Biomedical Signal Processing and Control 68 (2021) 102682

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

An efficient classification framework for breast cancer using hyper parameter tuned Random Decision Forest Classifier and Bayesian Optimization

Pratheep Kumar P [a,*], Mary Amala Bai V [b], Geetha G. Nair [c]

[a] Noorul Islam Centre for Higher Education, Department of Computer Science and Engineering, Kumaracoil, Tamil Nadu, India
[b] Department of Information Technology, Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu, India
[c] Women and Children Hospital, Thycaud, Thampanoor, Kerala, India

ARTICLE INFO

Keywords:
Breast cancer
Feature weighting
Random Decision Forest Classifier
Kernel neutrosophic C-Means clustering
Bayesian Optimization algorithm

ABSTRACT

Decision tree algorithm is one of the algorithm which is easily understandable and interpretable algorithm used in both training and application purpose during breast cancer prognosis. To address this problem, Random Decision Forests are proposed. In this manuscript, the breast cancer classification can be determined by combining the advantages of Feature Weight and Hyper Parameter Tuned Random Decision Forest classifier. Here the Kernel Neutrosophic C-Means Clustering is used as Feature Weight, which allocates greater weights to applicable features and smaller weights to less applicable features. Then Random Decision Forest classifier model are optimized with the help of the Bayesian Optimization algorithm to obtain optimal hyper tuning parameters. By this, the accurate classification of breast cancer is successfully achieved. Then the efficiency of the proposed system is executed in python. The performance analysis are executed in Wisconsin prognostic Breast Cancer (WPBC) dataset, 70 % training and remaining 30 % testing is compared with the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, the accuracy analysis of proposed feature weight and Random Decision Forest Classifier with Bayesian Optimization (FW + BOA-RDF) in Breast Cancer Wisconsin (Prognosis) Data Set is 6.66 %, 12.659 % and 37.618 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM, FW + GA-SVM respectively. The performance analysis in Wisconsin prognostic Breast Cancer (WPBC) dataset, 75 % training and the remaining 25 % testing is compared at Wisconsin Diagnostic Breast Cancer (WDBC) dataset, the accuracy analysis of FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 3.7146 %, 5.27398 % and 4.4413 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM, FW + GA-SVM respectively.

1. Introduction

Breast cancer usually begins with some uncontrolled cells in the breast region to grow peculiarly [1]. This uncontrolled breast cells are divided rapidly than healthy cells. It begins to form a lump or mass with the condition. Invasive ductal carcinoma is one of the types of breast cancer caused by milk-producing ducts cells [2,3]. Invasive lobular carcinoma is other types of breast cancer caused by a glandular tissue called lobules [4,5]. Early detection of breast cancer is essential for health concern. Early treatment of Breast cancer can decrease the growth of cancer tissue to other parts of the body [6]. The breast cancer may be benign that is not dangerous to health or malignant type [7]. Benign type is generally non-cancerous, which grows merely in one part of the body and can be treated with medicines or surgery [8]. The malignant type is generally cancerous tumor that spreads throughout the body [9]. Numerous processes and specializations are commenced in favor of investigation and collection of information about brain cancer; however it is extremely demands assignment for doctor of medicine to appreciate each and every special features of cancer from the most extensive data.

Mammography is an extensively used imaging modality for screening the breast cancer. But it cannot detect the masses accurately. Therefore different types of algorithm are used to overcome the deficiency. Decision trees algorithm is an easy-to-understand and interpretable algorithm used in both training and application purpose during breast cancer prognosis [10]. But it has a limit due to over fitting, which

* Corresponding author.
E-mail address: gptcpala.pk@gmail.com (P.K. P).

https://doi.org/10.1016/j.bspc.2021.102682
Received 19 February 2021; Received in revised form 9 April 2021; Accepted 24 April 2021
Available online 12 May 2021

produces high variance. To address this problem, Random Decision Forests are proposed. This Random Decision Forests classifier reduces the variance produced by the decision trees algorithm by majority voting for the final prediction. The correlation among decision trees is randomly decreased by choosing the features that recover the predictive power and create maximum performance. But random decision forests also cause uncertainty. So, hyper parameter tuning is the fundamental steps required on machine learning practice, like Random Decision Forests classifier to reduce uncertainty problems. Various methods have been evolved for breast cancer classification. But, there is a range for designing the suitable process to enlarge and execute the most effectual prognostic system of breast cancer. This motivated us to do the research area in breast cancer prognosis and diagnosis.

In this manuscript, the prognosis and diagnosis of breast cancer can be determined by combining the advantages of Kernel Neutrosophic C-Means Clustering Based Feature Weighting and Hyper Parameter Tuned Random Decision Forest classifier using Bayesian Optimization algorithm. This proposed kernel neutrosophic C-mean grouping-based feature weighting assigns maximum weights to relevant characteristics and minimum weights to less applicable features. This is different from selecting subsets of features and produces greater ranking precision to select subsets of features. In this work, Random Decision Forest classifier model are optimized with the help of Bayesian Optimization algorithm for getting optimal hyper tuning parameters to accurately prognosis and diagnosis of the breast cancer.

The main contributions of this manuscript are summarized as below:

- In this manuscript, combining the advantages of Kernel Neutrosophic C-Means Clustering Based Feature Weighting and Hyper Parameter Tuned Random Decision Forest classifier Using Bayesian Optimization algorithm are proposed Breast Cancer Prognosis and diagnosis.
- For Breast cancer Prognosis and diagnosis, Kernel Neutrosophic C-Means Clustering (KNCM) is used for feature weighting [11]. These weighted features are applied in the Random Decision Forest classifier (RDF).
- Generally Random Decision Forest classifier does not reveal the adoption of optimization techniques for computing the optimal parameters to ensure the accurate classification of breast cancer.
- Therefore in this work, the proposed Bayesian optimization algorithm [12] is utilized for optimizing the Random Decision Forest classifier [13].
- The proposed model is implemented in Python and efficiency of the proposed system is analyzed by evaluation metrics like balanced error rate, precision, recall, F-score, specificity, accuracy.
- Then comparison of evaluation metrics is analyzed with various existing feature weighting and classifier for breast cancer, namely feature weighting and Back propagation neural network with Ant lion Optimization algorithm (FW + ALO-BPNN) [14], feature weighting and support vector machine with Salp swarm Optimization algorithm (FW + SSA-SVM) [15], feature weighting and support vector machine with genetic Optimization algorithm (FW + GA-SVM) [16].
- Then the comparison of the evaluation metrics is analyzed with the various existing classifier, namely, Hybridized neural network and decision tree based classifier [17] and Random Forest-based rule extraction classifier [18].
- For demonstrating the efficiency of the proposed system in the prognosis and diagnosis of breast cancer, this manuscript performs a series of comparative studies by experimenting with two breast cancer data sets, i.e, the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [19] and Wisconsin prognostic Breast cancer dataset (WPBC) [20] obtained as UCI machine learning repository.

The rest of this manuscript is mentioned as follow. The Literature survey is described on section 2. Section 3 is about Proposed KNCM Based Feature Weighting and Hyper Parameter Tuned RDF-BOA for Breast Cancer Prognosis and diagnosis. Result and discussion are presented on section 4 and Lastly, Conclusions are presented on Section 5.

2. Literature survey

Among the numerous researches works on breast cancer Prognosis and diagnosis with different optimization algorithm, some of the most recent research works are reviewed here in this section.

Bhardwaj et al. [21] has presented a genetically optimized neural network (GONN) algorithm to perform the classification of breast cancer. This breast cancer classification system utilized a genetic algorithm (GA) for optimizing the artificial neural network parameter by implementing the novel crossover and mutation operators of GA. The calculation complexity also increases the breast cancer classification method. Furthermore, to perform real-life problems, it was necessary to evaluate the effectiveness of breast cancer techniques indicated by imaging systems such as mammography, ultrasound, and so on.

Acharya et al. [22] has presented the automatic characterization of malignant breast lesion. This technique utilizes shear wave elastography (SWE) to assess the discrete wave coefficients at three different levels. These features were removed as coefficients. The important characteristics were removed with sequential forward selection techniques and classified with a Relief classification system. The different classifiers for classifying benign and malignant lesions utilize the classified characteristics. This process reached classification accuracy of 93.59 %.

Sheikhpour et al. [23] have utilized PSO-KDE models that hybridize particle swarm optimization (PSO) and classifier based non-parametric kernel density estimation (KDE) for cancer diagnosis. Particle swarm optimization was utilized for determining the kernel bandwidth and chooses the subset of features on classifier based on kernel density estimation. The performance of classification and number of features selected were the principles utilized for designing the intention function of PSO-KDE.

Dora et al. [24] have suggested the algorithm based on Gauss-Newton representation for the classification of breast cancer. It utilizes sparse representation with selection of training samples. Until now, the sparse representation has been used effectively on pattern identification. This algorithm achieved the highest classification accuracy.

Phan et al. [25] have utilized a hybrid model of GA and SVM, for feature weighting and parameter optimization. The GA-SVM model reaches an important enhancement in sort performance across the entire data sets compared to Grid Search. Genetic algorithms (GA) are named as powerful tools to solve nonlinear optimization issues on maximum scale.

Fondon et al. [26] have presented an automatic classification of tissue malignancy for the diagnosis of breast carcinoma. Breast cancer was the second leading cause of cancer death among women. Their early diagnosis was utmost significance to avoid preventable deaths. Though, the assessment of malignancy of tissue biopsies was difficult and dependent on observer. The accuracy level ranges from 75.8 % when cross-validation was executed 5 times to 75 % with the exterior set of novel images and 61.11 %.

Raghavendra et al. [27] have suggested a growth of the papillary index of the breast for the differentiation of benign and malignant lesions with ultrasound imaging. Papillary lesions of the breast consist of a wide spectrum of pathologies ranging as benign to malignant. Papillary lesions of the breast have a variety of radiographic features on presentation; thus the differentiation among benign and malignant according to the characteristics of the image. The evolved model was analyzed with great collection of ultrasound images of papillary breast lesions.

Meiburger et al. [28] have utilized the detection of breast lesions with texton and features of local configuration patterns by ultrasound images. Breast cancer is the cancer that occurs most frequently on women throughout the world. When mammography leftovers the gold

standard on breast cancer detection, ultrasound was a significant imaging modality of detection and diagnosis of cancer. This includes the detection of breast lesions on ultrasound images with text filter banks, local configuration pattern characteristics and classification, without using any segmentation system.

3. Proposed KNCM based feature weighting and hyper parameter tuned RDF-BOA for breast Cancer

In this section, the accurate Prognosis and diagnosis of breast cancer is explained by combining the advantages of Kernel Neutrosophic C-Means Clustering Based Feature Weighting and Hyper Parameter Tuned Random Decision Forest Using Bayesian Optimization. The overall workflow of the proposed Breast Cancer Prognosis and diagnosis is given below in Fig. 1.

3.1. Input acquisition

In this step, the processed input data are taken from the breast cancer dataset to classify breast cancer [19,20]. The characteristics of dataset are multivariate and details of the statistical values of the attributes are given below in Table 1.

The output of feature extraction is given to the feature weighted input for providing relevant weight to the appropriate features. The Flowchart for Proposed KNCM Based Feature Weighting and Hyper Parameter Tuned RDF-BOA for Breast Cancer classification shows as Fig. 2.

Table 1
Detail information about attributes.

Attribute	Description	Mean value for dataset 1	Mean value for dataset 2
Feature 1	Thickness of clump	6.442	4.563
Feature 2	Uniformity of cell size	4.153	3.468
Feature 3	Uniformity of cell shape	3.954	4.123
Feature 4	Marginal adhesion	3.554	2.163
Feature 5	Single epithelial cell size	1.536	2.111
Feature 6	Bare nuclei	3.123	2.896
Feature 7	Bland chromatin	4.151	3.635
Feature 8	Normal nucleoli	1.966	1.123
Feature 9	Mitoses	1.713	1.963

3.2. Feature weighting using KNCM

In this process, feature weight is used for calculating the approximate optimal vector degree using the individual features in the training set based on the Truth degree, Intermediacy degree, Falsity degrees of the object membership values for the cluster. A feature weighting algorithm generally assigns the largest weights to the pertinent features and the smallest weights to less pertinent and redundant features. So feature weighting and feature subsets selection are different in nature, where feature weights are restricted and values should lies between 0 and 1. The Feature weighting works based on the principle of data clustering.

In this process, Data clustering categorized the input data into diverse category depends on few similarity features on breast region. The similarity can be determined with the help of correlation features. In this work, data clustering algorithm named Kernel Neutrosophic C-

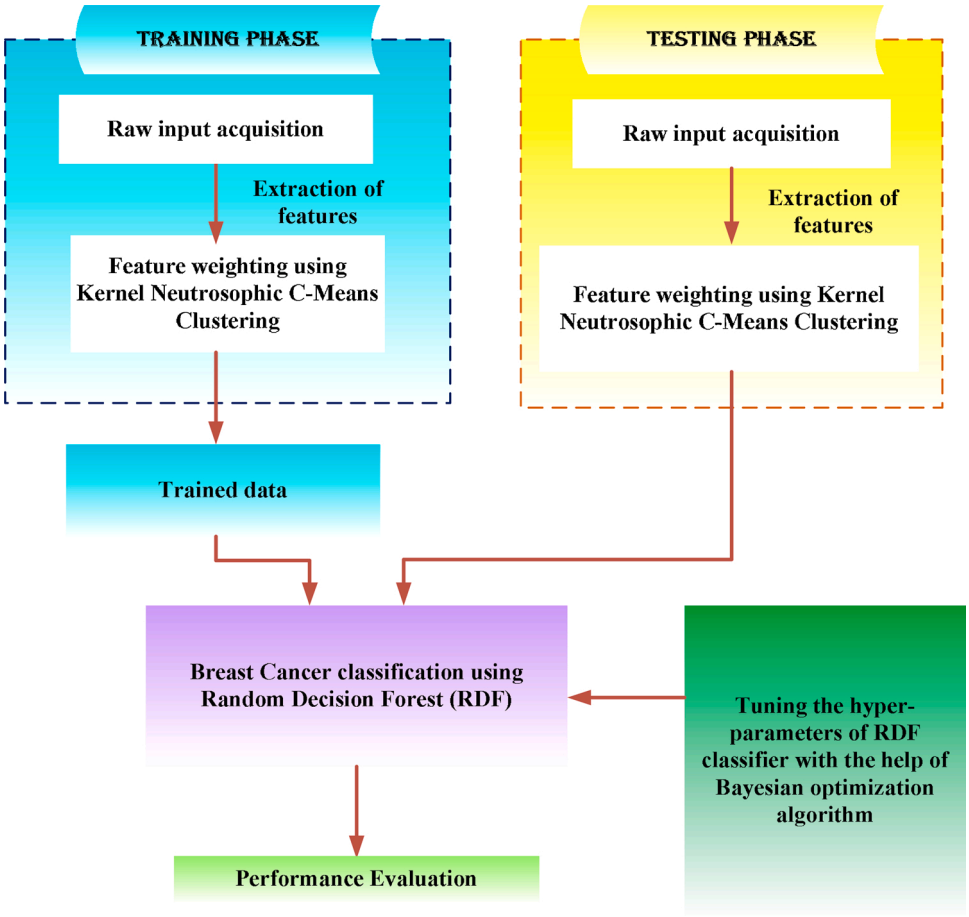

Fig. 1. Overall workflow for Breast Cancer classification.

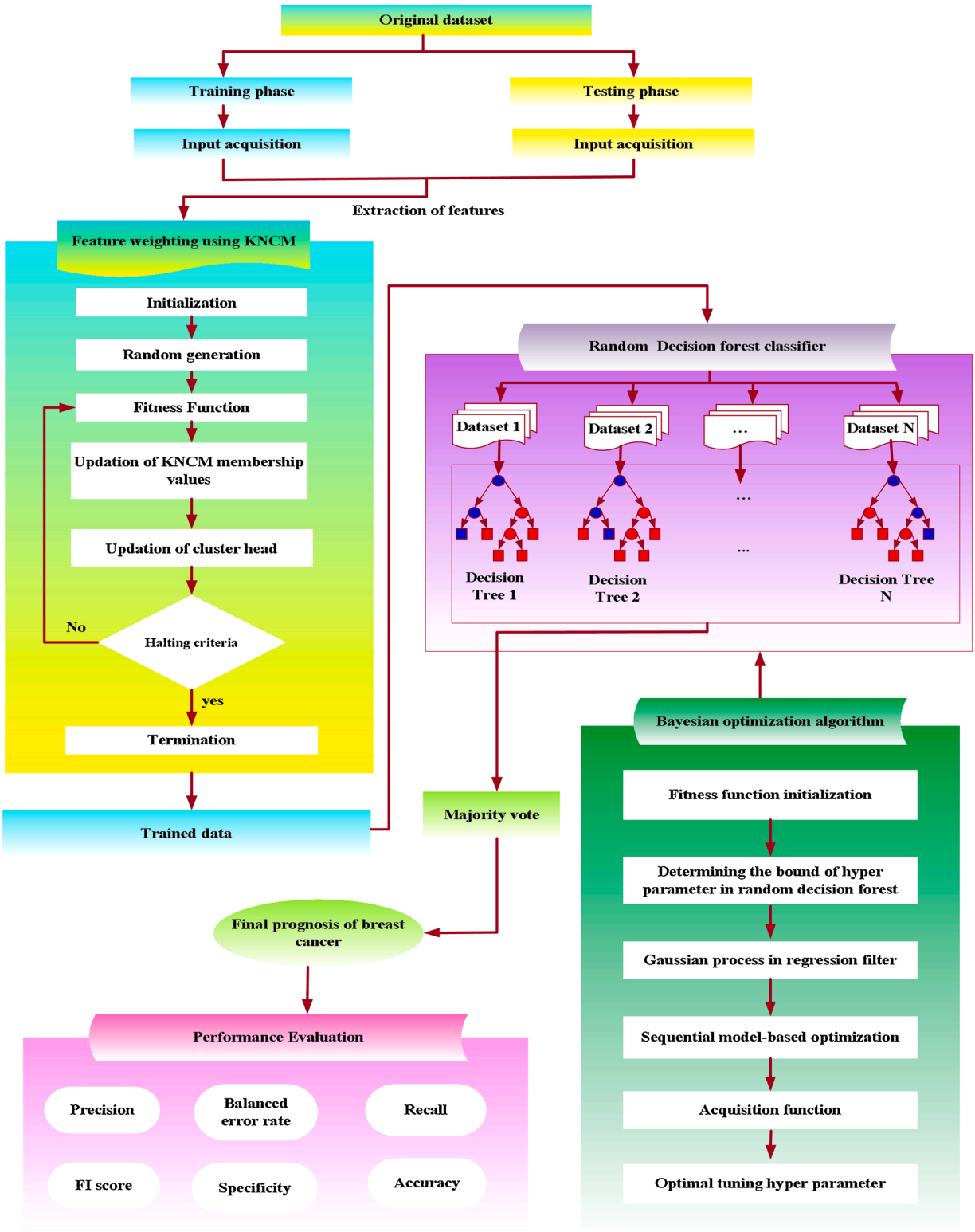

Fig. 2. Flowchart for Proposed KNCM Based Feature Weighting and Hyper Parameter Tuned RDF-BOA for Breast Cancer classification.

Means Clustering (KNCM) is used for feature weighting [11]. This Kernel Neutrosophic C-Means Clustering normally extracts the information from feature extraction. From the feature extraction information, it initializes the partition matrix and cluster centers. Then update the membership values of the Kernel Neutrosophic C-Means Clustering and clusters centers, which are used for determining the clustering object. The step by step procedure of the feature extraction based on Kernel Neutrosophic C-Means Clustering are given below

3.2.1. *Initialization*

In this step, the Kernel Neutrosophic C-Means Clustering partition matrix is initialized based on the following Eq. (1).

$$D_{p,q} = \begin{bmatrix} \vartheta_{1,1} & \cdots & \vartheta_{A,1} \\ \vdots & \ddots & \vdots \\ \vartheta_{1,n} & \cdots & \vartheta_{A,n} \end{bmatrix} \quad (1)$$

where feature weighting vector degree $\vartheta_{p,q} = [TR_{p,q}, IN_{p,q}, FA_{p,q}]$, based on the membership of object q to clusters p. Similarly $TR_{p,q}$ represents Truth degree of the object membership values for cluster. $IN_{p,q}$ represents intermediacy degree of the object membership values of cluster, $FA_{p,q}$ represents Falsity degrees of object membership values for cluster. Then similarly, the initial cluster centers $CC_P(Z)$ are initialized based on membership values of initial partition matrix are given below in the following Eq. (2).

$$CC_P(Z) = \frac{\sum_{p=1}^{A} [N_1 \times TR_{p,q}]^F \times [\vartheta_{p,q}]}{\sum_{p=1}^{A} [N_1 \times TR_{p,q}]^F} \quad (2)$$

where F represents the degree of Fuzziness, and the value of degree of Fuzziness as $F \geq 1$, Z is number of iteration, N_1, N_2, N_3 are weighting parameters, $\eth$ is a halting criteria and the values lies between 0 and 1.

3.2.2. Random generation

After the initialization process, the input parameters are randomly generated. In this step, the highest fitness values are chosen depend on fitness function. Here, randomly generate the population with Kernel Neutrosophic C-Means Clustering partition matrix $D_{p,q}$ and cluster centers $CC_P(Z)$.

3.2.3. Fitness function

A random number of solutions are generated from the initialized values. The fitness function can be evaluated by looping on iteration value Z which is used to minimize the objective function based on Gaussian kernel and given below in Eq. (3).

$$\begin{aligned} F(D_{p,q}) = & \sum_{p=1}^{A}\sum_{q=1}^{n} 2[N_1 \times TR_{p,q}]^F (1 - G(\vartheta_{p,q}, CC_P(Z)) \\ & + \sum_{p=1}^{A} 2[N_2 \times IN_{p,q}]^F (1 - G(\vartheta_{p,q}, CC_{PMAX}(Z)) \\ & + ED^2 \sum_{p=1}^{A} [N_3 \times FA_{p,q}]^F 2AX \; weighting \; parameter. \end{aligned} \quad (3)$$

where ED^2 is a Euclidean distance between object q and cluster p. G is a Gaussian kernel function and its value is one.

3.2.4. Updation of KNCM membership values

In this step, Kernel Neutrosophic C-Means Clustering (KNCM) membership values are updated with (Z-1)th iteration based on the Truth degree, Intermediacy degree, Falsity degrees of the object membership values for cluster are given below the following Eq. (4–6).

$$TR_{p,q} = \frac{N_2 N_3 [\vartheta_{p,q} - CC_P(Z)]^{-\frac{2}{F-1}}}{\sum_{q=1}^{n} [\vartheta_{p,q} - CC_P(Z)]^{-\frac{2}{F-1}} + [\vartheta_{p,q} - CC_{PMAX}(Z)]^{-\frac{2}{F-1}} + [ED]^{-\frac{2}{F-1}}} \quad (4)$$

$$IN_{p,q} = \frac{N_1 N_3 [\vartheta_{p,q} - CC_{PMAX}(Z)]^{-\frac{2}{F-1}}}{\sum_{q=1}^{n} [\vartheta_{p,q} - CC_P(Z)]^{-\frac{2}{F-1}} + [\vartheta_{p,q} - CC_{PMAX}(Z)]^{-\frac{2}{F-1}} + [ED]^{-\frac{2}{F-1}}} \quad (5)$$

$$FA_{p,q} = \frac{N_1 N_2 [ED]^{-\frac{2}{F-1}}}{\sum_{q=1}^{n} [\vartheta_{p,q} - CC_P(Z)]^{-\frac{2}{F-1}} + [\vartheta_{p,q} - CC_{PMAX}(Z)]^{-\frac{2}{F-1}} + [ED]^{-\frac{2}{F-1}}} \quad (6)$$

where $CC_{PMAX}(Z)$ is calculated with the help of cluster center with greatest $TR_{1(p,q)}$ and second greatest value $TR_{1(p,q)}$ of Truth degree of the object membership values for cluster and it is given below in the following Eq. (7).

$$CC_{PMAX}(Z) = \frac{TR_{1(p,q)} + TR_{1(p,q)}}{2} \quad (7)$$

3.2.5. Updation of cluster head

In this step, cluster head are updated with the iteration value based on the Eq. (2).

3.2.6. Termination

In this step, if Kernel Neutrosophic C-Means Clustering $TR_{p,q}(Z-1) - TR_{p,q}(Z) < \eth$, that means it satisfy the halting condition. Otherwise it goes step 4. Then allocate every data into class with greatest value based on the feature weighting $\vartheta_{p,q_{argmax}} = max[TR_{p,q}, IN_{p,q}, FA_{p,q}]$.

Finally, a set of feature weights are produced automatically in the Kernel Neutrosophic C-Means Clustering process, and these important features obtained can be used for building classification models.

3.3. Breast cancer classification using RDF-BOA

The Random Decision Forest is a supervised learning algorithm that is utilized in classification and regression. But the random decision forest algorithm is mainly used for classification problems. It is a technique that works with assembling numerous decision trees through the training phase and which receives the maximum votes is chosen by the random forest as the final decision. The decision tree is a tree-shaped diagram; every branch of the tree indicates the probable decision for determining the course of action. The main problem with Random decision forest algorithm during classification and detection is to find out the exact target features values from given set of training objects and their feature values. The exact target characteristics can be selected with the help of entropy, information gain, leaf node, decision node, and root node characteristics. The entropy is used to calculate the homogeneity of the data set. The information gain is a measure of the decrease on entropy after the data set is divided according to its target classification. The entropy can be calculated with the help of following Eq. (8).

$$Entropy = \sum_{a}^{z} -M_a \log_2 M_a \quad (8)$$

where a is the total information regarding samples from z occurences, z is a number of occurrences. M_a is the approximated number that the certain result will come out in z occurences. The information gain can be calculated with the help of following Eq. (9).

$$Information\ gain = High\ entropy - Low\ entropy \quad (9)$$

In the random decision forest classifier, the leaf node characteristic carries the classification or the decision result. Decision node has two or more branches. The topmost decision node is generally termed as root node characteristics. The problem statement regarding Random decision forest is to classify the different types of breast cancer Prognosis and diagnosis depends on different features. Normally the data are divided based on the highest information gain conditions. During dividing the entropy values has decreased significantly. If node has previously reached an entropy value of zero, then there is no need of node split for particular branch. Then based on other features, the node divided

process continuous till the value of entropy as 0. This is a way of predicting classification accuracy in random decision forest classifier. The random forest works based on the majority votes from different decision tree for accurate classification of breast cancer Prognosis and diagnosis. It conquers the over fitting issue by averaging or combining the outcomes of different decision trees. But the optimal parameters cannot be detected with the aid of the random decision forest classifier. So Artificial intelligence based optimization algorithm can be employed in the random decision forest classifier due to their availability, applicability and global perspective.

In this work, Bayesian optimization [12] can be utilized to optimize the random decision forest classifier for finding the optimal parameters (best number of trees and leaves per tree in the forest). Here Bayesian optimization is used for tuning the hyper parameters used in random decision forest classifier. Generally for hyper parameter tuning in machine learning, grid search, manual search and random search method are used for parameter configuration. But the search has its own drawbacks regarding time and there is no strategy-based informed search. So overcome this drawbacks, the Bayesian optimization Algorithm are used. The Bayesian optimization Algorithm assembled a probability model for finding the value, which automatically diminishes the objective function based on the precedent estimation outcome of the objective. In this paper, Bayesian optimization Algorithm is chosen because it has its own advantage; it requires less iteration than other tuning method like grid search and random search and it find out the optimal hyper parameters. The step by step procedure for random decision forest classifier using Bayesian optimization algorithm (RDF-BOA) are given below

3.3.1. Initialization

In this step, the appropriate parameter is initialized randomly in the initialization process, which is used for training the data and given below in the following equation (10).

$$Training\ data = \{ (r_1, s_1), \quad (r_2, s_2), \quad \ldots\ldots, \quad (r_w, s_w) \} \tag{10}$$

And it also initializes the hyper parameter settings δ and loss function LF.

3.3.2. Random generation

After the initialization process, the input parameters are randomly generated. In this step, the highest fitness values are chosen based on specific hyper-parameter setting.

3.3.3. Fitness function

The random number of solution is generated from the initialized values. The fitness function used to minimize the objective function based on the following Eq. (11).

$$fitness\ function\left(\frac{Y}{\delta}\right) = \begin{cases} D(\delta), & Y < Y^* \\ G(\delta), & Y \geq Y^* \end{cases} \tag{11}$$

where $D(\delta)$ represents the density estimation based on the loss value during observations, $G(\delta)$ is generated by the remaining observations value of the loss, Y represents the loss and it is given below in Eq. (13) and Y^* denotes the particular quantiles.

3.3.4. Sequential model-based optimization

Sequential model-based optimization is one of the concise forms of Bayesian optimization Algorithm (BOA) for tuning the random decision forest classifier hyper parameters. Sequential model-based optimization work based on the following procedure. First it finds the optimal hyper-parameter setting δ^* by building the Gaussian process GP_Z with sampled point and it can be determined with the help of the following Eq. (12).

$$\delta^* = argmin GP_{Z-1}(\delta) \tag{12}$$

Then find out the value of loss under the hyper-parameter setting δ^* and the loss value can be determined with the help of following Eq. (13).

$$Y = LF(\delta^*) \tag{13}$$

This corresponding loss value and the hyper-parameter setting δ^* are stored in the corresponding trails and it is denoted as CT.This corresponding trails are used for parameter settings and evaluations purposes. The Updation of CT are determined with the help of following Eq. (14)

$$CT = CT \cup (\delta^*, Y) \tag{14}$$

Then finally build the Gaussian process GP_Z model based on updated CT.

3.3.5. Acquisition function

The Acquisition Function of Bayesian optimization Algorithm (BOA) is used to determine the next iteration of the search process. In this paper, the expected improvement is chosen as acceptable performance criterion. The improvement can be determined with the help of Loss value and it is determined with following Eq. (15).

$$D(\delta) = \max(Y_{MIN} - Y(\delta), 0) \tag{15}$$

With the help of Eq. (15), the expected improvement can be determined with the help of following Eq. (16).

$$Exp\ Imp = \int_{-\infty}^{Y_{MIN}} \max(Y_{MIN} - Y(\delta), 0) \times fitness\ function\left(\frac{Y}{\delta}\right) dY \tag{16}$$

where Y_{MIN} denotes the minimum loss value in CT, $Y(\delta)$ represents hyper-parameter setting and δ^* denotes loss value.

3.3.6. Termination

In this step, the optimal hyper-parameter is selected in random decision forest classification with the help of the Bayesian optimization Algorithm (BOA). Finally random decision forest classifier classifies the Breast Cancer with the help of Bayesian optimization Algorithm.

4. Result and discussion

In this section, the simulation performance of combining the advantages of Kernel Neutrosophic C-Means Clustering based Feature Weighting and Hyper Parameter Tuned Random decision Forest is analyzed using Bayesian Optimization for Breast Cancer Prognosis and diagnosis. The simulations are conducted on PC with the Intel Core i5, 2.50 GHz CPU, 8GB RAM and Windows 7. The proposed method is simulated using the python. Here evaluation metrics like balanced error rate, precision, recall, F-score, specificity, accuracy are analyzed. The performance analyses are compared the proposed method with feature weighting and classification. First, the comparison of evaluation metrics is analyzed with the various existing feature weighting and classifier with optimization algorithm for breast cancer, namely feature weighting and Back propagation Neural Networks using Ant Lion Optimization of breast cancer classification (FW + ALO-BPNN) [14], Salp chain-based optimization of support vector machines and feature weighting for medical diagnostic information systems (FW + SSA-SVM) [15] and Feature weighting and SVM parameters optimization depends on genetic algorithms for classification (FW + GA-SVM) [16].

Then comparison of evaluation metrics is analyzed with the various existing classifier, namely Hybridized neural network and decision tree based classifier [17] and Random Forest-based rule extraction classifier [18]. The simulation parameters of the proposed algorithm are demonstrated at Table 2.

4.1. Dataset description

For comparison analysis, the proposed method is implemented into

Table 2
Simulation parameter.

Parameter	Value
Degree of Fuzziness	$F \geq 1$
weighting parameter N_1, N_2, N_3	0.6,0.3,0.1
Number of iteration Z	100
Halting criteria $\mathfrak{Z}$	0 to 1
Euclidean distance ED^2	2.24
Gaussian kernel G	1

two dataset. The two dataset utilized on this paper are Breast Cancer Wisconsin (Diagnostic) Data Set and Breast Cancer Wisconsin (Prognosis) Data Set.

4.1.1. Breast Cancer wisconsin (Diagnostic) data set

Features are taken as digitized image of a fine needle aspiration (FNA) of breast mass from UCI machine learning repository. The characteristics of multivariate data set depict the cell nuclei features values within the image. The data are taken from 569 persons. The features details are given in the Table 1 for classify the benign and malignant types of breast cancer. The Breast Cancer Wisconsin (Diagnostic) Data Set contain benign cases are 357 and malignant cases are 212.

4.1.2. Breast Cancer wisconsin (Prognosis) data set

The Breast Cancer Wisconsin (Prognosis) dataset, features are also extracted as digitized image of a fine needle aspirate of breast mass (FNA from UCI's machine learning repository). Here the data is taken from 198 people and the details about the characteristics are provided on Table 1. Breast Cancer Wisconsin (Prognosis) Data Set contain benign cases are 151 and malignant cases are 47.

4.2. Performance metrics

To measure the Precision, Recall, F-Measure, Accuracy, Specificity, Balanced Error rate, the confusion matrix are used. For measuring confusion matrix, True Negative, True Positive, False Negative and False Positive values are needed.

- True Positive (δ): Benign properly recognized into benign.
- True Negative (γ): Malignant properly recognized into malignant.
- False Positive (β): Malignant imperfectly recognized into benign.
- False Negative (α): Benign imperfectly recognized into malignant.

4.2.1. Precision

The precision are also called as Positive predictive values, can be determined with the help of Eq. (17)

$$\text{Precision} = \frac{\delta}{\delta + \beta} \tag{17}$$

4.2.2. Recall

Recall is also called as Sensitivity, can be determined with the help of Eq. (18)

$$\text{Recall} = \frac{\delta}{\delta + \alpha} \tag{18}$$

4.2.3. F-score

F score can be determined with the help of following Eq. (19)

$$\text{F} - \text{score} = \frac{\delta}{\delta + \frac{1}{2(\beta+\gamma)}} \tag{19}$$

4.2.4. Accuracy

The accuracy values can be determined with the help of following Eq. (20)

$$\text{Accuracy} = \frac{\delta + \gamma}{\delta + \gamma + \beta + \alpha} \tag{20}$$

4.2.5. Specificity

Specificity is also named as true negative rate, which can be determined with the help of equation (21)

$$\text{Specificity} = \frac{\gamma}{\gamma + \beta} \tag{21}$$

4.2.6. Balanced error rate

The balanced error rate can be determined with the help of following Eq. (22)

$$\textit{Balanced error rate} = 1 - 0.5 \times \frac{\text{Recall} + \text{Specificity}}{100} \tag{22}$$

4.3. Comparison of performance analysis with various feature weighting and classifier for breast cancer

In this section, the prognosis and diagnosis of breast cancer is analyzed with the aid of two data set such as Breast Cancer Wisconsin (Diagnostic) Data Set [19], Breast Cancer Wisconsin (Prognosis) Data Set [20]. Table 3 shows the data set for 70 % training and the remaining

Table 3
70 % training result for Breast Cancer.

Feature weighting and classifier with optimization algorithm for breast cancer		Wisconsin Prognosis dataset	Wisconsin Diagnosis dataset
FW + ALO-BPNN	Precision	0.97	0.72
	Recall	0.96	0.75
	F-Measure	0.97	0.72
	Training Accuracy	0.989	0.891
	Testing Accuracy	0.964	0.75
	Specificity	0.925	1
	Balanced Error rate	0.035	0.16
FW + SSA-SVM	Precision	0.94	0.6
	Recall	0.94	0.6
	F-Measure	0.94	0.6
	Training Accuracy	1	1
	Testing Accuracy	0.941	0.6
	Specificity	0.925	1
	Balanced Error rate	0.035	0.16
FW + GA-SVM	Precision	0.96	0.52
	Recall	0.96	0.67
	F-Measure	0.96	0.59
	Training Accuracy	0.994	0.949
	Testing Accuracy	0.959	0.666
	Specificity	0.925	1
	Balanced Error rate	0.035	0.16
FW-KNCM + BOA-RDF (Proposed)	Precision	0.97	0.84
	Recall	0.96	0.8
	F-Measure	0.97	0.75
	Training Accuracy	1	0.992
	Testing Accuracy	0.964	0.8
	Specificity	0.925	1
	Balanced Error rate	0.035	0.16

30 % of the tests for the Breast Cancer Wisconsin (prognosis) and Breast Cancer Wisconsin (diagnosis) data sets, respectively. Table 4 shows the data set for 75 % training and remaining 25 % for both Breast Cancer Wisconsin (Prognosis) and Breast Cancer Wisconsin (Diagnosis) Data Set, respectively.

From Table 3, shows dataset for 70 % training and 30 % testing, the precision analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 16.66 %, 40 % and 61.53 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Recall analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 6.66 %, 33.33 % and 19.4 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The F-Measure analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 4.16 %, 25 % and 27.11 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Training Accuracy analysis of proposed FW + BOA-RDF for Breast Cancer Wisconsin (Prognosis) Data Set is 11.35 % and 4.53 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Testing Accuracy of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 6.66 %, 33.33 % and 20.12 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively.

From Table 3, shows 70 % training and 30 % testing, precision analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 0.001 %, 3.19 % and 1.04 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Recall analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 0.0001 %, 2.12 % and 0.001 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The F-Measure analysis of proposed FW + BOA-RDF in Bayesian Optimization for Breast Cancer Wisconsin (Diagnostic) Data Set is 0.001 %, 3.19 % and 1.04 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Training Accuracy analysis of proposed FW + BOA-RDF for the Breast Cancer Wisconsin (Diagnostic) Data Set is 1.112 % and 0.6 % higher than existing methods like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Testing Accuracy of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 0.0001 %, 2.44 % and 0.521 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively.

From Table 4, shows 75 % training and 25 % testing, the precision analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 26.08 %, 7.44 % and 10.30 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Recall analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 16.66 %, 31.25 % and 10.52 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The F-Measure analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 15.71 %, 26.56 % and 15.71 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Training Accuracy analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 13.09 % and 0.7 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Testing Accuracy of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 13.88 %, 28.12 % and 7.89 % higher than existing method FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively.

From Table 4, shows 75 % training and 25 % testing, precision analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 3.15 %, 4.25 % and 1.03 % higher than existing method FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Recall analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 3.15 %, 4.25 % and 1.03 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The F-Measure analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 3.15 %, 4.25 % and 3.15 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Training Accuracy analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 1.0101 % and 0.502 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. The Testing Accuracy of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 2.944 %, 4.482 % and 1.45 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively.

Fig. 3 shows the Roc curve for Feature weighting and classifier with optimization algorithm for breast cancer classification. Here the ROC of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Diagnostic) Data Set is 3.157 %, 4.2553 % and 2.0833 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively.

Table 4
75 % training result for Breast Cancer.

Feature weighting and classifier with optimization algorithm for breast cancer		Wisconsin Prognosis dataset	Wisconsin Diagnosis dataset
FW + ALO-BPNN	Precision	0.95	0.69
	Recall	0.95	0.72
	F-Measure	0.95	0.7
	Training Accuracy	0.99	0.878
	Testing Accuracy	0.951	0.72
	Specificity	0.946	1
	Balanced Error rate	0.02	0.16
FW + SSA-SVM	Precision	0.94	0.64
	Recall	0.94	0.64
	F-Measure	0.94	0.64
	Training Accuracy	1	1
	Testing Accuracy	0.937	0.64
	Specificity	0.946	1
	Balanced Error rate	0.02	0.16
FW + GA-SVM	Precision	0.97	0.74
	Recall	0.97	0.76
	F-Measure	0.97	0.7
	Training Accuracy	0.995	0.986
	Testing Accuracy	0.965	0.76
	Specificity	0.946	1
	Balanced Error rate	0.02	0.16
FW-KNCM + BOA-RDF (Proposed)	Precision	0.98	0.87
	Recall	0.98	0.84
	F-Measure	0.98	0.81
	Training Accuracy	1	0.993
	Testing Accuracy	0.979	0.82
	Specificity	0.946	1
	Balanced Error rate	0.02	0.16

4.4. Classification error percentage

In this section, the classification error percentage is computed based on the ratio between the total number of classified samples of breast cancer and the incorrectly classified samples. The classification error percentage comparison is given in Table 5.

The proposed FW + BOA-RDF with the various existing Feature weighting and classifier for breast cancer, namely FW + ALO-BPNN [14], FW + SSA-SVM [15] and FW + GA-SVM [16] respectively. The

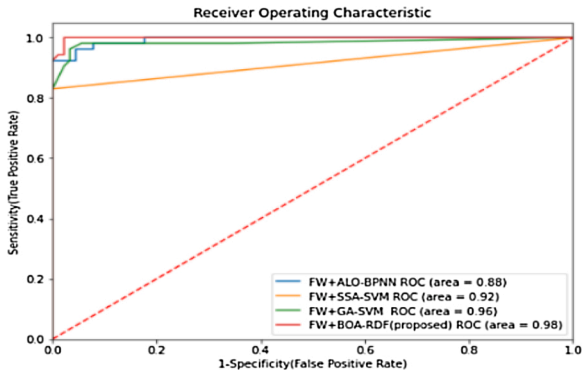

Fig. 3. Roc curve for Feature weighting and classifier with optimization algorithm for breast cancer classification.

Table 5
Comparison of Classification Error percentage.

Classifier	Classification Error percentage
FW + ALO-BPNN	0.048951049
FW + SSA-SVM	0.062937063
FW + GA-SVM	0.034965035
FW + BOA-RDF (proposed)	0.020979021

classification error percentage is calculated with the help of python. From Table 5, the Classification Error percentage analysis of proposed FW + BOA-RDF is 57.14 %, 66.66 % and 40 % lower than existing method like FW + ALO-BPNN [14], FW + SSA-SVM [15] and FW + GA-SVM [16] respectively.

4.5. Comparison of performance analysis with various classifier used for breast cancer

In this section, the performance like accuracy, precision, recall and f-measure analysis is compared with various classifiers used in breast cancer classification. Here the proposed Random Decision Forest based Bayesian Optimization classifier is compared with Hybridized neural network and decision tree based classifier and Random Forest-based rule extraction classifier. Figs. 4–8 demonstrates that performance analysis of various classifiers used for breast cancer.

Fig. 4 illustrates that accuracy analysis of various classifier used for breast cancer. Accuracy of proposed Random Decision Forest Classifier with Bayesian Optimization for Breast Cancer (RDF-BOA) is 5.376 %, and 3.1578 % higher than existing method like Hybridized neural network-decision tree based classifier (HNN-DTC) and Random Forest-based rule extraction classifier (RF-REC) respectively.

Fig. 5 demonstrates that F-measure analysis of various classifiers

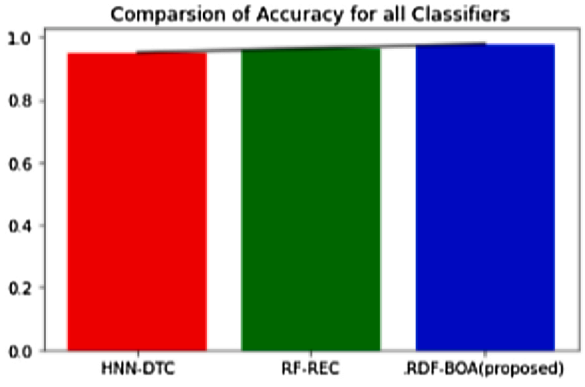

Fig. 4. Accuracy analysis.

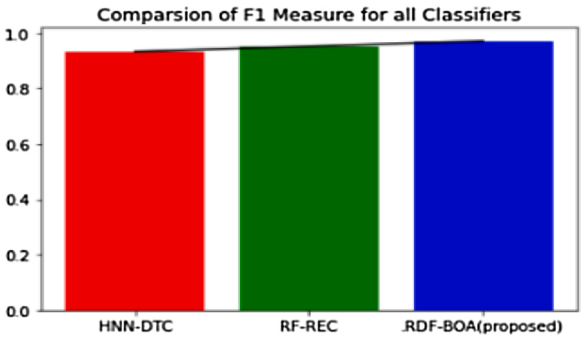

Fig. 5. F-measure analysis.

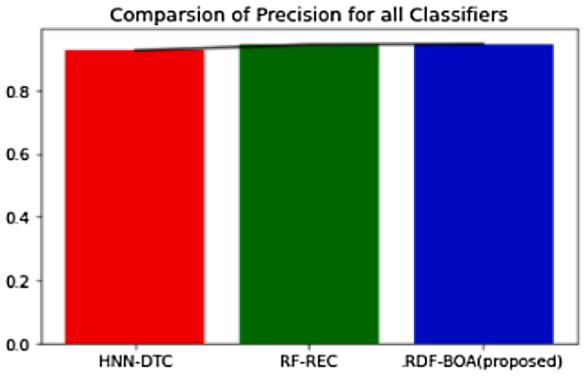

Fig. 6. Precision analysis.

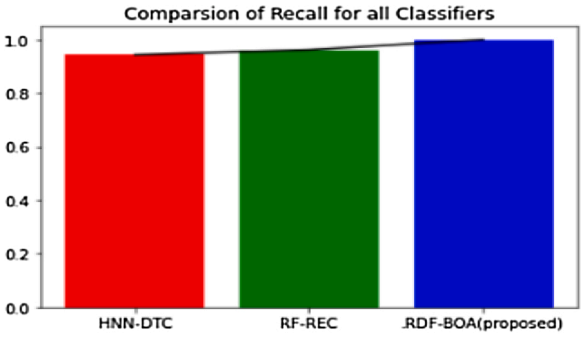

Fig. 7. Recall analysis.

used for breast cancer. The F-measure of proposed Random Decision Forest Classifier with Bayesian Optimization for Breast Cancer (RDF-BOA) is 8.1489 %, and 5.319 % higher than existing method like Hybridized neural network-decision tree based classifier (HNN-DTC) and Random Forest-based rule extraction classifier (RF-REC) respectively.

Fig. 6 demonstrates that precision analysis of various classifiers used for breast cancer. The precision of proposed Random Decision Forest Classifier with Bayesian Optimization for Breast Cancer (RDF-BOA) is 8.152 %, and 1.632 % higher than existing method like Hybridized neural network-decision tree based classifier (HNN-DTC) and Random Forest-based rule extraction classifier (RF-REC) respectively.

Fig. 7 portrays that Recall analysis of various classifiers used for breast cancer. The Recall of proposed Random Decision Forest Classifier with Bayesian Optimization for Breast Cancer (RDF-BOA) is 11.11 %, and 5.263 % higher than existing method like Hybridized neural network-decision tree based classifier (HNN-DTC) and Random Forest-based rule extraction classifier (RF-REC) respectively.

Fig. 8 portrays that ROC analysis of various classifiers used for breast cancer. The ROC of proposed Random Decision Forest Classifier with

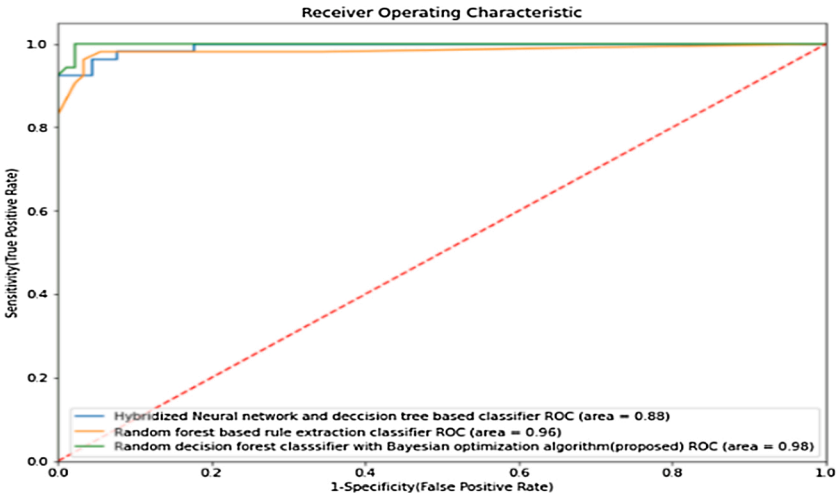

Fig. 8. Roc Curve.

Bayesian Optimization for Breast Cancer (RDF-BOA) is 11.3636 %, and 2.0833 % higher than existing method like Hybridized neural network-decision tree based classifier (HNN-DTC) and Random Forest-based rule extraction classifier (RF-REC) respectively.

5. Conclusion

In this manuscript, combining the advantages of Feature Weighting and Hyper Parameter Tuned Random Decision Forest classifier are accurately Prognosis and diagnosis of the breast cancer. Here Kernel Neutrosophic C-Means Clustering is used as Feature Weighting. Then Random Decision Forest classification model are optimized with the help of the Bayesian Optimization algorithm to obtain optimal hyper tuning parameters. In this manuscript, feature weight and optimal classification are provided the best results based on prognosis and diagnosis. During the performance analysis of the Wisconsin prognostic Breast Cancer (WPBC) dataset, 70 % training and remaining 30 % testing is compared with the WDBC dataset, the precision analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 16.66 %, 11.53 % and 58.16 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. Similarly the performance analysis of the Wisconsin prognostic Breast Cancer (WPBC) dataset, 75 % training and remaining 25 % testing is compared with the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, the precision analysis of proposed FW + BOA-RDF in Breast Cancer Wisconsin (Prognosis) Data Set is 7.27 %, 2.75 % higher than existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM respectively. Finally, the simulation outcomes demonstrate that the proposed method generates less error 57.14 %, 66.66 % and 40 % lower than the existing method like FW + ALO-BPNN, FW + SSA-SVM and FW + GA-SVM classifier on breast cancer Prognosis and diagnosis, which carries a significant role on clinical radiological diagnosis.

Funding information

None.

CRediT authorship contribution statement

Pratheep Kumar P: Conceptualization, Methodology, Supervision. **Mary Amala Bai V:** Writing - original draft. **Geetha G. Nair:** Writing - review & editing.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] A.T. Boraei, P.K. Singh, M. Sechi, S. Satta, Discovery of novel functionalized 1, 2, 4-triazoles as PARP-1 inhibitors in breast cancer: Design, synthesis and antitumor activity evaluation, Eur. J. Med. Chem. 182 (2019) 111621.
[2] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, U.R. Acharya, Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images, Pattern Recognit. Lett. (2020).
[3] Y. Guo, Y. Hu, M. Qiao, Y. Wang, J. Yu, J. Li, C. Chang, Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma, Clin. Breast Cancer 18 (3) (2018) e335–e344.
[4] M. Inoue, H. Nakagomi, H. Nakada, K. Furuya, K. Ikegame, H. Watanabe, M. Omata, T. Oyama, Specific sites of metastases in invasive lobular carcinoma: a retrospective cohort study of metastatic breast cancer, Breast Cancer 24 (5) (2017) 667–672.
[5] Y. An, J.R. Adams, D.P. Hollern, A. Zhao, S.G. Chang, M.S. Gams, P.E. Chung, X. He, R. Jangra, J.S. Shah, J. Yang, Cdh1 and Pik3ca mutations cooperate to induce immune-related invasive lobular carcinoma of the breast, Cell Rep. 25 (3) (2018) 702–714.
[6] S. Al-Mahmood, J. Sapiezynski, O.B. Garbuzenko, T. Minko, Metastatic and triple-negative breast cancer: challenges and treatment options, Drug Deliv. Transl. Res. 8 (5) (2018) 1483–1507.
[7] M.A. Mohammed, B. Al-Khateeb, A.N. Rashid, D.A. Ibrahim, M.K. Abd Ghani, S. A. Mostafa, Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images, Comput. Electr. Eng. 70 (2018) 871–882.
[8] R. Rodrigues, S. Geyl, J. Albouys, C. De Carvalho, M. Crespi, T. Tabouret, A. Taibi, S. Durand-Fontanier, R. Legros, M. Dahan, P. Carrier, Effect of implementing a regional referral network on surgical referral rate of benign polyps found during a colorectal cancer screening program: a population-based study, Clin. Res. Hepatol. Gastroenterol. (2020).
[9] T. Fujioka, K. Kubota, M. Mori, Y. Kikuchi, L. Katsuta, M. Kasahara, G. Oda, T. Ishiba, T. Nakagawa, U. Tateishi, Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network, J. Radiol. 37 (6) (2019) 466–472.
[10] M.M. Ghiasi, S. Zendehboudi, Application of decision tree-based ensemble learning in the classification of breast Cancer, Comput. Biol. Med. (2020) 104089.
[11] Y. Akbulut, A. Şengür, Y. Guo, K. Polat, KNCM: kernel neutrosophic c-means clustering, Appl. Soft Comput. 52 (2017) 714–724.
[12] H.M. Torun, M. Swaminathan, A.K. Davis, M.L.F. Bellaredj, A global Bayesian optimization algorithm and its application to integrated system design, Ieee Trans. Very Large Scale Integr. 26 (4) (2018) 792–802.
[13] B.A. Akram, A.H. Akbar, O. Shafiq, HybLoc: hybrid indoor Wi-Fi localization using soft clustering-based random decision forest ensembles, IEEE Access 6 (2018) 38251–38272.
[14] S. Dalwinder, S. Birmohan, K. Manpreet, Simultaneous feature weighting and parameter determination of Neural Networks using Ant Lion Optimization for the classification of breast cancer, Biocybern. Biomed. Eng. 40 (1) (2020) 337–351.
[15] A.Z. Ala'M, A.A. Heidari, M. Habib, H. Faris, I. Aljarah, M.A. Hassonah, Salp chain-based optimization of support vector machines and feature weighting for medical diagnostic information systems. Evolutionary Machine Learning Techniques, Springer, Singapore, 2020, pp. 11–34.

[16] A.V. Phan, M. Le Nguyen, L.T. Bui, Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems, Appl. Intell. 46 (2) (2017) 455–469.
[17] A. Suresh, R. Udendhran, M. Balamurgan, Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers, Soft comput. 24 (11) (2020) 7947–7953.
[18] S. Wang, Y. Wang, D. Wang, Y. Yin, Y. Wang, Y. Jin, An improved random forest-based rule extraction method for breast cancer diagnosis, Appl. Soft Comput. 86 (2020) 105941.
[19] https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic).
[20] http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Prognostic).
[21] A. Bhardwaj, A. Tiwari, H. Bhardwaj, A. Bhardwaj, A genetically optimized neural network model for multi-class classification, Expert Syst. Appl. 60 (2016) 211–221.
[22] U.R. Acharya, W.L. Ng, K. Rahmat, V.K. Sudarshan, J.E. Koh, J.H. Tan, Y. Hagiwara, C.H. Yeong, K.H. Ng, Data mining framework for breast lesion classification in shear wave ultrasound: a hybrid feature paradigm, Biomed. Signal Process. Control 33 (2017) 400–410.
[23] R. Sheikhpour, M.A. Sarram, R. Sheikhpour, Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer, Appl. Soft Comput. 40 (2016) 113–131.
[24] L. Dora, S. Agrawal, R. Panda, A. Abraham, Optimal breast cancer classification using Gauss–Newton representation based algorithm, Expert Syst. Appl. 85 (2017) 134–145.
[25] A.V. Phan, M. Le Nguyen, L.T. Bui, Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems, Appl. Intell. 46 (2) (2017) 455–469.
[26] I. Fondón, A. Sarmiento, A.I. García, M. Silvestre, C. Eloy, A. Polónia, P. Aguiar, Automatic classification of tissue malignancy for breast carcinoma diagnosis, Comput. Biol. Med. 96 (2018) 41–51.
[27] U. Raghavendra, J.E.W. Koh, A. Gudigar, W.Y. Chan, M.T.R. Hamid, K. Rahmat, F. Fadzli, K.H. Ng, C.P. Ooi, E.J. Ciaccio, H. Fujita, Development of breast papillary index for differentiation of benign and malignant lesions using ultrasound images, J. Ambient Intell. Humaniz. Comput. (2020) 1–9.
[28] U.R. Acharya, K.M. Meiburger, J.E.W. Koh, E.J. Ciaccio, N. Arunkumar, M.H. See, N.A.M. Taib, A. Vijayananthan, K. Rahmat, F. Fadzli, S.S. Leong, A novel algorithm for breast lesion detection using textons and local configuration pattern features with ultrasound imagery, IEEE Access 7 (2019) 22829–22842.