Testing
WBC	95.92%	97.6%	97.36%	97.4%
BCD	94.96%	96.58%	97.09%	98.4%
PCB	95.57%	97.7%	96.8%	97.7%
WBCD	95.37%	96.12%	96.47%	97.0%

and increasing the accurate medical analysis accordingly. The obtained results from the learning and testing procedure on the datasets are explained in Table 1, which proves a set of results achieved by RBF neural network model without EBL and with EBL algorithm. The RBF model was applied once more with the production of the ensemble boosting learning algorithm with the aim of obtaining greater accuracy. The average accuracy results have been achieved based on different types of the breast cancer dataset, as mentioned in Section III. The results of the proposed method are shown in Table 2.

The outcomes of the RBFNN predictor with ensemble boosting learning evidenced an enhancement diagnosis when the RBFNN method predicted the datasets with ensemble boosting learning results. Interestingly, the ensemble boosting learning algorithm increases the diagnosis accuracy with 98.4% rate. The best diagnosis results using RBF are demonstrated in Figure 4, while the improvement prediction accuracy is illustrated in Figure 5, as shown.

Figures 4 and 5 show the both learning and testing results of the RBFNN model without learning improvement and with learning improvement using ensemble boosting learning algorithm. The 10-fold cross-validations were considered and the diagnosis accuracy based on RBFNN without EBL achieved 97.6%, 96.58%, 97.7% and 96.12% for the WBC, BCD, and WBCD datasets respectively. In contrast, the average testing analysis results based on RBFNN with
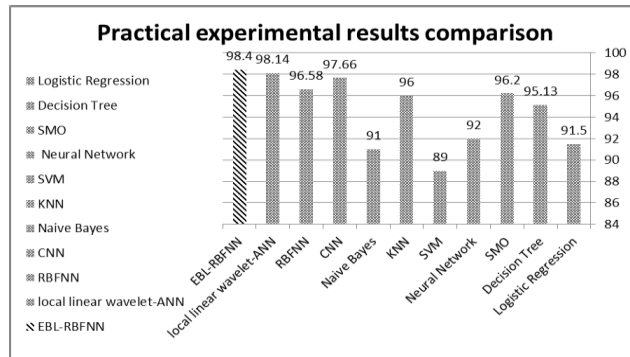
EBL obtained 97.4%, 98.4%, 97.7% and 97.0% for the WBC, BCD, BCP, and WBCD datasets respectively.

It can be noted that, through Figures 4 and 5, the best result achieved using RBFNN classifier with EBL is 97.09% in the learning results and 98.4% in the testing experimentations. The deduction is that there is an improvement when using the ensemble boosting learning method. The results of the RBFNN with ensemble boosting learning are improved, and the breast tumor analysis is more precise when using the RBFNN results with the ensemble boosting learning method in the BCD dataset.

The proposed model has been compared with different breast cancer classification models in term of diagnosis accuracy using UCI Breast cancer dataset. The comparison between the proposed method and the state of the art methods that were discussed in Section II is shown in Table 4.

Table 4 shows the comparison of diagnosis accuracy for the state-of-the-art classification algorithms with EBL-RBF neural network method). It can be noted that the obtained result of the proposed method using EBL-RBF neural network is better than several diagnosis classification algorithms in term of prediction accuracy with 98.4% ratio. In addition, the practical experimental results between the proposed EBL-RBFNN with other classification methods such as random forest, SVM, RBFNN, NN, etc. represented in different views as shown in Figure 6.

Figure 6 demonstrates the comparison of practical experimental results between the proposed EBL-RBFNN with other classification methods in terms of Breast cancer diagnosis

TABLE 3. T-test Statistical significance outcomes.

Dataset Results between RBF-NN boosting ensemble & RBF-NN	Paired Differences							Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	
Breast Cancer WBC Original	-.17000	1.84213	.58253	-1.48778	1.14778	-.292	9	0.777
Breast Cancer Diagnosis	1.84000	2.42634	.76728	.10430	3.57570	2.398	9	0.040
Breast Cancer Prognostic	.01000	3.30066	1.04376	-2.35115	2.37115	.010	9	0.993
Breast Cancer WBC Diagnostic	.88000	3.20513	1.01355	-1.41282	3.17282	.868	9	0.408

TABLE 4. Comparison between the proposed method and other state of the art methods.

Reference Method	Classifier Name	Accuracy Results
Yusuff et al.[45]	Logistic Regression	91.5%
Asri et al.,[46]	Decision Tree	95.13%
Chaurasia and Pal[47]	SMO	96.2%
Rani[48]	Neural Network	92%
Ali and W.Z [49]	SVM	89%
Potdar and Kinnerkar[29]	KNN	96%
Banu & Ponniah [26]	Naive Bayes	91%
Karthik et al. [50]	CNN	97.66%
Senapati and Dash[51]	local linear wavelet-ANN	98.14%
Proposed Method	EBL-RBFNN	98.4%

TABLE 5. Comparison between the proposed method and other ensemble learning with breast cancer classifier methods.

Reference Method	Ensemble learning Classifier	Accuracy Results
[52]	EBL- Random Forest	96.4861%
[52]	EBL-SVM	97.89%
[53]	EBL-KNN	97.076%
[52]	EBL-Naive Bayes	96.3397%
[54]	EBL-CNN	96.39%
[55]	EBL-ResNet	88.75%
[55]	EBL- DeNt	91.25%
[52]	EBL-Logistic Regression	96.9253%
[52]	EBL-BayesNet	96.9253%
Proposed Method	EBL-RBFNN	98.4%

accuracy. The results of the experiment indicated that the EBL-RBFNN achieved the best results (98.4%) compared with other classifiers methods.

In Table 5, the proposed method considered several neural networks based on ensemble learning method. Therefore, the proposed method has been compared with the combination of ensemble learning method and other machine learning methods to investigate the effectiveness of the proposed method with other ensemble learning with breast cancer classifier methods.

It can be noted that the obtained result of the proposed method using EBL-RBF neural network is better than other ensemble learning with breast cancer classifier methods.

The T-testing algorithm was performed between the data obtained from the experiment using RBFNN as statistical significance testing and testing based on the RBFNN with ensemble boosting learning method, and it presented the

enhancements obtained by using the hybrid RBFNN with ensemble boosting learning method only with Breast Cancer Diagnosis (BCD) dataset. The low T-test results (normally less than 0.05) shows that the two methods are significantly changed. This condition was emphasized in evaluation measures based on results obtained in Table 3 regarding the reliability of values (0.040). It indicates that the RBFNN with the ensemble boosting learning method obtained enhancement significant on the diagnosis results.

VI. CONCLUSION AND FUTURE WORK

This research tried to investigate the breast cancer disease diagnosis to adjust the prediction process based on the RBF neural network and ensemble boosting learning methods. The quality of breast cancer disease prediction was emphasized using integrated EBL RBF neural network algorithms. In this study, the experiments were conducted based on four types of UCI breast cancer datasets. The experiments concluded that the diagnosis results have been applied by using EBL-RBFN neural network algorithm with different types of breast cancer features and datasets. Also, the deep examinations highlighted that the integrated RBFN and ensemble boosting method obtained the best result in term of diagnosis accuracy with Breast Cancer Diagnosis dataset (BCD) with 98.4%. Through this research, it can be noted that the limitation of this study is that the system needs to deal with the doctor to support the diagnostic stance of the cancer stage. This process requires several of steps that may delay the detection of cancer which starting from collecting the patient's laboratory data and inserting them into the proposed system then predict and detect the cancer stage. The current research needs an intelligent prediction system that interacts with the patient and disease without a doctor's intervention to facilitate the early detection of cancer.

In future, more focus should be on how to improve the current study using some of the optimization techniques to enhance the diagnosis accuracy of the breast cancer disease.

REFERENCES

[1] A. E. Giuliano, K. K. Hunt, K. V. Ballman, P. D. Beitsch, P. W. Whitworth, and P. W. Blumencranz, "Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: A randomized clinical trial," *Jama*, vol. 305, pp. 569–575, Feb. 2011.

[2] R. Etzioni, N. Urban, S. Ramsey, M. McIntosh, S. Schwartz, and B. Reid, "Early detection: The case for early detection," *Nature Rev. cancer*, vol. 3, p. 243, 2003.

[3] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, 2015.

[4] D. Weller, P. Vedsted, G. Rubin, F. M. Walter, J. Emery, S. Scott, C. Campbell, R. S. Andersen, W. Hamilton, F. Olesen, P. Rose, S. Nafees, E. van Rijswijk, S. Hiom, C. Muth, M. Beyer, and R. D. Neal, "The aarhus statement: Improving design and reporting of studies on early cancer diagnosis," *Brit. J. Cancer*, vol. 106, no. 7, pp. 1262–1267, Mar. 2012.

[5] J. A. ALzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna, and C. Thaventhiran, "Boosted neural network ensemble classification for lung cancer disease diagnosis," *Appl. Soft Comput.*, vol. 80, pp. 579–591, Jul. 2019.

[6] A. Kumar, M. Ramachandran, A. H. Gandomi, R. Patan, S. Lukasik, and R. K. Soundarapandian, "A deep neural network based classifier for brain tumor diagnosis," *Appl. Soft Comput.ing*, vol. 82, Sep. 2019, Art. no. 105528.

[7] S. B. Sakri, N. B. Abdul Rashid, and Z. Muhammad Zain, "Particle swarm optimization feature selection for breast cancer recurrence prediction," *IEEE Access*, vol. 6, pp. 29637–29647, 2018.

[8] K. Liu, G. Kang, N. Zhang, and B. Hou, "Breast cancer classification based on fully-connected layer first convolutional neural networks," *IEEE Access*, vol. 6, pp. 23722–23732, 2018.

[9] F. Liu and M. Brown, "Breast cancer recognition by support vector machine combined with daubechies wavelet transform and principal component analysis," in *Proc. Int. Conf. ISMAC Comput. Vis. Bio-Eng.*, 2018, pp. 1921–1930.

[10] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin, "Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features," *IEEE Access*, vol. 7, pp. 105146–105158, 2019.

[11] J. T. Kwak and S. M. Hewitt, "Nuclear architecture analysis of prostate cancer via convolutional neural networks," *IEEE Access*, vol. 5, pp. 18526–18533, 2017.

[12] G. Manogaran, P. M. Shakeel, A. S. Hassanein, P. M. Kumar, and G. C. Babu, "Machine learning approach-based gamma distribution for brain tumor detection and data sample imbalance analysis," *IEEE Access*, vol. 7, pp. 12–19, 2018.

[13] P. Mohamed Shakeel, T. E. E. Tobely, H. Al-Feel, G. Manogaran, and S. Baskar, "Neural network based brain tumor detection using wireless infrared imaging sensor," *IEEE Access*, vol. 7, pp. 5577–5588, 2019.

[14] H. H. Sultan, N. M. Salem, and W. Al-Atabany, "Multi-classification of brain tumor images using deep neural network," *IEEE Access*, vol. 7, pp. 69215–69225, 2019.

[15] Y. Sun, "The neural network of one-dimensional convolution—An example of the diagnosis of diabetic retinopathy," *IEEE Access*, vol. 7, pp. 69657–69666, 2019.

[16] A. Dagliati, S. Marini, L. Sacchi, G. Cogni, M. Teliti, V. Tibollo, P. De Cata, L. Chiovato, and R. Bellazzi, "Machine learning methods to predict diabetes complications," *J. Diabetes Sci. Technol.*, vol. 12, no. 2, pp. 295–302, May 2017.

[17] D. N. Ganesan, D. K. Venkatesh, D. M. A. Rama, and A. M. Palani, "Application of neural networks in diagnosing cancer disease using demographic data," *Int. J. Comput. Appl.*, vol. 1, no. 26, pp. 81–97, Feb. 2010.

[18] V. Chaurasia, S. Pal, and B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *J. Algorithms Comput. Technol.*, vol. 12, no. 2, pp. 119–126, Feb. 2018.

[19] A. Frank. (2010). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml

[20] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, "Machine learning with applications in breast cancer diagnosis and prognosis," *Designs*, vol. 2, no. 2, p. 13, May 2018.

[21] A. Hamza, "An enhanced breast cancer diagnosis scheme based on two-step-SVM technique," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, pp. 158–165, 2017.

[22] P. Mirajkar and P. Lakshmi, "Prediction of cancer risk in perspective of symptoms using Naïve Bayes classifier," *Int. J. Eng. Re-Search Comput. Sci. Eng.*, vol. 4, no. 9, pp. 2320–2394, 2017.

[23] S. Goyal, "Naïve Bayes model based improved k-nearest neighbor classifier for breast cancer prediction," in *Proc. Int. Conf. Adv. Informat. Comput. Res.*, 2019, pp. 3–11.

[24] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of heart disease using K- nearest neighbor and genetic algorithm," *Procedia Technol.*, vol. 10, pp. 85–94, 2013.

[25] Y.-Q. Liu, C. Wang, and L. Zhang, "Decision tree based predictive models for breast cancer survivability on imbalanced data," in *Proc. 3rd Int. Conf. Bioinf. Biomed. Eng.*, Jun. 2009, pp. 1–4.

[26] B. Banu and P. Thirumalaikolundusubramanian, "Comparison of Bayes classifiers for breast cancer classification," *Asian–Pacific J. Cancer Prevention: APJCP*, vol. 19, no. 10, p. 2917, 2018.

[27] A. F. Akinsola, M. A. Sokunbi, F. M. Okikiola, and I. O. Onadokun, "Data mining for breast cancer classification," *Int. J. Eng. Comput. Sci.*, vol. 6, no. 8, pp. 22250–22258, 2017.

[28] Collaborative Group on Hormonal Factors in Breast Cancer, "Breast cancer and hormone replacement therapy: Collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer," *Lancet*, vol. 350, pp. 1047–1059, 1997.

[29] K. Potdar and R. Kinnerkar, "A comparative study of machine learning algorithms applied to predictive breast cancer data," *Int. J. Sci. Res.*, vol. 5, no. 9, pp. 1550–1553, 2016.

[30] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Proc. ICML*, 1997, pp. 412–420.

[31] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, 1994, pp. 487–499.

[32] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. C. Shih, J. Tomaszewski, F. A. González, and A. Madabhushi, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, no. 1, p. 46450, Apr. 2017.

[33] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, and Y.-K. Seong, "A deep learning framework for supporting the classification of breast lesions in ultrasound images," *Phys. Med. Biol.*, vol. 62, no. 19, pp. 7714–7728, Sep. 2017.

[34] G. Wimmer, A. Vécsei, M. Häfner, and A. Uhl, "Fisher encoding of convolutional neural network features for endoscopic image classification," *J. Med. Imag.*, vol. 5, no. 03, p. 1, Sep. 2018, 034504.

[35] H. Chen, Q. Dou, X. Wang, J. Qin, and P. A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1160–1166.

[36] G. I. Salama, M. Abdelhalim, and M. A.-E. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers," *Breast Cancer (WDBC)*, vol. 32, p. 2, Jan. 2012.

[37] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proc. Nat. Acad. Sci. USA*, vol. 87, pp. 9193–9196, 1990.

[38] K. P. Bennett, "Decision tree construction via linear programming," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, 1992.

[39] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optim. Methods Softw.*, vol. 1, pp. 23–34, Jan. 1992.

[40] D. S. Broomhead and D. Lowe, "Radial basis functions, multi-variable functional interpolation and adaptive networks," Royal Signals Radar Establishment Malvern, Malvern, U.K., Tech. Rep. 4148, 1988.

[41] M. J. Powell, "Radial basis function for multivariable interpolation: A review," in *Proc. IMA Conf. Algorithms Approximation Functions Data*, 1985, pp. 143–167.

[42] M. Klaseen and Y.-H. Pao, "The functional link net in structural pattern recognition," in *Proc. IEEE TENCON: IEEE Region 10 Conf. Comput. Commun. Syst. Conf.*, vol. 2, Sep. 1990, pp. 567–571.

[43] M. A. Arbib, "The mirror system, imitation, and the evolution of language," in *Imitation in Animals and Artifacts*. Cambridge, MA, USA: MIT Press, 2002, pp. 229–280.

[44] A. H. Osman and A. A. Alzahrani, "New approach for automated epileptic disease diagnosis using an integrated self-organization map and radial basis function neural network algorithm," *IEEE Access*, vol. 7, pp. 4741–4747, 2019.

[45] H. Yusuff, N. Mohamad, U. Ngah, and A. Yahaya, "Breast cancer analysis using logistic regression," *Int. J. Res. Rev. Appl. Sci.*, vol. 10, pp. 14–22, 2012.

[46] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Comput. Sci.*, vol. 83, pp. 1064–1069, 2016.

[47] V. Chaurasia and S. Pal, "A novel approach for breast cancer detection using data mining techniques," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 1, p. 17, Jan. 2017.

[48] K. U. Rani, "Parallel approach for diagnosis of breast cancer using neural network technique," *Int. J. Comput. Appl.*, vol. 10, no. 3, pp. 1–5, Sep. 2010.

[49] E. E. E. Ali and W. Z. Feng, "Breast cancer classification using support vector machine and neural network," *Int. J. Sci. Res.*, vol. 5, no. 3, pp. 1–6, 2016.

[50] S. Karthik, R. S. Perumal, and P. C. Mouli, "Breast Cancer Classification Using Deep Neural Networks," in *Knowledge Computing and Its Applications*. Singapore: Springer, 2018, pp. 227–241.

[51] M. R. Senapati and P. K. Dash, "Local linear wavelet neural network based breast tumor classification using firefly algorithm," *Neural Comput. Appl.*, vol. 22, nos. 7–8, pp. 1591–1598, Apr. 2012.

[52] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *Eur. J. Oper. Res.*, vol. 267, no. 2, pp. 687–699, Jun. 2018.

[53] A. S. Assiri, S. Velastin, and S. Nazir, "A hybrid ensemble method for accurate breast cancer tumor classification using state-of-the-art classification learning algorithms," MDPI, Basel, Switzerland, Tech. Rep. 2019110341, 2019. [Online]. Available: https://www.preprints.org/manuscript/201911.0341/v1

[54] R. Rasti, M. Teshnehlab, and S. L. Phung, "Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks," *Pattern Recognit.*, vol. 72, pp. 381–390, Dec. 2017.

[55] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, "EMS-net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images," *Neurocomputing*, vol. 366, pp. 46–53, Nov. 2019.

AHMED HAMZA OSMAN received the bachelor's degree in computer science from the International University of Africa, the master's degree in computer science from the Sudan University of Science and Technology, Sudan, and the Ph.D. degree in computer science with excellent academic achievements from Universiti Teknologi Malaysia (UTM).

He was the Head of the Computer Science Department, Faculty of Computer Studies, International University of Africa. He currently works as an Associate Professor with King Abduaziz University (KAU), Saudi Arabia. His research interest includes information retrieval, plagiarism detection, soft computing, data mining, natural language processing, and text summarization.

HANI MOETQUE ABDULLAH ALJAHDALI was born in Jeddah, Saudi Arabia, in 1983. He received the B.Sc. degree in computer science from King Abdulaziz University, Jeddah, in 2005, and the M.Sc. degree in information technology and the Ph.D. degree in computer science from the University of Glasgow, in 2009 and 2015, respectively.

From 2005 to 2007, he was with Saudi Electricity Company as a Budget and System Analyst. In 2011, he has appointed as a Lecturer at the Department of Information Systems, King Abdulaziz University. He is currently an Associate Professor with the Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh. His research interests include information security, human–computer interaction, software engineering, big data, and data mining.

. . .