Indian Journal of Science and Technology, Vol 9(14), DOI: 10.17485/ijst/2016/v9i14/76802, April 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Breast Cancer Detection using a Neuro-fuzzy based Classification Method

Soumadip Ghosh[1*], Sushanta Biswas[2], Debasree Chanda (Sarkar)[2] and Partha Pratim Sarkar[2]

[1]Department of CSE, Academy of Technology, Hooghly - 712121, West Bengal, India; soumadip.ghosh@gmail.com
[2]Department of Engineering and Technological Studies, University of Kalyani, Nadia - 741245, West Bengal, India

Abstract

Background/Objectives: Detection and analysis of critical diseases such as breast cancer is a significant domain of data mining analysis and research. In this research study, we propose a neuro-fuzzy classification method for breast cancer detection. **Methods/Statistical Analysis:** The proposed neuro-fuzzy method considers the pattern-wise degree of memberships of breast cancer databases to the existing data classes that are accomplished using a fuzzification method. The method produces a membership matrix with an element count identical to the product of the number of data records and data classes present. These matrix elements are then used as input to a neural network. **Findings:** We apply our method to three UCI databases namely WBC, WDBC and MM. The research work aims to recognize breast cancer disease using the proposed method and then compare its performance with two well-known classification algorithms namely Multilayer Perceptron and Support Vector Machine. We use here 10-fold cross validation technique for performing simulation. Different measures, for instance, root-mean-square error, kappa statistic, accuracy, false-positive rate, true-positive rate, precision, recall and f-measure are used to perform numerical analysis of the simulated results. All these evaluation measures support the supremacy of our proposed method. **Application/Improvements:** The suggested method has great potential in terms of classification capability and predictive power to use in the fields of Medical Science and Bioinformatics.

Keywords: Breast Cancer, Classification, Data Mining, Multilayer Perceptron, Neuro-Fuzzy, Support Vector Machine

1. Introduction

Data mining[1,2] applications are becoming popular in the Medical Science and the Bioinformatics research field for diagnosis of critical diseases. Breast cancer[3] has perhaps become an exceedingly intense subject for determining cures apart from AIDS in the present decade. It is a kind of cancer disease rising from the breast tissue cells and the disease is more usual in adult females than in adult males.

Furthermore, the disease is very destructive to women since it can lead to the loss of a breast or the consequence is much more severe. A medical survey[4] conducted by experts reveals that the disease occurs in 22.9% of all types of cancers in the women and it also results in 13.7% of demises from them. Medical diagnosis and existence rates for the disease mostly are determined by the type and period of cancer, method of treatment and geographical location of the patient.

The diagnosis of breast cancer disease can be done by classification which is an important branch of data mining. It is essentially the method of determining a classifier or model that describes and discriminates several data classes from each other. Initially, the classification procedure applies some pre-processing tasks (data cleaning, data transformation etc.) to the original data. Then, the method divides this pre-processed data set into two different sections specifically the training set and the test set. They should be independent of each other to avoid biases. Classification consists of two different steps. The initial step develops a classification model indicating a

*Author for correspondence

well-defined set of classes. Therefore, this is the training phase, where the goal is to construct a model by reviewing a specified training set accompanied by their related class label attributes. After that, the classification model is applicable for prediction called the testing phase. This step estimates the accuracy of the derived model using the test set.

Artificial Neural Network (ANN)[6–8] means to develop an artificial system that could achieve human brain like reasoning capability. It is widely known for excellent precision and extraordinary learning capability even if the minor information is accessible. Multilayer Perceptron (MLP)[9,10] model is one of the proficient methods of classification from the ANN domain. An MLP model contains numerous layers of nodes organized in a directed graph structure, with connections between the adjacent layers. MLP uses the back propagation technique to train and build the neural network model.

Support Vector Machine (SVM)[11] is another powerful model that may construct single hyper plane or multiple hyper planes in a high-dimensional feature space for regression analysis, classification task or other kinds of analysis tasks. SVM utilizes a hyper plane to discriminate between classes. When classes intersect with each other, the common idea is to employ a hyper plane for minimizing the error of data points alongside the borderline between data classes; the points are called the support vectors.

Due to the presence of ambiguity in the training dataset, overlying borders among classes and vagueness in describing characteristics some uncertainties can still occur at any stage of a classification system. The fuzzy set[12–14] theory approach is flexible enough to manage the various facets of uncertainties about real life situations. ANN and fuzzy set theory based combined approach is the neuro-fuzzy[15,16] technique. A typical neuro-fuzzy classification based approach exploits these two techniques in an efficient way. It combines the human like rational thinking of fuzzy based systems with the connective structure and the learning ability of ANNs to develop a Neuro-Fuzzy System (NFS).

The present study proposes a NFS model for detection of breast cancer. We use NFS, MLP and SVM classifiers for performance comparison using three benchmark databases of the University of California, Irvine (UCI). These databases are namely Wisconsin Breast Cancer (WBC), Wisconsin Diagnostic Breast Cancer (WDBC) and Mammographic Mass (MM). We investigate classifier performances using measures such as Root-Mean-Square Error (RMSE), accuracy, False-Positive Rate (FP-Rate), True-Positive Rate (TP-Rate), kappa statistic, recall, precision and f-measure.

In[17] presented the "rule-based expert systems" to perform clinical diagnosis. The study inspired the utilization of knowledge in Medicine and Bioinformatics.

In[18] developed a decision tree-based method to breast cancer prediction. It took a group of people involved in tobacco and alcohol consumption in their past lives. The study demonstrated that several complex genetic variations are responsible for several types of cancer diseases.

In[19] designed a strong analytical model to forecast breast cancer durability using a training set that comprised gene signature types from the cancer patients. They primarily recognized several such signature types called "attractor meta-genes" in an investigation of various tumour types. They also tested these signatures using a computing framework named "Sage Bionetworks" to demonstrate the usefulness of this method.

In[20] used different ANN-based models to detect and analyze the disease of breast cancer. They applied four different classifiers of neural network, specifically Multi-Layer Perceptron (MLP), Radial Basis Function Network (RBFN), Leaning Vector Quantization (LVQ) and Competitive Learning Network (CLN). The simulated results showed that LVQ was the best performer using the test dataset, followed by other classifiers such as CLN, MLP and RBFN.

In[21] analyzed procedures for extracting strong fuzzy rules using a familiar neuro-fuzzy based software tool named "NEFCLASS". They designed a rule base for breast cancer diagnosis consisting of 9 rules and the positive and negative predictive values of this rule base were 75% and 93% respectively.

In[22] combined neural network and fuzzy rule-based system in their proposed design. The research study proved that a combined approach could be beneficial for performance improvement of data analysis and decision-making systems.

In[23] a software-based tool named "Breast Cancer Risk Evaluator" was established to estimate the risk of breast cancer by processing the information systems database of numerous patients. After information processing, the system generated vital information to forecast the present stage of cancer in the patient.

In[24] statistical information was taken out from the Region of Interest (known as ROI) in the parenchymal

region of breast. It employed K-NN classifier through three dissimilar distance metrics namely Euclidean, Cosine, City-block and their combination to perform classification. These features are fed into the K-NN classifier to categorize the ROI into any of the three breast tissue classes like 'dense', 'fatty' and 'glandular'. The accuracy obtained for combined K-NN was 91.16%.

In[25] an attempt was made to develop a diagnosis tool to detect breast cancer by wavelet transform. The tool was proposed to perform multi-scalar contrast improvement at dissimilar scales of wavelet. The procedure was sustained by means of segmentation techniques that combine details of the core image structure and its boundaries. The procedure approved the use of an algorithm for image processing and tried to detect together micro-classification and masses in the breast tissue image.

In[26] analyzed a breast cancer dataset using four decision tree classifiers namely J48 (C4.5), Classification and Regression Trees (CART), Alternating Decision Tree (AD-Tree) and Best First Tree (BF-Tree). Based on the simulated results, the performance of J48 was better compared to other algorithms for given chosen data set.

In[27] proposed a technique of breast cancer diagnosis that covered five stages of breast cancer detection based on mammography. Neural networks were used to train the system to detect cancer using a dataset for breast cancer patients. The combination of multiple methods could solve the problems of breast cancer detection with a higher degree of precision.

In[28] mammogram images were developed from real-time and standard imaging databases for the patients. The major goals of the proposed methods were to diagnose breast cancer using Fuzzy Enhanced Mammogram Segmentation schemes named FEM1 and FEM2 by Matlab programming environment. For the evaluation of performance, statistical measures such as Similarity Index (SI), Correct Detection Ratio (CDR) and Under Segmentation Error (USE) were evaluated. Based on the result, it was found that the CDR for FEM1 was 87% whereas FEM2 exhibited only 77% and also took 6.25 times lesser execution time. Lastly, Support Vector Method (SVM) was used to categorize whether the mammogram under test was 'normal' or 'abnormal'. As a result, FEM1 outperformed other classifiers.

The research studies performed in[29,30] also added immeasurably in the diagnosis of breast cancer disease using several data modelling techniques.

2. Proposed Neuro-Fuzzy Classification Method

In this research study, we propose a neuro-fuzzy based classification method to detect the breast cancer disease. The study uses the pattern-wise degree of belongings of the breast cancer data patterns (i.e., data records) to all the classes that are attained using a fuzzification procedure. The present work is an extension of the previous research work as specified in[31]. We have used a sigmoidal Membership Function (MF) for fuzzification. The fuzzification procedure creates a membership matrix with an overall number of elements identical to the product of the number of data patterns and data classes present in the database. The proposed neuro-fuzzy classification method divides into three different phases described below.

2.1 Phase 1

The first phase named fuzzification phase takes a sample dataset comprising multiple data patterns, fuzzifies the data pattern values with a sigmoidal MF and then computes the degree of membership of individual patterns to various classes. Let us consider that we have a dataset consisting of Q input patterns and P data classes. We now define the dataset in terms of the input pattern vector z where 'T' is the matrix transpose operator:

$$z = \left[z_1, z_2, ..., z_Q \right]^T \tag{1}$$

The phase essentially builds a membership matrix of order $(Q \times P)$ from the input pattern vector z that consists of the degree of the memberships of Q different patterns to P number of classes. Each item in this matrix is a membership function of the form $m_{i,j}(z_i)$, where z_i is the *i-th* feature value of the pattern vector z with indices $i = 1, 2, \ldots, Q$ and $j = 1, 2, \ldots, P$. Thus, we can describe the membership function as:

$$m_{i,j}(z_i) = \text{degree of membership of the input pattern } i \text{ to the} \tag{2}$$

As already stated, we have used the well-known sigmoidal MF for fuzzification. It is asymmetric in nature and depends on two different parameters a and b as specified by the equation:

$$m_{i,j}(z_i) = m_{i,j}(z_i; a, b) = \frac{1}{1 + e^{-a(z_i - b)}} \tag{3}$$

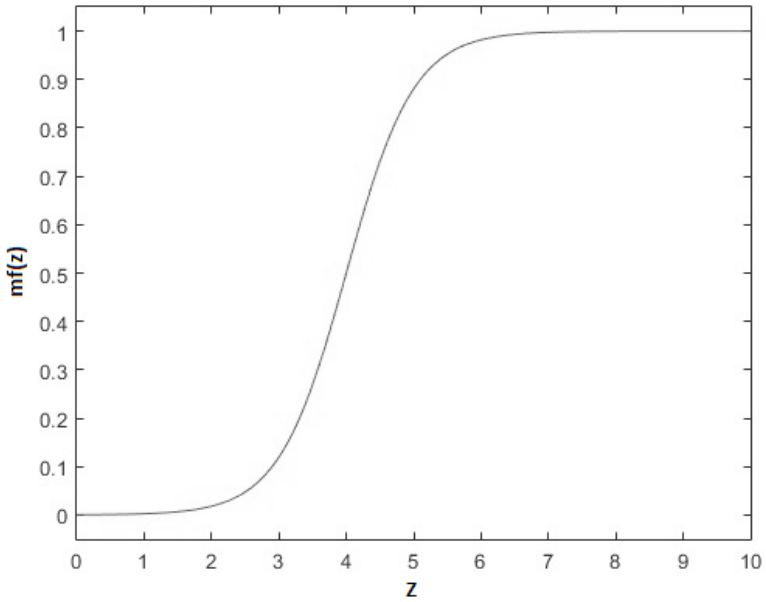

Figure 1. A typical sigmoidal membership function.

The parameter *a* here controls the slope at the cross-over point $z_i = b$. The sigmoidal MF is either open to its right or left subject to the sign of the parameter *a*; and thus can represent 'very positive' or 'very negative'. It is also easier to model and train in an ANN model. So, a neuro-fuzzy model with sigmoidal MF is proposed and used in this work. A typical sigmoidal MF is shown below in Figure 1.

If we modify the values of the two parameters a and b, it is easier to find the desired MF that offers more flexibility for the classification task. We apply the above MF to the input pattern vector z for developing a membership matrix. The resulting matrix appears like this:

$$M(z)=\begin{bmatrix} m_{1,1}(z_1) & m_{1,2}(z_1) & m_{1,3}(z_1) & \cdots & m_{1,P}(z_1) \\ m_{2,1}(z_2) & m_{2,2}(z_2) & m_{2,3}(z_2) & \cdots & m_{2,P}(z_2) \\ m_{3,1}(z_3) & m_{3,2}(z_1) & m_{3,3}(z_3) & \cdots & m_{3,P}(z_3) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ m_{Q,1}(z_Q) & m_{Q,2}(z_Q) & m_{Q,3}(z_Q) & \cdots & m_{Q,P}(z_Q) \end{bmatrix} \quad (4)$$

Here $m_{i,j}(z_i)$ is the degree of membership of the *i-th* pattern in input vector *z* to the class *j* using *i =1, 2,…, Q* and *j =1, 2,…, P*. For example, $m_{3,6}(z_3)$ denotes the degree of membership of 3rd pattern to the class 6. The membership matrix is used as input to an ANN model as presented below.

2.2 Phase 2

The second phase builds an MLP classification model or classifier. In this phase, the above matrix is transformed into a $(P \times Q)$ vector through carrying out transpose operation. This vector is then fed to an MLP classifier. The classifier consists of a hidden layer in between the input layer and the output layer. The node count in input layer equals the number of elements in the matrix as stated above and the node count in output layer equals the number of obtainable data classes. The given ANN model uses gradient descent as the learning rule. The output layer and the hidden layers of the model use tan sigmoid as the activation function. The selection of the number of Processing Elements (PEs) present in the hidden layers is also an essential parameter. Subsequently a thorough investigation helps us in selecting the number of PEs present in the hidden layers[32]. The hidden layer consists of some PEs, denoted by, *h* is given by the following equation as:

$$h = \sqrt{(number\ of\ inputs * number\ of\ outputs)} \quad (5)$$

2.3 Phase 3

The third or final phase makes use of the defuzzification procedure that is conceptually just opposite to that of the first phase. In this phase, the proposed classi-

fier performs a rigid (i.e., hard) classification by using a maximum (MAX) operation to produce the activation output of the given MLP model. An input pattern will be associated with a particular class *j* provided that the pattern has the highest class membership value with respect to class *j* compared to the other classes. Therefore, an unknown pattern *x* is assigned to *class j* based on the concept of "highest class membership value" if and only if:

$$M_j(x) \geq M_i(x) \quad \forall i \in (1, 2, ..., P) \, and \, i \neq j \tag{6}$$

Here $M_i(x)$ is the activation output of the *i-th* node in the final or output layer of the MLP model.

The block diagram of our proposed Neuro-Fuzzy System model is shown below in Figure 2.

3. Methodology

The methodology relies on the NFS, MLP and SVM models or classifiers. These classifiers examine the three benchmark UCI breast cancer datasets, namely Wisconsin Breast Cancer (WBC), Wisconsin Diagnostic Breast Cancer (WDBC) and Mammographic Mass (MM) for classification. The broad level stages of the proposed methodology are described here in detail.

3.1 Stage 1

Initially, the following preprocessing methods are employed to each of the breast cancer dataset before classification task:

3.1.1 Stage-1a. Data Cleaning

It represents the preprocessing of data for diminishing noise and managing the missing values. The arithmetic mean usually substitutes a missing value for an attribute based on statistics.

3.1.2 Stage-1b. Data Transformation

The method attempts to normalize the dataset as because the neural network model requires measurement of distance for classification analysis. It transforms the values of the database attribute to a small-scale interval like -1.0 to +1.0.

3.2 Stage 2

Afterward, every single data set is distributed into two separate subsets, namely the training data set and the test data set. We employ the well-known 10-fold Cross-Validation (CV) method to generate the training and test

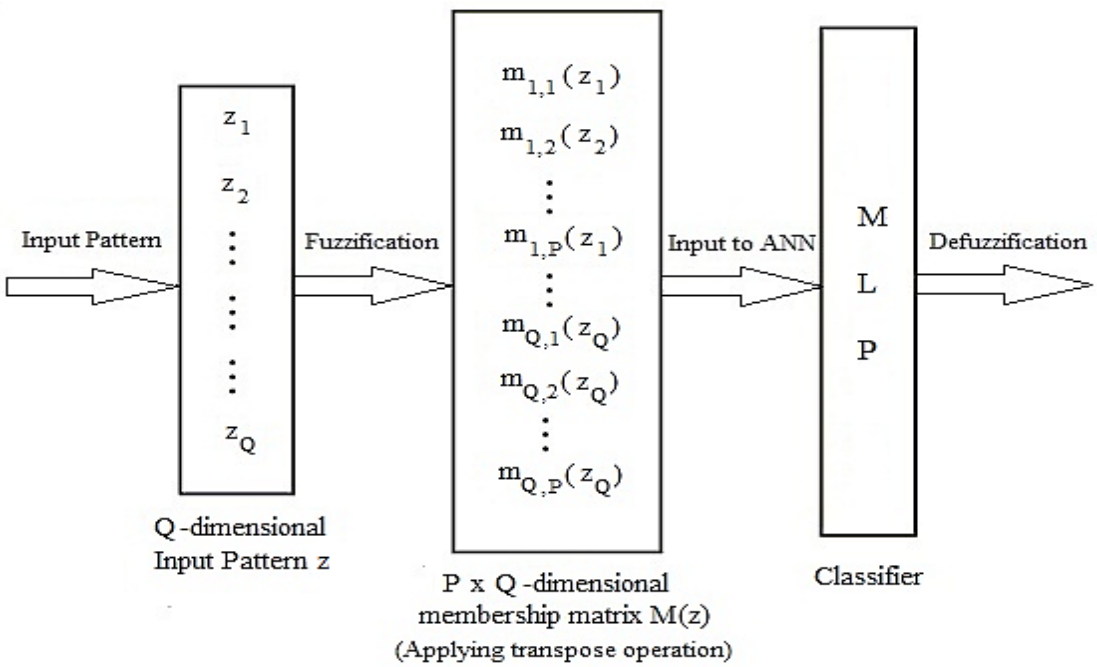

Figure 2. Proposed Neuro-Fuzzy System model.

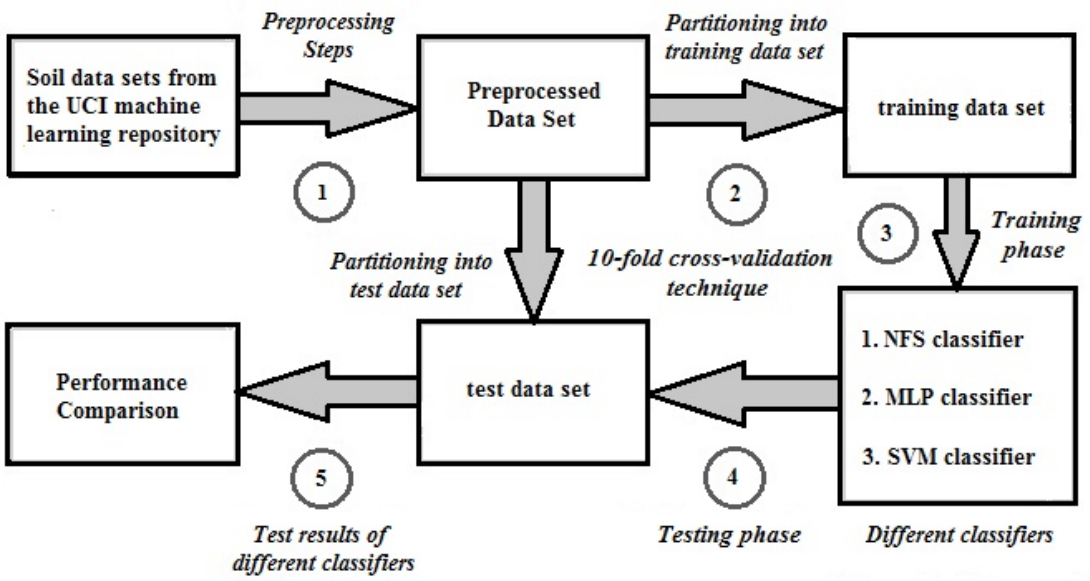

Figure 3. Broad level stages of the proposed methodology.

datasets so that they should be independent of each other for avoiding biases.

3.3 Stage 3

The proposed Neuro-Fuzzy System (NFS) utilizes the training set for building a classification model. The training set is also given to the MLP and SVM techniques individually for developing other models.

3.4 Stage 4

The three classification models (NFS, MLP and SVM) later applied to the test dataset for evaluating the performance of each classifier.

3.5 Stage 5

Some well-known metrics, for example, RMSE, kappa statistic, accuracy, FP-Rate, TP-Rate/recall, precision and f-measure are used to perform quantitative analysis of the results generated by these classifiers.

The broad level stages of the proposed methodology are depicted below in Figure 3.

4. Results and Discussion

The three classification techniques namely NFS, MLP and SVM are trained and tested on three benchmark breast cancer datasets using MATLAB software (version R2015a).

The MLP classification method uses the same set of configuration parameters that we employ in the MLP network structure of the proposed NFS model. Furthermore, the optimal configuration for developing an SVM classifier is described here. We apply different types of kernel-based techniques in the simulation. Finally, an SVM model with a polynomial kernel is selected as the best option among all. A nonlinear version of SVM can be represented by using a suitable kernel function K as:

$$K(x_i, x_j) = \phi(x_i).\phi(x_j) \tag{7}$$

Here (x) signifies a non-linear mapping function employed to map the records in the database. An SVM model using a polynomial kernel of degree (exponent) d can be defined as:

$$K(x_i, x_j) = (x_i . x_j + 1)^d \tag{8}$$

4.1 Performance Measures

After building the models by the classification techniques mentioned above, they are applied to the test dataset for performance evaluation. We have estimated the performances of these classification models based on several evaluation metrics presented below.

4.1.1 Root-Mean-Square Error (RMSE)

The RMSE[33] evaluation metric denotes a long-familiar performance measure of the dissimilarity between the

values predicted by a classifier and the values set up on the model being demonstrated. The RMSE metric of a model's assessment concerning the calculated variable $e_{classifier}$ implies the square root of the mean-squared inaccuracy:

$$rmse = \frac{\sqrt{\sum_{k=1}^{n}(e_{discovered,k} - e_{classifier,k})^2}}{n} \quad (9)$$

Here $e_{discovered}$ are the discovered values and $e_{classifier}$ are the predicted values for $\forall k$.

4.1.2 Kappa Statistic

The kappa statistic[34], represented by κ, is a well-known performance metric in statistics. It denotes the relative measure of reliability within different raters or judges. The following equation estimates the value of κ as:

$$K = \frac{prob(O) - prob(C)}{1 - prob(C)} \quad (10)$$

Here prob *(O)* is the probability of witnessed settlements amongst the judges and prob (C) is the probability of settlements estimated by coincidence. If $\kappa = 1$, the judges have approved each other's decision. If $\kappa = 0$, then the judges do not agree with each other.

4.1.3 Confusion Matrix

In the soft computing field, the confusion matrix[35] is a specific table layout illustrating a classification algorithm's performance. It is a tabular representation that permits more thorough analysis than accuracy. Each attribute of the matrix denotes the patterns in the anticipated data class whereas every tuple designates the patterns in the definite class. Table 1 below displays a confusion matrix layout using a two-class classifier is having the following cells as:

- True-Positive (TP) indicates the amount of 'positive' patterns categorized as 'positive.'
- False-Positive (FP) means the amount of 'negative' patterns categorized as 'positive.'
- False-Negative (FN) denotes the amount of 'positive' patterns categorized as 'negative.'
- True-Negative (TN) implies the amount of 'negative' patterns categorized as 'negative.'

A two-class confusion matrix defines several standard terms. The accuracy (i.e. classification accuracy) is the

Table 1. A confusion matrix layout for a two-class classifier

		Predicted Class	
		positive	negative
Actual Class	positive	tp	fp
	negative	fn	tn

sum of the correctly classified examples divided by the total number of examples present. The following equation calculates this as:

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (11)$$

The precision denotes the ratio of the predicted positive examples found to be correct, as calculated using the following equation:

$$precision = \frac{tp}{tp + fp} \quad (12)$$

The FP-Rate (i.e. False-Positive Rate) indicates the ratio of negative examples incorrectly classified as positive, as given by the equation:

$$fp - rate = \frac{fp}{fp + tn} \quad (13)$$

The TP-Rate (i.e., True-Positive Rate) or recall is the ratio of positive occurrences discovered correctly, as estimated using the equation:

$$recall = tp - rate = \frac{tp}{tp + fn} \quad (14)$$

In some situations, a large precision value may be more relevant while sometimes large recall value might be more prominent. Nevertheless, in the most demonstrations, one should attempt to increase both values. The integrated arrangement of these values is the f-measure and typically articulated as a harmonic mean of both the values:

$$f - measure = \frac{2 * precision * recall}{precision + recall} \quad (15)$$

4.2 Results and Performance Analysis

In the testing phase, NFS, MLP and SVM classifiers are applied to three UCI benchmark datasets for investigation and performance analysis described below. The

results that we report here solely based on the simulation experiment that we have taken.

4.2.1 Wisconsin Breast Cancer (WBC) Database

We have used the benchmark Wisconsin Breast Cancer (WBC)[36] database of UCI for analysis of breast cancer. The dataset has 699 rows and 11 columns. The class label column has two values, specifically Benign (i.e. normal) and Malignant (i.e. cancerous). The statistical distribution of data tuples per class with percentage values are - Benign: 458 (65.5%) and Malignant: 241 (34.5%). There are sixteen samples in the dataset consisting of a single missing attribute value, marked by the symbol '?'. We resolve these missing values during the pre processing phase. We provide here descriptions of the attributes present in the database.

The first attribute denotes a sample code number and appropriately not considered for classification. The attributes having serial numbers starting from 2 to 10 are important in the classification procedure. These are the input attributes in the database. The clump thickness property indicates that the cancerous cells are grouped in multiple layers, while normal (Benign) cells tend to combine in a single layer.

Uniformity of cell size/shape denotes the similarity in size and shape of normal cells while the cancer cells tend to change in size and shape. Marginal adhesion in breast tissue cells is an important characteristic to ascertain cancer. The loss of adhesion is a symptom of malignancy as the cancer cells lose the ability to stay together. The feature single epithelial cell size denotes the homogeneity of cell size. Epithelial cells that are considerably large might be a malignant cell.

The term 'bare nuclei' indicate a type of nuclei that is not bounded by the cell cytoplasm. We usually witness them in the benign tumors. The bland chromatin describes an identical texture of the nucleus commonly seen in the benign cells. In cancer cells, the chromatin becomes harsher.

The normal nucleoli are microscopic structures observed in the cells and are very negligible in size. Usually, nucleoli in the cancerous cells are more noticeable and at times there are more than one. Mitoses are nuclear division combined with cytokines that create two identical cells during the prophase stage. Diagnosticians count mitoses to decide the degree of cancer. The last or the eleventh attribute is the class attribute and it has got two values, specifically Benign and Malignant. These two values are interpreted by the numeric values 2 and 4 respectively.

Table 2. Comparison of accuracy, RMSE and kappa statistic for WBC database

Classifier/Model	Accuracy	RMSE	Kappa statistic
NFS	97.8 %	0.1062	0.9784
MLP	86.3 %	0.2132	0.8614
SVM	87.6 %	0.1922	0.8762

Each of the three classifiers namely NFS, MLP and SVM is applied individually for classifying the test dataset. We have used measures suchlike classification accuracy, RMSE and kappa statistic for performance evaluation as shown below in Table 2.

From Table 2, we can recognize that the accuracy of the NFS classifier is 97.8%. The MLP model has a classification accuracy of 86.3%; while the SVM model is having the accuracy of 87.6%. Certainly, NFS has better classification accuracy than MLP and SVM. Then we investigate the performance comparison of these classifiers established on the RMSE and the Kappa statistic index.

The experiment has used the well-known measure like the RMSE value that is likely to be very low. Kappa statistic is an excellent estimator for inter-rater agreement between classes. As it is evident from Table 2, kappa statistic indexes of the designated algorithms are nearby 0.81-1.0. Following the definition of the kappa statistic, the result is considered to be 'almost perfect agreement'. The result indicates that NFS gives the best performance with the RMSE value of 0.1062 and a kappa statistic index of 0.9784, followed by SVM having 0.1922 as RMSE value and 0.8762 as a Kappa statistic index. Whereas, MLP holds the last position is having the largest RMSE value (0.2132) and the smallest kappa statistic index (0.8614). The Figure 4 below shows this statistical evidence using a 3-D column diagram.

Then we have conducted the performance evaluation of these models in Table 3 below using the values of TP-Rate/recall, FP-Rate, precision and f-measure built from the confusion matrix. For assessing the classifier performance, we should presume greater magnitudes for TP-Rate/recall, precision and f-measure; and smaller quantity for FP-Rate.

In Table 3, the magnitudes of TP-Rate/recall, FP-Rate, precision and f-measure metrics for the NFS model are 97.7%, 3.2%, 97.6% and 97.6% individually. On the other hand, for the MLP classifier these values are 86.2%,

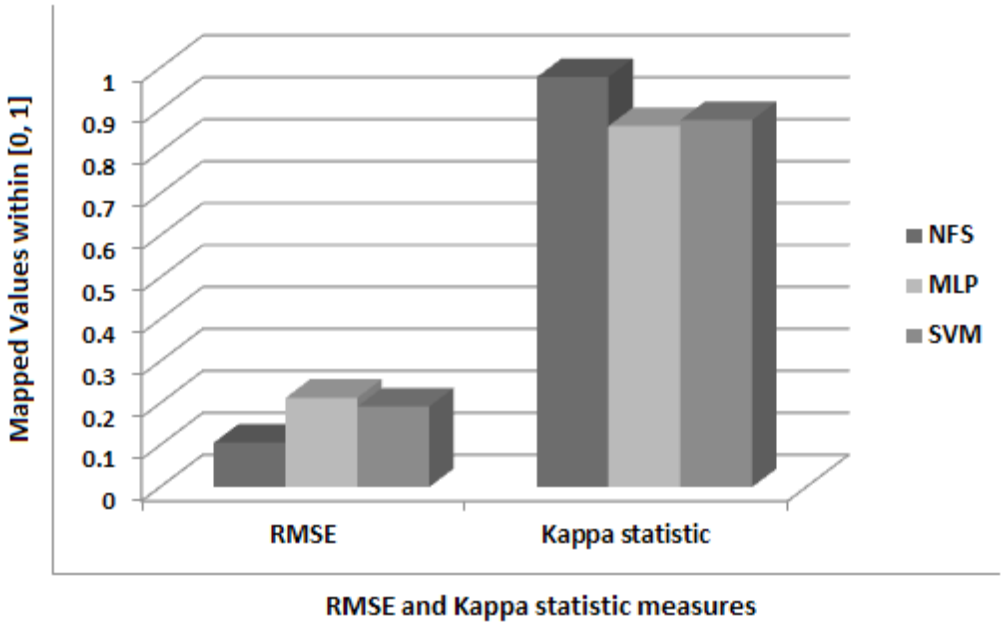

Figure 4. Comparison of RMSE and kappa statistic using WBC database.

Table 3. Comparison of TP-Rate/recall, FP-Rate, precision and f-measure for WBC database

Classifier /Model	TP-Rate/ Recall	FP-Rate	Precision	F-Measure
NFS	97.7 %	3.2 %	97.6 %	97.6 %
MLP	86.2 %	12.8 %	86.2 %	86.2 %
SVM	87.5 %	11.7 %	87.4 %	87.4 %

12.8%, 86.2% and 86.2% respectively. The SVM model has the values of TP-Rate/recall, FP-Rate, precision and f-measure as 87.5%, 11.7%, 87.4% and 87.4% separately. Certainly, the NFS model has the best values for TP-Rate/recall, precision and f-measure and the lowest value for FP-Rate metric compared to others. We currently validate this information using a 3-D column diagram as given below in Figure 5.

Hence, based on all these performance measures, the NFS model has specified the finest performance.

4.2.2 Wisconsin Diagnostic Breast Cancer Database

Subsequently, we have used the benchmark UCI database named Wisconsin Diagnostics Breast Cancer (WDBC)[37] for classification performance analysis. The dataset consists of 569 instances and 32 attributes. These attributes include the ID number, diagnosis (class attribute) and 30 real-valued input attributes or input features. The class column has two values, specifically Benign (B) and Malignant (M). The numbers of instances per class with percentage values are - Benign: 357 (62.7%) and Malignant: 212 (37.3%). The attributes from 3 to 32 of this dataset are calculated from the digitized imagery of the Fine Needle Aspirate (FNA) of breast tissue mass. They identify the features of the cell nuclei present in this digitized image. The significances of each of the attributes in the WDBC dataset are described below in details.

- ID number.
- Diagnosis (M = Malignant, B = Benign).
- Exactly 10 real-valued characteristics are calculated for each of the cell nucleus as:
  - The radius (the mean value of distances from the center to points on the perimeter).
  - The texture (the standard deviation of gray-scale values).
  - The perimeter.
  - The area.
  - The smoothness (the local variation in the radius lengths).

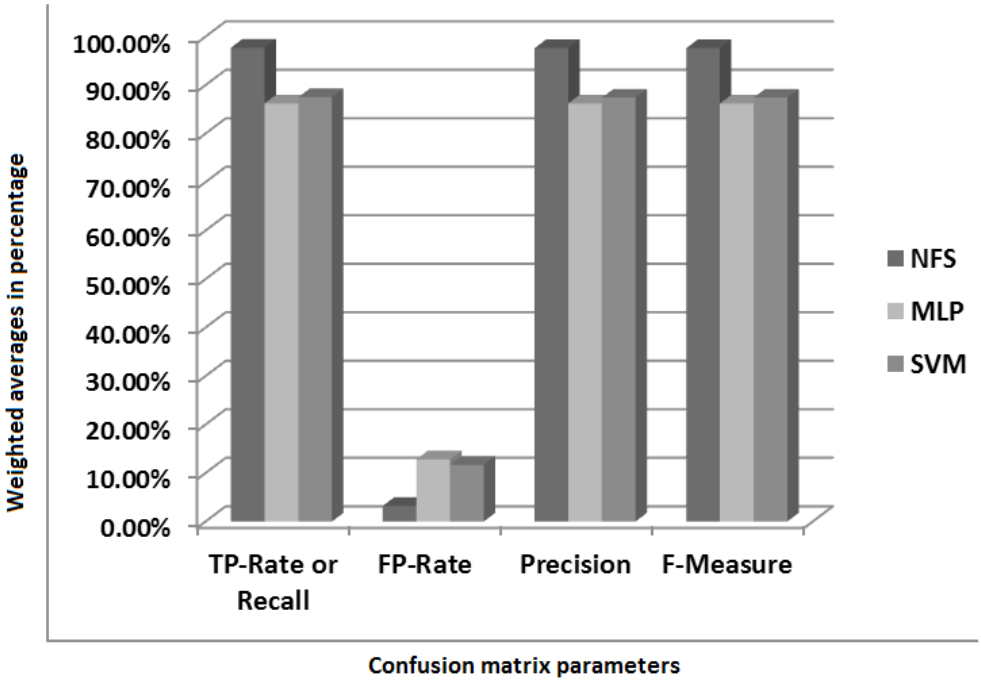

Figure 5. Comparison of TP-Rate/recall, FP-Rate, precision and f-measure using WBC database.

- The compactness ([perimeter]2/area - 1.0).
- The concavity (the severity of concave portions of the contour).
- The concave points (the number of concave portions of the contour).
- The symmetry.
- The fractal dimension (coastline approximation - 1).

Primarily mean value, Standard Error (SE) value and the largest value (mean of the three largest values) of these features are computed for each digitized image, resulting in exactly 30 features. For example, field 3 is the mean radius value, field 13 is the radius SE value and field 23 is the largest radius value. All the feature values here are coded with four significant digits.

The first attribute ID number is the id number of each tuple in the database and considered to be irrelevant to the classification process; hence the attribute is not considered for classification. The second attribute diagnosis denotes the class attribute and takes in exactly two values, namely, Malignant and Benign denoted by 'M' and 'B' respectively.

So the three classification models, namely NFS, MLP and SVM are applied to the test dataset for classification. We have evaluated the performance of these classifiers along the base of different performance measures such-like accuracy, RMSE and the kappa statistic index as stated below in Table 4.

From Table 4, we can understand that the classification accuracy of the NFS classifier is 95.9%. MLP model has classification accuracy of 84.4%; while SVM model has an accuracy of 82.7%. Surely, NFS has much better accuracy than MLP and SVM. Then we have made a performance comparison analysis among these classifiers using evaluation metrics like RMSE and kappa statistic. The kappa statistic value of the selected classification algorithms is about 0.81-1.0. By the definition of the kappa statistic, the result is 'almost perfect agreement'. The results show that the NFS model gives the best performance with a RMSE value of 0.1254 and a kappa statistic index of 0.9504. Followed by MLP is having a RMSE magnitude of 0.2265 and a kappa statistic index of 0.8552 and SVM holds the last place with

Table 4. Comparison based on accuracy, RMSE and kappa statistic for WDBC database

Classifier/Model	Accuracy	RMSE	Kappa statistic
NFS	**95.9 %**	**0.1254**	**0.9504**
MLP	84.4 %	0.2265	0.8552
SVM	82.7 %	0.2624	0.8302

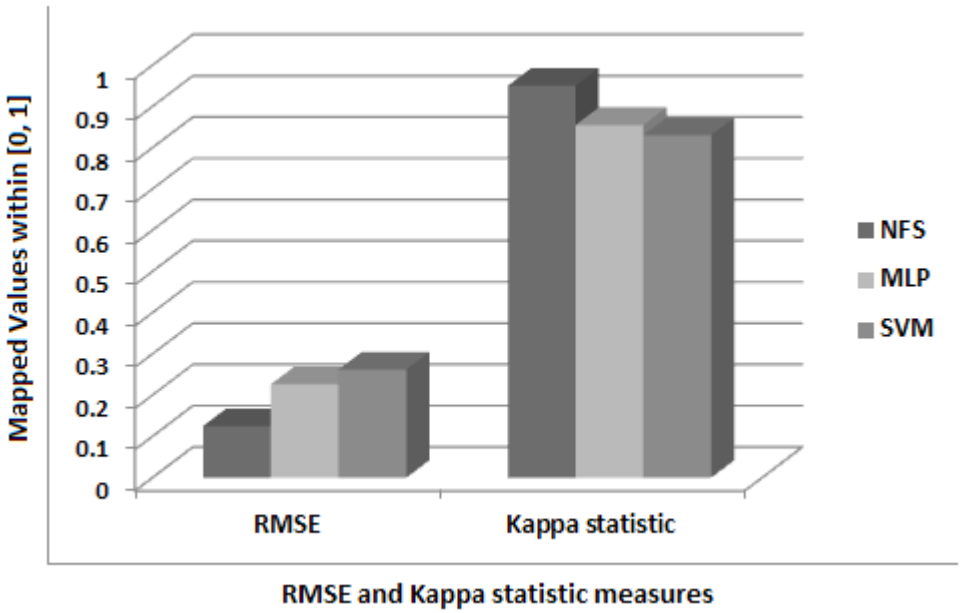

Figure 6. Comparison of RMSE and kappa statistic using WDBC database.

the highest RMSE (0.2624) and the lowest kappa statistic (0.8302). The Figure 6 below displays the detailed statistical information using a 3-D column diagram.

Next, we have performed a performance comparison of these classification models using the numeric values of TP-Rate/recall, FP-Rate, precision and f-measure constructed from the confusion matrix in Table 5 below.

In Table 5, the magnitudes of TP-Rate/recall, FP-Rate, precision and f-measure metrics for the NFS classification model are 95.7%, 4.6%, 95.8% and 95.7% individually. However, MLP classifier is having these values as 84.2%, 13.2%, 84.2% and 84.2% respectively. The SVM model has the values of TP-Rate /recall, FP-Rate, precision and f-measure as 82.6%, 14.5%, 82.5% and 82.5% separately. Indeed, the NFS model has the highest values for TP-Rate/recall, precision and f-measure and the lowest value for the FP-Rate. We now ascertain this statistical evidence using a 3-D column diagram as indicated below in Figure 7.

Table 5. Comparison of TP-Rate /recall, FP-Rate, precision and f-measure for WDBC database

Classifier /Model	TP-Rate/ Recall	FP-Rate	Precision	F-Measure
NFS	**95.7 %**	**4.6 %**	**95.8 %**	**95.7 %**
MLP	84.2 %	13.2 %	84.2 %	84.2 %
SVM	82.6 %	14.5 %	82.5 %	82.5 %

Therefore, established on all these evaluation criteria, the NFS model has given the best performance.

4.2.3 Mammographic Mass Database

Lastly, we have used the Mammographic Mass (MM)[38] database of UCI for our research study. Numerous tools involving Computer Aided Diagnosis (CAD) are utilized to decrease the high number of needless breast biopsies in the last decades. These systems are a collaborative endeavor of many health groups. CAD based systems should assist physicians to decide whether to do a biopsy on any localized abnormal structural change in breast tissues using mammogram test as an alternative. The researchers typically employ a tool called Breast Imaging-Reporting and Data System (BI-RADS) for quality assurance. The dataset contains 961 tuples and has 6 attributes. The first two attributes in this dataset are the assessment of BI-RADS tool and the age of the patient. The remaining attributes are the different BI-RADS features associated with a mammography mass, such as Shape, Margin and Density and finally the class attribute named Severity. The class label column contains two values, specifically Benign and Malignant. The statistical distribution of data tuples per class with percentage values are - Benign: 516 (53.7%) and Malignant: 445 (46.3%). There are some missing attribute values in this dataset that are resolved during the

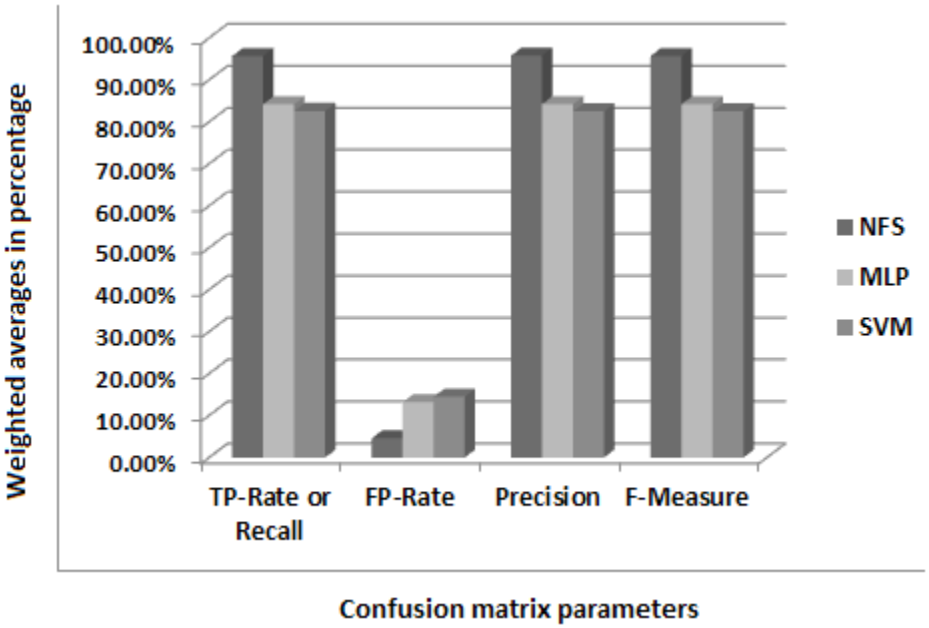

Figure 7. Comparison of TP-Rate/recall, FP-Rate, precision and f-measure using WDBC database.

preprocessing stage. The implications of the attributes in the dataset are designated below.

The first attribute denotes the BI-RADS assessment that lies within 1 to 5. A higher value is likely to increase the chance of malignity. The second attribute named Age represents the patient's age in years. The third column named Shape designates the shape of a mammography mass and its value lies within 1 to 4. The fourth column named Margin denotes mass margin, and its value ranges from 1 to 5. The fifth column named Density identifies the density of mammography mass and its value lies within 1 to 4. The sixth attribute Severity denotes the class attribute and brings in exactly two values, namely, Benign and Malignant indicated by numeric values '0' and '1' respectively. All these features are associated with digital mammography assessment performed by CAD-based tools like BI-RADS.

As usual the three classifiers, namely NFS, MLP and SVM are applied to the test dataset for classification. We have measured the operation of these classifiers along the theme of different performance evaluation metrics such-like classification accuracy, RMSE and the kappa statistic index as presented below in Table 6.

From Table 6, we can realize that the classification accuracy of the NFS classifier is 88.7%. The MLP model has the accuracy of 77.9%; while the SVM model has an accuracy of 78.6%. Evidently, NFS has better classification accuracy than MLP and SVM. Then we inspect the performance comparison of these classifiers established on the RMSE and the kappa statistic index as collected from Table 6. We could see that the kappa statistic index of the selected algorithms is ranging from 0.61 to 0.80. Following the definition of the kappa statistic, this is considered to be 'substantial'. Based on the above result, NFS again gives the best performance with 0.2759 as the RMSE value and 0.7938 as the kappa statistic index. Followed by SVM are having 0.3623 as its RMSE value and 0.6354 as the kappa statistic index. MLP has held the last position with the maximum RMSE value (0.3753) and the lowermost kappa statistic index (0.6025). The Figure 8 below demonstrates this information using a 3-D column diagram.

Table 6. Comparison based on accuracy, RMSE and kappa statistic for MM database

Classifier/Model	Accuracy	RMSE	Kappa statistic
NFS	**88.7 %**	**0.2759**	**0.7938**
MLP	77.9 %	0.3753	0.6025
SVM	78.6 %	0.3623	0.6354

Afterward, we have conducted the performance comparison of these classifiers in Table 7 below based on the

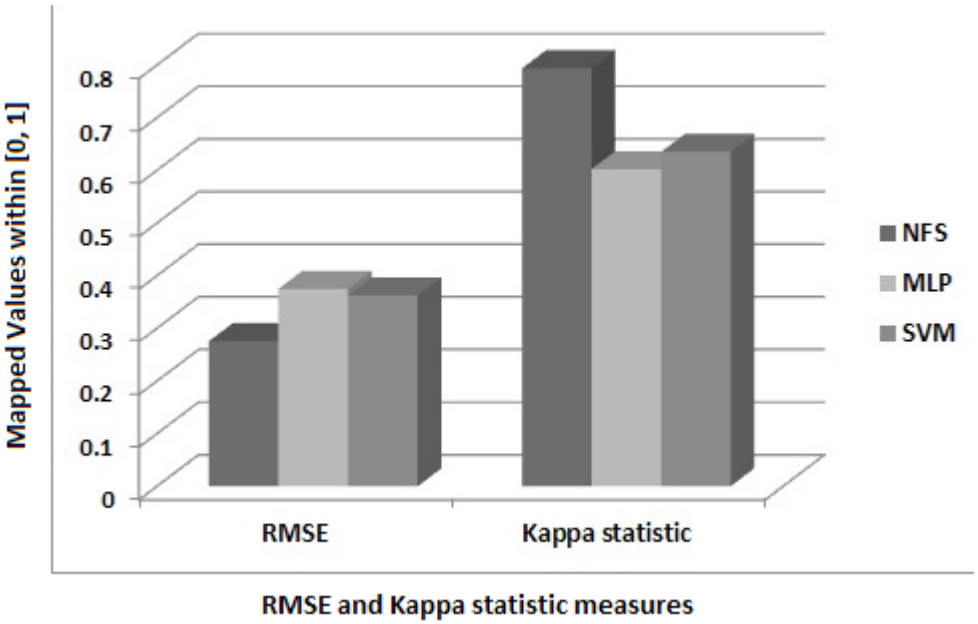

Figure 8. Comparison of RMSE and kappa statistic using MM database.

information about TP-Rate/recall, FP-Rate, precision and f-measure metrics created from the confusion matrix.

In Table 7, the magnitudes of TP-Rate/recall, FP-Rate, precision and f-measure metrics for the NFS model are 88.7%, 10.8%, 87.6% and 87.6% correspondingly. Whereas, the MLP classifier consists of these values as 77.8%, 20.1%, 77.8% and 77.8% respectively. The SVM model has the values of TP-Rate/recall, FP-Rate, precision and f-measure as 78.6%, 19.3%, 78.5% and 78.5% individually. Certainly, the NFS model has the highest values TP-Rate/recall, FP-Rate, precision and f-measure and the lowest value for FP-Rate among all. We now validate this statistical information in the arrangement of a 3-D column diagram as shown below in Figure 9.

Thus, based on all these assessment measures, the NFS model has established the highest performance.

Table 7. Comparison of TP-Rate/recall, FP-Rate, precision and f-measure for MM database

Classifier / Model	TP-Rate/ Recall	FP-Rate	Precision	F-Measure
NFS	**88.7 %**	**10.8 %**	**87.6 %**	**87.6 %**
MLP	77.8 %	20.1 %	77.8 %	77.8 %
SVM	78.6 %	19.3 %	78.5 %	78.5 %

As a whole, in respect of all the performance evaluation metrics used for the three UCI datasets, we have obtained superior results for the proposed NFS classification method compared to MLP and SVM-based models. The NFS technique has the largest values for accuracy, kappa statistic, TP-Rate/recall, precision and f-measure and the smallest values for RMSE and FP-Rate metrics during the simulation experiment. Indeed, the NFS method showing an excellent predictive ability and reduced error rate has outperformed MLP and SVM models in all respects. Therefore, our proposed neuro-fuzzy based method has been very much efficient in predicting the presence (i.e. Malignant) or absence (i.e. Benign) of the breast cancer disease.

5. Conclusion

The present study proposes a neuro-fuzzy based classification method for breast cancer detection and established its efficiency successfully using three UCI datasets, namely, WBC, WDBC and MM. The method utilizes and integrates the primary benefits of Artificial Neural Networks such as immense parallelism, adaptivity, robustness and optimality with the vagueness and imprecision management capability of fuzzy sets. Furthermore, the proposed classification model builds a

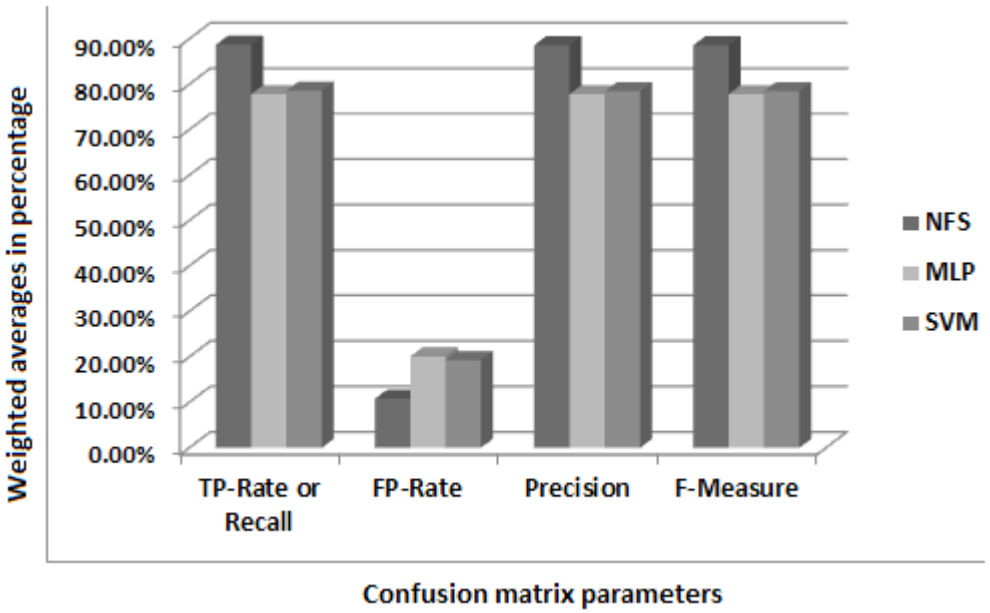

Figure 9. Comparison of TP-Rate/recall, FP-Rate, precision and f-measure using MM database.

membership matrix that offers information of the pattern-wise degree of membership of a data pattern to all the classes present.

The research work intends to analyze and investigate the operational aspects of the proposed NFS technique in comparison with MLP and SVM classifiers. We have used different metrics such as accuracy, RMSE, kappa statistic index, TP-Rate/recall, FP-Rate, precision and f-measure to accomplish performance evaluation. The proposed neuro-fuzzy classification method has an accuracy of 97.8% using the WBC dataset, 95.9% using the WDBC dataset and 88.7% using the MM dataset. These results are significantly better than that of MLP and SVM algorithms (accuracy is more than 10%). Considering the three benchmark UCI datasets, the NFS classifier also has the lowest RMSE value and the highest f-measure and kappa statistic values compared to MLP and SVM. Thus, it is concluded that the proposed technique has great potential in predicting breast cancer disease and thus can be used in the Medical Science and the Bioinformatics research field.

7. References

1. Agrawal R, Imielinski T, Swami A. Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering. 1993 Dec; 5(6):914–25.
2. Chen MS, Han J, Yu PS. Data mining: An overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering. 1996 Dec; 8(6):866–83.
3. Coyne E, Borbasi S. Living the experience of breast cancer treatment: The younger women's perspective. Australian Journal of Advanced Nursing. 2009 Jun; 26(4):6–13.
4. Boyle P, Levin B. World Cancer Report 2008. International Agency for Research on Cancer (IARC). World Health Organization; 2008. p. 1–260.
5. Han J, Kamber M. Data mining concepts and techniques. 2nd ed. 2006.
6. Rojas R. The backpropagation algorithm. Neural networks. A systematic introduction. Berlin: Springer-Verlag; 1996.
7. Haykin S. Neural networks: A comprehensive foundation. Prentice Hall; 2nd ed. 1998.
8. Bose NK, Liang P. Neural network fundamentals with graphs, algorithms and applications. McGraw-Hill; 1996.
9. Werbos PJ. The roots of back propagation: From ordered derivatives to neural networks and political forecasting. New York: John Wiley and Sons; 1994.
10. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986 Oct; 323:533–6.
11. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995 Sep; 20(3):273–97.
12. Zadeh LA. Fuzzy sets. Information and Control. 1965; 8(3):338–53.
13. Liu B. Uncertain theory: An introduction to its axiomatic foundations. Berlin: Springer-Verlag; 2004.

14. Dubois D, Prade H. Fuzzy sets and systems. New York: Academic Press; 1980.
15. Jang JSR, Sun CT, Mizutani E. Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence. Prentice Hall; 1997.
16. Lin CT, Lee CSG. Neural Fuzzy Systems: A neuro-fuzzy synergism to intelligent systems. Prentice Hall; 1996.
17. Clancey WJ, Shortliffe EH, Eds. Readings in medical artificial intelligence: The first decade. Reading, Mass: Addison-Wesley; 1984.
18. Anunciacao O, Gomes BC, Vinga S, Gaspar J, Oliveira AL, Rueff J. A data mining approach for the detection of high-risk breast cancer groups. Advances in Intelligent and Soft Computing. Springer-Verlag. Berlin Heidelberg. 2010; 74:43–51.
19. Cheng WY, Yang THO, Anastassiou D. Development of a prognostic model for breast cancer survival in an open challenge environment. Sci Transl Med. 2013 Apr; 5(181):1–12.
20. Janghel RR, Shukla A, Tiwari R, Kala R. Breast cancer diagnosis using Artificial Neural Network model. Proceedings of the 3rd IEEE International Conference on Information Sciences and Interaction Sciences; Chengdu, China. 2010 Jun 23-25. p. 89–94.
21. Keleş A, Keleş A. Extracting fuzzy rules for diagnosis of breast cancer. The Turkish Journal of Electrical Engineering and Computer Science. 2013 Aug; 21(1):1495–503.
22. Nauck D, Klawonn P, Kruse P. Foundations of Neuro Fuzzy Systems. Wiley: Chichester; 1997.
23. Venkatesan E, Velmurugan T. Performance analysis of decision tree algorithms for breast cancer classification. Indian Journal of Science and Technology. 2015 Nov; 8(29):1–8.
24. Vaidehi K, Subashini TS. Breast tissue characterization using combined K-NN classifier. Indian Journal of Science and Technology. 2015 Jan; 8(1):23–6.
25. Aarthy SL, Prabu S. An approach for detecting breast cancer using wavelet transforms. Indian Journal of Science and Technology. 2015 Oct; 8(26):1–7.
26. Venkatesan E, Velmurugan T. Performance analysis of decision tree algorithms for breast cancer classification. Indian Journal of Science and Technology. 2015 Nov; 8(29): 1-8.
27. Moh'd Rasoul A. Al-Hadidi, Mohammed Y. Al-Gawagzeh, Bayan A. Alsaaidah. Solving mammography problems of breast cancer detection using Artificial Neural Networks and image processing techniques. Indian Journal of Science and Technology. 2012 Apr, 5(4):2520–8.
28. Garge DM, Bapat VN. A low cost wavelet based mammogram image processing for early detection of breast cancer. Indian Journal of Science and Technology. 2009 Sep; 2(9):63–5.
29. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Computational Biology. 2011 Oct; 7(10):1–8.
30. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Elsevier. 2011 Mar; 144(5):646–74.
31. Ghosh S, Biswas S, Sarkar D, Sarkar PP. A novel neuro-fuzzy classification technique for data mining. Egyptian Informatics Journal. Elsevier. 2014 Nov; 15(3):129–47.
32. Elisseeff V, Paugam-Moisy H. Size of multilayer networks for exact learning: Analytic approach. Advances in Neural Information Processing Systems, Cambridge, MA: The MIT Press. 1997; 9:162–8.
33. Armstrong JS, Collopy F. Error measures for generalizing about forecasting methods: Empirical Comparisons. International Journal of Forecasting. 1992 Jun; 8(1):69–80.
34. Carletta J. Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics. 1996 Jun; 22(2):249–54.
35. Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment. 1997 Oct; 62(1):77–89.
36. Breast Cancer Wisconsin (Original) UCI Dataset, 1992 Jul. Available from: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29/
37. Breast Cancer Wisconsin (Diagnostic) UCI Dataset, 1995 Nov. Available from: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29/
38. Mammographic Mass UCI Dataset. 2007 Oct. Available from: https://archive.ics.uci.edu/ml/datasets/Mammographic+Mass/