Intelligent Prediction of Breast Cancer: A Comparative Study

Merhan A. Abd-Elrazek [1], Ahmed A. Othman [2], Mohamed H. Abd Elaziz[3]and Mohamed N. Abd-Elwhab[4]

[1]Computer and Control Dept., Faculty of Engineering, Suez Canal University, Egypt

[2]Information System Dept., Faculty of Computers & Informatics, Suez Canal University, Egypt

[3]Information Systems Dept., Faculty of Computers & Information, Ain shames University, Egypt

[4]Electric Engineering Dept., Faculty of Engineering, Suez Canal University, Egypt

eng.merhanahmed@eng.suez.edu.eg, a.othman@ci.suez.edu.eg, mhaziz@cis.asu.edu.eg, mohamed.nabil@eng.suez.e

Abstract

Breast cancer is defined as the growth of breast tissue in abnormal way that performing tumors. Breast cancer is one of the most spreading cancer between women and may considered as the first cause of their death. Not all tumors in breast are classified as breast cancer. However, they must be examined by physicians even if it may be normal tumors. Therefore, Detecting breast cancer in an early stage could increases the percent of surviving and may be saves more lives. In this paper, we propose a complete comparative study between different classifications techniques used to predict breast cancer. A set of popular supervised machine learning and data mining techniques will be used to predict breast cancer (benign or malignant). Three different techniques are proposed namely: Classification without feature selection (CWFS), Feature Selection Classification (FSC) and Normalization and Feature Selection Classification (NFSC). The results of accuracy, specificity, precision and sensitivity are calculated and recorded for each system. Hence, our results when compared with the up to date techniques show higher accuracy.

Keywords: Health care, Breast Cancer, Machine Learning, Classification Algorithm, Medical Records.

1. Introduction

Patients` medical records have a large amount of data such as doctor and nurse notes, imaging studies, laboratory results, medications and progression notes. Moreover, the data are collected from different sources and data types and cannot be managed and processed easily without computer aided systems. Therefore, machine learning and data mining have been widely used in health care process. Both computer aided systems and intelligent techniques are used to perform intelligent decision systems (IDS) [1]. Those systems used in health care field for medical purposes such as support intelligent diagnosing [2], disease early detection [3] [4]. And management and financial purposes such as manage hospital resources, control treatment cost and allocate medical team.

Machine learning may be categorized as supervised and unsupervised algorithms. They are used to manage and process huge amount of healthcare systems data. The unsupervised techniques are based on managing the data sets based on their similarity and repeated structure trying to find hidden pattern in unlabeled dataset which is defined as clustering. On the other hand, supervised techniques processed the datasets based on its previously known output. If the output is continuous, the regression algorithms are considered the best choice. However, classification algorithms are used to predict discrete outputs. Both categories are

widely used in the health care field [5], [6], [7]. Moreover, supervised and unsupervised techniques can be combined and worked together for providing better results [8]. These technique is called semi-supervised learning technique. For that, combining unlabeled data with some labelled data may results improving in the prediction accuracy [9].

Predicting breast cancer is a sensitive research area needs highly accurate prediction models. So, many intelligent techniques invented and developed for such target. One of the intelligent techniques is supervised machine learning which has its benefits such as: easy to understand, efficient training algorithm, tolerate noisy inputs but it also has its limitations [10]. So, artificial intelligent researchers interest in innovate solutions continuously develop its rules and working strategy by selection and recombination to find remarkable pattern on the existing dataset features to create high accurate decision system using robust algorithm as GA (genetic Algorithm) [11]. Predicting the occurrence of cancer is not the only interest of the researchers. However, predicting the surviving probability take an important place in the health field. Delen et al. [12], consider patient as survival after 60 month from the date of diagnosis. Using large dataset 200,000 record, they trained a model under the supervision of an expert. Moreover, researchers assigned a weight and a value for each feature according to its importance and effect on the model. The system is merged with meta-heuristic population-based optimization and ensemble learning to predict cancer recurrence during 5 years of diagnosis [13]. Combining Feature extraction and selection techniques improving model accuracy and saving training time. As well, The K-means algorithm is utilized to recognize the hidden patterns of the benign and malignant tumors separately then support vector machine (SVM) is used to obtain the new classifier to differentiate the incoming tumor [14]. Heng et al. used Jointly Sparse Discriminant Analysis (JSDA) to extract the key factors to enhance prediction model results [15]. This paper will discuss data mining and machine learning algorithms popularly used in predicting breast cancer.

Moreover, the effect of data preprocessing techniques, feature selection and normalization on the overall accuracy of the predicting model. This paper is organized as the following: In section 2, a background review about feature selection and recent machine learning technique is presented. In Section 3, a survey of the related work and their results, contribution and algorithms is provided. In Section 4, the proposed techniques are explained and presented. In Section 5, the experimental results are presented and discussed. Section 6 contains the conclusion of the paper.

2. Background Review

This section contains a brief description for features selection process and an explanation for the different classification techniques which are used in this work.

2.1. Feature selection

Feature selection is process of extracting most relevant features of the model for accurate prediction and removing unneeded, irrelevant and redundant attributes from data. Performing feature selection in model construction simplifies the model understanding for the researchers and minimize the processing time. Moreover, avoid the curse of dimensionality and over fitting problem [16]. Feature selection performed using three categories:

2.1.1 Wrapper Methods: wrapper feature selection methods combine random features subset to train the model. The error rate of the trained model decide which feature combination to be used. As, the lower error rate combination kept and the higher error rate combination features removed.

2.1.2 Filter Methods: filter feature selection method selects features by applying score for each feature. Highly ranked features considered to be kept to be final features. Filter feature selection could be used as preprocessing step for wrapper method in large dataset.

2.1.3 Embedded Methods: embedded feature selection method selects the features during the model construction.

2.2. Classification and Regression Trees (CART)

Classification and regression trees are designed to predict response for the input data by following the decision of the tree from the root to the leaf node. Classification trees predict nominal values (yes or no) while regression trees predict continuous values. Decision trees are easy to understand and considered an efficient training algorithm. However, they have a drawback when dealing with missing values as its class must be mutually exclusive [10]. CART techniques provide promising results in cancer prediction with microarray analysis [17] especially when combined with other techniques [18]. Moreover, when CART used as a part of CAD system with more relevant quantitative features for classification beside ultra sound images, it helps provide more better diagnose results [19]. It is possible to accurately build automatic classifiers from data if these data come from a single observer, in [20] the data based on FNA fine needle aspiration.

2.3. Artificial Neural Network (ANN)

ANN is a computational model based on the structure and functions of biological neural networks. This formation leads to use its benefits as tolerate noisy inputs. As well, it is used in classification and regression models. ANN has the ability to represent boolean functions [10] to create models that overcomes the difficulty of prognostic prediction by developing model its accuracy is near the actual data prediction [21]. Moreover, researchers keep improve techniques as Memetic Pareto Artificial Neural Network (MPANN) for better generalization and obtain much lower computational cost compared with other algorithms [22] [23]. Radial Basis Function Neural Network provide better result when compared with SVM [24].

2.4. Naive Bayes(NB)

Naive Bayes is one of the classification techniques that based on applying Bayes' theorem. As it works on minimize misclassification probability, it performs well when predicting a class that had been derived from the same data. However, it performed much worse than other techniques when used with classes never been trained because of its strong feature independence assumptions. To overcome this problem, so many techniques are developed to perform better results with (NB) [25] [26]. Besides, the selected feature in the algorithm play an important role in the accuracy of the prediction [27]. Therefore, (NB) is integrated with other algorithms such as decision tree to select a subset of attributes for the production of naive assumption of class conditional independence to increase the accuracy of the results [28].

2.5. Random Forest Trees (RF)

RF is an ensemble model that based on multiple decision trees. It is considered as one of the most accurate prediction techniques that work efficiently with large datasets with huge input variables. To improve the results of random forest pre- diction for multiclass disease classification [29], correlation- based feature selection, symmetrical uncertainty and gain

ratio are used to create improved model of RF. As the performance of the RF algorithm is affected by increase the strength and decrease the correlation of individual trees of the forest and to improve the function which determines how the outputs of the base classifiers are combined. Many modification used to achieve this by modify the node splitting and the voting procedure [30].

2.6. Support Vector Machine (SVM)

SVM is a machine learning algorithm that analyses data for classification and regression analysis. It is a supervised learning method that looks at data and sorts it into one of two categories. It generates a map of the sorted data with the margins between the two categories as far apart as possible. SVM works really well with clear margin of separation. It is effective in high dimensional spaces and cases where number of dimensions is greater than the number of samples. It uses a subset of training points in the decision function (called support vectors) which save memory. On the other hand, it needs a high training time when performed on large data set. As well, it has low accuracy when performed on noisy data set (i.e. target classes are overlapping). SVM doesn't provide probability estimates directly. Tingting et al. [31] uses three SVM classifiers combined with radial basis function (RBF) networks, and self-organizing maps (SOMs). The model evaluated using three error estimators to improve the prediction of breast cancer diagnosis dataset and the performance was over 98

2.7. Fuzzy Logic

Fuzzy sets are started with a set of if-then rules that converts to their mathematical equivalent. It provides simple, easy to train and flexible systems. It provides promising classification results especially when combined with other algorithms such as genetic algorithm [31] [32]. As well, it is used in predicting disease risk factors [33] using experts information.

3. Related Work

Many comparative researches had been proposed to evaluate the performance of breast cancer prediction systems. Sivagami [34] compared Support Vector Machine (SVM), Multilayer Perceptron Neural Network and One R and J48 Decision Tree Induction using linear, polynomial and RBF kernel with different parameter settings for *d*, *gamma* and C-regularization parameters. The parameters *d* and *gamma* are associated with polynomial kernel and RBF kernel respectively to determine the presence or the absence of Breast Cancer. The results of the algorithms show a higher performance of SVM algorithm with RBF kernel with accuracy up to 95 %.

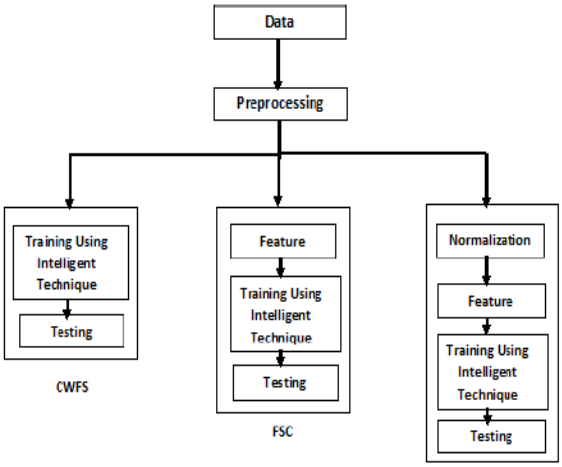

Figure 1: The proposed algorithms steps

Moreover different machine learning algorithms had been used in breast cancer prediction during the past 10 years. Kharya et.al. [10] states that Artificial Neural Network (ANN) considered as the most widely used technique. On one hand, SVM could not be used with large datasets due to its high computation in the training phase. On the other hand, bayesian network performs prediction under uncertainty with incomplete data. As well, decision trees are powerful algorithms used to predict breast cancer. There are four main trees

algorithms namely; j48, Classification and Regression Trees (CART), Alternating Decision Tree (AD Tree) and Best First Tree (BF Tree). The j48 algorithm has the best performance based on the proposed dataset [35].

The prediction accuracy varies from a model to another based on the input data type, the number of trained records and the used Machine learning algorithm [36]. Therefore, the integration of multidimensional heterogeneous data with the application of different techniques for feature selection and classification can provide promising results.

The behavior of a real ant colony simulated as Ant-Miner algorithm. It used to extract classification rules to be applied to unseen data as a decision aid. This algorithm produces good accuracy result with reducing the number of rules [37]. In order to increase the number of samples of the minority classes, some preprocessing techniques such as mega-trend diffusion (MTD) were developed and combined with the algorithms for better results [38]. Recently, Kulkarni et al. [39] used about 19 classifiers in combination with data pre-processing methods (string to nominal, PKI discretization, Numeric to binary, standardization, normalization and discretization). They found that a maximum accuracy of 74% achieved by using jRip Classifier with Numeric to Binary data pre-processing. For continuous real values, Ant-miner algorithm used discretization preprocessing. Although data pre-processing may leads to better results [37], some algorithms working well on the row datasets without the need to pre-processing such as the continuous and discrete features [40].

4. Proposed Techniques

In this section, three different techniques were developed and tested using two different sets of data. In the first technique, classification were performed without any feature selection technique. During the rest of this paper, we will call this technique CWFS (Classification without feature selection). However, a combination of feature selection techniques were used at the second technique. Therefore, we will call the second technique FSC (Feature Selection Classification). Moreover, the third technique used normalization technique along with the feature selection techniques. Hence, we will call this technique NFSC (Normalization and Feature Selection Classification), the proposed algorithms described in Fig.1. The results generated by the previous three algorithms are recorded and compared. The proposed algorithms explained in details in the next section and exposed to explain the five stages used in the algorithms:

- Preprocessing: Performed before using any of the three techniques.
- Normalization: Performed only at the third technique (NFSC).
- Feature selection: Performed at the second and the third techniques (FSC and NFSC).
- Training stage: performed at the three techniques
- Testing stage: Performed at the end of each technique.

4.1. Data Pre-processing

In this stage, the available data is examined and divided into training and testing data. The process is proceeds as follows:

4.1.1. Read and store the available data set (Algorithm 1, *RDM*)

Algorithm 1 Pre-processing Algorithm

1: ———— **Preprocessing** ————
2: *RDM* – Read the available data matrix M.
3: **for** each record in M **do**
4: *RMR* – Remove the record contains any missing value.
5: **end for**
6: **if** The algorithm is FSC **then**
7: ———— **Feature Selection** ————
8: *FSM* – $M_N = Feature\_selection(M)$
9: *SSM* – Save the new selected matrix M_N.
10: **end if**
11: **if** The algorithm is NFSC **then**
12: ———— **Normalization** ————
13: **for** each Column in M **do**
14: *CMV* – Calculate the maximum value M_X of the column and the minimum value M_M.
15: *CNV* – Calculate the normalized values using equation 1.
16: **end for**
17: *SNM* – Save the new normalized matrix M_Z.
18: *FSN* – $M_N = Feature\_selection(M_Z)$
19: *SNM* – Save the new matrix M_N.
20: ———— **End Normalization** ————-
21: **end if**
22: *DRM* – Divide the available matrix M_N randomly using k-fold cross validation.
23: *SOM* – Save the generated training matrix M_T and the testing matrix M_S.

Figure 2: Data preprocessing steps

4.1.2. Process the data and delete the record if it contains any missing value (Algorithm 1, *RMR*).

4.1.3. Randomly divide the data into training and testing data using K-fold cross validation (90% of the data are used for training and the remaining 10% for are used for testing) (Algorithm 1, *DRM*).

4.1.4. Save the generated training and testing matrices (Algorithm 1, *SOM*).

4.2. Classification without feature selection (CWFS)

In this technique, the process is divided into two main phases: Training and testing phases.

4.2.1. Training Phase: The target matrix M_T is generated to be used at the training process (Algorithm 2, *GTM*). As well, the training matrix M_T generated from the previous step is loaded to be used to train 9 different classification techniques (Algorithm 2, *LTM*). The matrix M_T is used as input to the technique *N* along with the target matrix *T* as output (Algorithm 2, *STR*). The generated training module *T M* is then saved in order to be used at the testing phase (Algorithm 2, *RTM*). The selected nine techniques used for training are namely, random forest, support vector machine, classification tree, regression tree, fuzzy rules, Naïve Bayes, Neural Network, K-nearest neighbor and tree bagger. The training process for each technique is proceeds as follows:

1) Random Forest (*RF*): create ensample model of 50 trees, using CART technique it is randomly select observation and features performing multiple CART and we could say the result is the mean of all predicted trees. Figure 3 shows the out-of-bag error decreases with the number of grown trees.

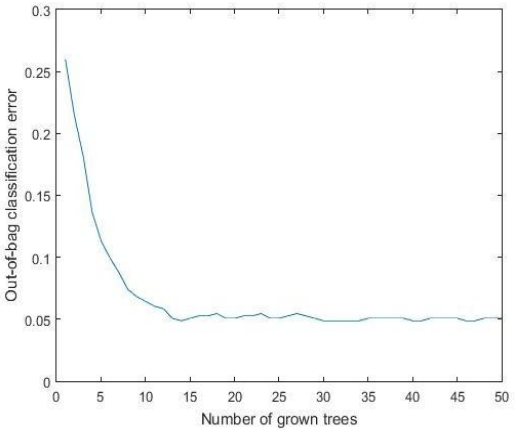

Figure 3: Out of bag errors

2) Classification tree (*CT*): each input variable represented as root and output variable represented as leaf node. The tree constructed by splitting the root nods in a process called recursive partitioning until meet stopping criteria. Figure 4 shows one of the training trees.

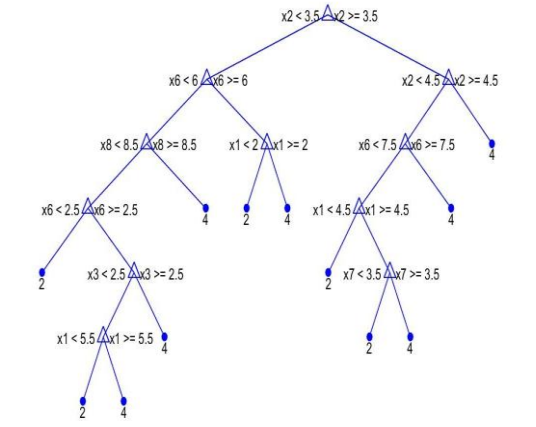

Figure 4: Classification tree training example

3) Fuzzy rules (*Fuzzy*): A Sugeno-type Fuzzy Inference System (FIS) is generated using subtractive clustering. The cluster center is set to 0.5.

4) Regression tree (*RT*): all records in the Training Set are grouped into one partition. The algorithm then starts dividing the data into two branches using each possible binary split on each field. After that, the algorithm selects the split that minimizes the sum of the squared deviations from the mean in the two separated partitions. This splitting rule is then applied to each of the new branches. This process continues until each node reaches a user- specified minimum node size and becomes a terminal node.

5) Support Vector Machine (*SVM*): A multi-class prediction model is created using error-correcting output codes (ECOC) model. ECOC uses SVM binary learners with one-versus-one coding design $\frac{K(K-1)}{2}$. For each binary learner (where K is the number of classes), one class is positive and another is negative and the software ignores the rest.
6) Neural Network (*NN*): The dataset is divided into 15% validation, 15% testing and the remaining 70% for training. Bayesian regularization back propagation function 'trainbr' is used for training. It minimizes a combination of squared errors and weights and then determines the correct combination to produce a well generalized network.
7) K-nearest neighbor (*KN*): predicting the most closest class for the tested record by calculating the shorter distance using distance measures as Euclidean.
8) Naïve Bayes (*NB*): *'fitcnb'* multi class training function is used for predicting the labels of the new data. A membership probabilities is predicted for each class and the class with the highest probability is considered the most likely class.
9) Tree bagger (*TB*): It combines the results of many decision trees to reduce the effects of over fitting and improves generalization.

4.2.2 Testing Phase: This phase starts by loading the testing matrix M_S generated at the preprocessing step (Algorithm 1, *SOM*, Algorithm 2, *LSM*). This matrix is used as input to the trained module *T M* and the resulted metrics *M C* is returned (Algorithm 2, *SSF*, *RMM*). As we mentioned before, the training process trying to figure out the relation between input and output data using the training algorithms. Each algorithm produce a model using the training matrix and the pre-defined target or output. This model used to predict the target of the testing matrix M_S. Different testing functions are used for each technique and the results are recorded in *M C*. The process is repeated 5 times and the average results are generated.

Algorithm 2 Training and Testing Algorithm

```
1: ———— Training ————
2: GTM –Generate the target matrix T
3: LTM – Load the training matrix M_T
4: CTF – Create the training function Train
5: for each algorithm in [CWFS, FSC, NFSC] do
6:    for each Technique N do
7:       STR – Train N with input M_T and output T.
         TM = Train (Technique N, M_T, T)
8:       RTM – Return the training module TM
9:    end for
10: end for
11: ———— Testing ————
12: LSM – Load the testing matrix M_S
13: CSF – Create the testing function Test
14: for each algorithm in [CWFS, FSC, NFSC] do
15:    for each Trained module TM do
16:       SSF – Test TM, with input M_S and save the output
          metrics at the matrix MC.
          MC=Test ( TM, M_S)
17:       RMM –Return the calculated metrics MC
18:    end for
19: end for
```

Figure 5: Training and testing steps

4.3.Feature Selection Classification (FSC)

In this technique, a set of feature selection techniques are applied on the matrix *M*. The new matrix M_N is then used for training and testing (Algorithm 1, *FSM, SSM*),

4.3.1 Feature selection*:* Feature selection is the process of selecting the most relevant features for the prediction model. First, the data matrix *M* is passed to a feature selection function (Algorithm 1, *FSM*). This function applies five different feature selection techniques on *M*. Each technique will generate its selected features and the feature that selected by at least two techniques is considered a final feature, figure 6. The five feature selection techniques used was:

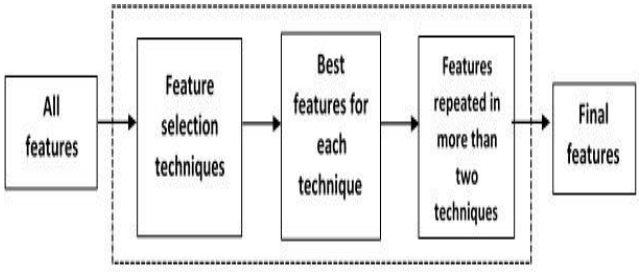

Figure 6: Feature selection process

- Laplacian score: [41]
- Spectral: [42]
- MCFC:
- Greedy: [43]
- feature similarity: [44]

4.3.2 Training*:* This process is performed exactly the same like what is explained in section IV-B1 and Algorithm

4.3.3 Testing*:* This process is performed exactly the same like what is explained in section IV-B2 and Algorithm

4.4.Normalization and Feature Selection Classification (NFSC)

4.4.1 Normalization*:* In this technique, the data matrix *M* is normalized using Min-Max normalization technique (Algorithm 1, *CMV, CNV*. Min-Max is a normalization strategy which linearly transforms x to y using the following equation:

$$y = \frac{x - min}{\max - min} \quad (1)$$

Where min and max are the minimum and maximum values in *X* and *X* is the set of observed values of *x* (*X* is a single column at *M*). It can be easily seen that if $x = min$ then $y = 0$ and if $x = max$ then $y = 1$. Therefore, the generated matrix M_N contains values that ranges between 0 and 1 (Algorithm 1, *SNM*).

4.4.2 Feature selection*:* This process is performed exactly the same like what is explained in section IV-C1. However, the feature selection matrix will applied on the matrix M_Z (generated after normalization) instead of *M* (Algorithm 1, *FSN*.

4.4.3 Training*:* This process is performed exactly the same like what is explained in section IV-B1 and Algorithm 2.

4.4.4 Testing*:* This process is performed exactly the same like what is explained in section

IV-B2 and Algorithm 2.

5. Experiments and Results

In this experiment, two data sets were used separately to evaluate the proposed techniques. The first one is Wisconsin Breast Cancer Database (WBC) that consists of 699 record with 9 features beside the ID field and their target class (2 for benign and 4 malignant). The second dataset is Wisconsin Diagnostic Breast Cancer (WDBC) consists of 569 record with 30 feature and their target class from machine learning repository (UCI). Each technique from the nine techniques is used inside the three proposed techniques (*CWF S*, *F SC*, *N F SC*) to evaluate both datasets. Hence, each technique is used tree times to come up with 27 different experiment. For each experiment, the process is repeated five times and the average results are calculated and presented.

5.1 Evaluation Metrics

All algorithms are evaluated according to the terms of accuracy, precision, sensitivity and specificity. These terms are calculated and constructed using the confusion matrix elements as follows:

- True positive (TP): Correct positive prediction
- False positive (FP): Incorrect positive prediction
- True negative (TN): Correct negative prediction
- False negative (FN): Incorrect negative prediction

4.1. Accuracy: Is calculated as the summation of all correct predictions divided by the total number of the dataset.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

4.2. Sensitivity: Is calculated as the summation of correct positive predictions divided by the total number of positives.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

4.3. Specificity: Is calculated as the summation of correct negative predictions divided by the total number of negatives.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4)$$

4.4. Precision is calculated as the summation of correct positive predictions divided by the total number of positive predictions.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

The model is simulated without feature selection (CWFS), with feature selection (FSC) and normalized dataset with feature selection (NFSC). Each proposed model is run for five iterations then calculate the mean and standard deviation of the used algorithms to calculate the evaluation parameters:

5.2 The Presented Results

In this section, the results of the three proposed algorithms *CW F S*, *F SC* and *N F SC* performed over the two data sets (WBC and WDBC) are calculated and presented. For each algorithm, the experiment is repeated five times and the average of the results are presented.

5.2.1 WBC results*:* The results of applying the three proposed algorithms on the first dataset (WBC) are presented as follows:

1) CWFS results: In this algorithm, most of the nine techniques managed to achieve a high prediction results for the evaluation metrics Figure 7. In this experiment, NB was the best algorithm with accuracy up to 97% ± 2.64 %, precision of 97% ± 3%, sensitivity of 98% ± 2% and specificity of 98% ± 2%. The KN, TB and SVM are then ordered respectively in terms of their accuracy. The lowest accuracy results were generated by fuzzy and NN algorithms. Moreover, fuzzy and neural networks in addition to regression tree algorithm had the lowest precision. Therefore, they had the minimum correctly predicted classes from positive predicted classes. As well, NN classified only 20% of the total predicted yes classes as true.

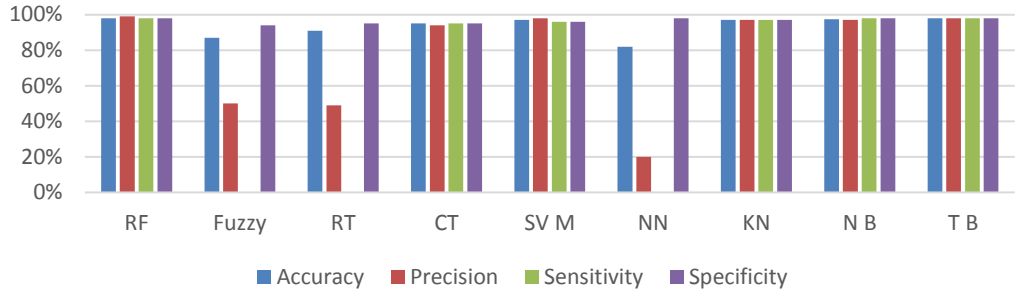

Figure 7: Results of the first training set without feature selection CWFS

2) FSC results: By performing feature selection on the training dataset II the best predicted algorithms were TB and RF with accuracy up to 98% ±1.68%. However, the results of their precision, sensitivity and specificity were 98% ± 2%. NB came at the third place with accuracy 97.35% ± 1.6%. Moreover, the applying of feature selection managed to improve the accuracy results for the fuzzy, RT and NN algorithms up to 5% compared to the results achieved by CWFS. However, the precision of fuzzy and RT did not affected by applying the feature selection. On the other hand, the precision of NN increases to reach over 39% ±22% of the truly predicted classes compared to CWFS. Moreover, the specificity and sensitivity have a little progress for most of algorithms, Figure 8.

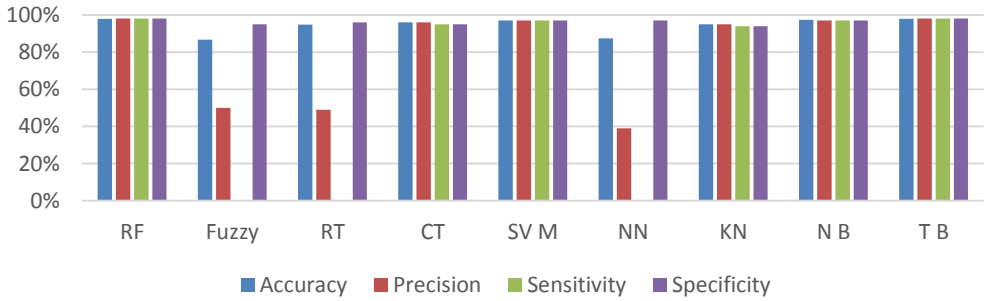

Figure 8: Results of the first training set with feature selection (FSC)

3) NFSC results: Figure 9 summarize the results of applying the third algorithm where normalizing the data is performed before the feature selection process. It can be noticed that the best results was achieved by TB algorithm with accuracy up to 97.09% ±1.02. The second place was gone to RF algorithm with accuracy up to 96.8% ± 1.2%. The remaining seven algorithms were ordered as NB with accuracy up to 96.21% ± 1.28% then KN with accuracy 95.34 ± 1.89% then SVM, CT, RT, fuzzy and NN.

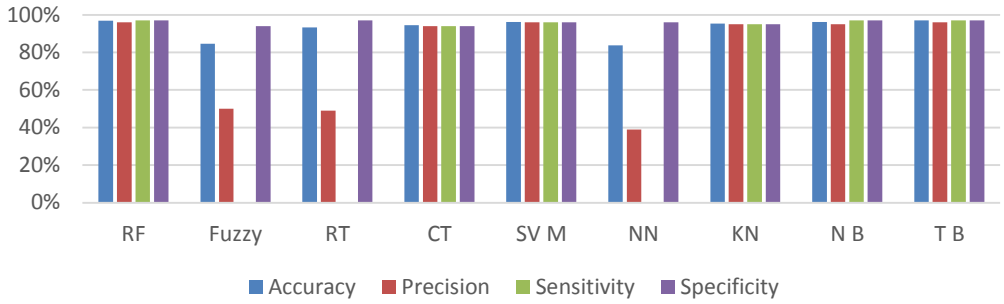

Figure 9: Results of the first training set with feature selection and normalization NFSC)

In conclusion, for the first data set it is clear that the best results were achieved by the RF technique at two out of the three proposed algorithm. It generates accuracy up 98 ±2%, 97.9 ± 1.67% and 96.8 ± 1.2% at CWFS, FSC and NFSC respectively, then NB accuracy is 97.36 ± 2.64%. However, TB technique generate the best results at the third algorithm and the same results as RF at the second algorithm. Moreover, it is noticed that applying feature selection at the second algorithm managed to increase the accuracy of RT, CT and NN. However, normalizing the data before the feature selection managed to increase the accuracy of RT, NN and KN only.

5.2.2 WDBC results

1) CWFS results: Figure 10 summarize the results of applying the first proposed algorithms on the WDBC dataset. It can be seen that, the best results was generated by TB technique with accuracy up to 96.84% 2.29%, precision up to 96 ± 3%, sensitivity up to 97 ± 3% and 97 ± 3% for specificity. RF technique came at the second place with accuracy up to 96.49 ± 3.04%. It is remarkable that, SVM algorithm generate a low accuracy of 67 ± 4.54%. As well, the lowest sensitivity and specificity was generated by SVM.

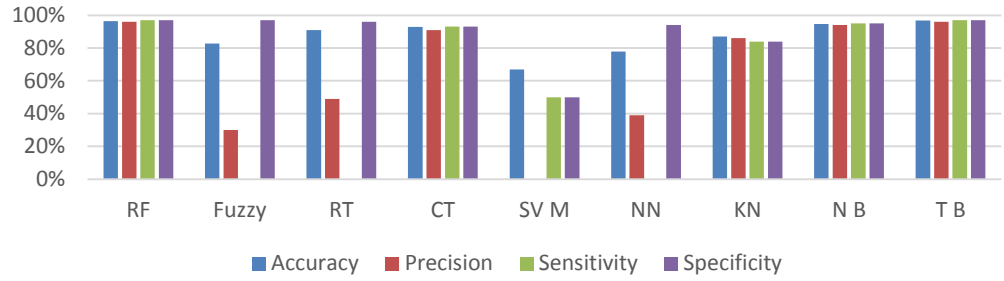

Figure 10: Results of the second training set without feature selection (CWFS)

2) FSC results: Figure 11 present the results of the nine techniques applied after the feature selection process was performed. The best accuracy was 96.14 ± 2.6% obtained by the two algorithms RF and TB. Also, they had the same predictive results for the terms sensitivity, specificity and precision. On the other hand, the lowest results was obtained by SVM technique with accuracy up to 61 ± 4.54%.

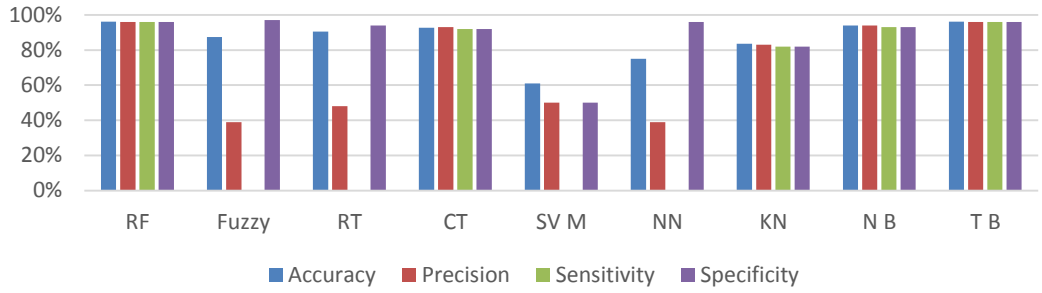

Figure 11: Results of the second training set with feature selection (FSC)

3) NFSC results: Figure 12 present the results of the nine algorithms applied on a set of selected features from normalized data. It is clear that the feature selection after data normalization managed to improve the results of the techniques. For example, the results of SVM increased from 67% at CWFS and 61% at FSC to up to 98 ± 1.57%. As well, there are a great improvement on the results of KN and a slight improvement on the results of NB.

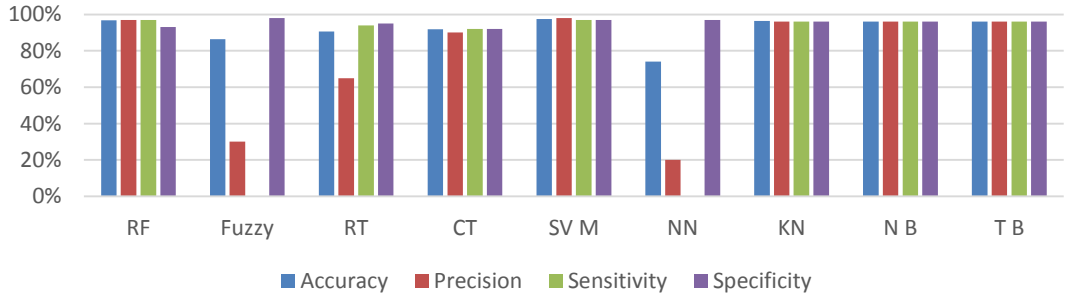

Figure 12*:* Results of the second training set with feature selection and normalization (NFSC)

In conclusion, the proposed algorithm generates different results for different techniques when applied on different data. The first dataset has a low number of features (9 features). Moreover, the values of the data are range from 1 to 10. Therefore, the systems with feature selection and/or normalization did not add a lot to the results. For example, RF generates the best result accuracy of 98% and standard deviation of 0.8% without feature selection and normalization. However, feature selection managed to improve the performance of some techniques such as RT and NN but some other techniques perform better without feature selection such as RF, Fuzzy, CT, SVM, K-nearest and Naïve Bayes . On the other hand, as the number of features increased at the second dataset, feature selection and normalization play a role in enhancing the performance of the overall model. For example, the accuracy of RF, SVM, K-nearest and tree bagger is increased when the data is normalized and the features are selected. Hence, the feature selection techniques and preprocessing techniques are positively affect the prediction accuracy according to the number of feature and the range of the feature values.

6. Conclusion

Predicting breast cancer in an early stages increases the percentage of surviving and saves more lives. By using data mining and machine learning techniques, early diagnose and effective treatment as well as surgical intervention may be achieved. A set of popular data mining and intelligent algorithms (e.g., Random Forest, Support Vector Machine,

Classification Tree, Regression Tree, Fuzzy, Naive Bayes, Neural Network, K- nearest and Tree Bagger) are used to create a classification and a prediction model to predict breast cancer. Three different algorithms are proposed and applied on two different sets of data. These algorithms are CWFS (Classification without feature selection), FSC (Feature Selection Classification and NFSC (Normalization and Feature Selection Classification). These algorithms are used to explore the effect of using the feature selection and normalization of the classification accuracy. At the first dataset, normalization and feature selection could not managed to improve the accuracy because of the small number of features available at the first data set (9 features). However, they managed to improve the results of the algorithms at the second data set with 30 features.

References

[1]. M. Abdel Badie and H. N. Elmahdy "Enhancing the life quality of elderly using Ambient Intelligent Technology (AmIT)," Egyptian Computer Science Journal Vol.41 No.3 September 2017.

[2]. F. Gorunescu, "Intelligent decision systems in medicineâ ĂŤa short sur- vey on medical diagnosis and patient management," in *E-Health and Bioengineering Conference (EHB), 2015*, pp. 1–9, IEEE, 2015.

[3]. H. Kaur and S. K. Wasan, "Empirical study on applications of data mining techniques in healthcare," *Journal of Computer science*, vol. 2, no. 2, pp. 194–200, 2006.

[4]. H. Mohsen, E. A. Eldahshan, E. M. El-horbaty and A. M. Salem " Classification of Brain MRI for Alzheimer's Disease Based on Linear Discriminate Analysis" Egyptian Computer Science Journal Vol. 41 No.3 September 2017.

[5]. R. Brydges, A. Dubrowski, and G. Regehr, "A new concept of unsuper- vised learning: directed self-guided learning in the health professions," *Academic Medicine*, vol. 85, no. 10, pp. S49–S55, 2010.

[6]. X. Wang, D. Sontag, and F. Wang, "Unsupervised learning of disease progression models," in *Proceedings of the 20th ACM SIGKDD interna- tional conference on Knowledge discovery and data mining*, pp. 85–94, ACM, 2014.

[7]. S. Anto, "Supervised machine learning approaches for medical data set classification-a review 1," 2011.

[8]. M. Nilashi, O. Ibrahim, H. Ahmadi, and L. Shahmoradi, "A knowledge- based system for breast cancer classification using fuzzy logic method," *Telematics and Informatics*, vol. 34, no. 4, pp. 133–144, 2017.

[9]. L. Peng, W. Chen, W. Zhou, F. Li, J. Yang, and J. Zhang, "An immune-inspired semi-supervised algorithm for breast cancer diagnosis," *Computer Methods and Programs in Biomedicine*, vol. 134, pp. 259–265, 2016.

[10]. S. Kharya, D. Dubey, and S. Soni, "Predictive machine learning tech- niques for breast cancer detection," *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 4, no. 6, pp. 1023– 1028, 2013.

[11]. T.-C. Chen and T.-C. Hsu, "A gas based approach for mining breast cancer pattern," *Expert Systems with Applications*, vol. 30, no. 4, pp. 674–681, 2006.

[12]. D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivabil- ity: a comparison of three data mining methods," *Artificial intelligence in medicine*, vol. 34, no. 2, pp. 113–127, 2005.

[13]. M. R. Mohebian, H. R. Marateb, M. Mansourian, M. A. Mañanas, and F. Mokarian, "A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (hpbcr) using optimized ensemble learning," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 75–85, 2017.

[14]. B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of k-means and support vector machine algorithms," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1476– 1482, 2014.

[15]. H. Kong, Z. Lai, X. Wang, and F. Liu, "Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning, " *Neurocomputing*, vol. 177, pp. 198–205, 2016.

[16]. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[17]. M. B. Al Snousy, H. M. El-Deeb, K. Badran, and I. A. Al Khlil, "Suite of decision tree-based classification algorithms on cancer gene expression data," *Egyptian Informatics Journal*, vol. 12, no. 2, pp. 73–82, 2011.

[18]. K.-H. Chen, K.-J. Wang, K.-M. Wang, and M.-A. Angelia, "Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data," *Applied Soft Computing*, vol. 24, pp. 773–780, 2014.

[19]. W. K. Moon, I.-L. Chen, J. M. Chang, S. U. Shin, C.-M. Lo, and R.-F. Chang, "The adaptive computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound," *Ultrasonics*, vol. 76, pp. 70–77, 2017.

[20]. N. Cruz-Ramírez, H.-G. Acosta-Mesa, H. Carrillo-Calvet, and R.-E. Barrientos-Martínez, "Discovering interobserver variability in the cyto- diagnosis of breast cancer using decision trees and bayesian networks," *Applied Soft Computing*, vol. 9, no. 4, pp. 1331–1342, 2009.

[21]. W. N. Street, "A neural network model for prognostic prediction.," in *ICML*, pp. 540–546, 1998.

[22]. H. A. Abbass, "An evolutionary artificial neural networks approach for breast cancer diagnosis," *Artificial intelligence in Medicine*, vol. 25, no. 3, pp. 265–281, 2002.

[23]. W. Duch and R. Adamczak, "Statistical methods for construction of neural networks.," in *ICONIP*, pp. 639–642, 1998.

[24]. T. Subashini, V. Ramalingam, and S. Palanivel, "Breast mass classifica- tion based on cytological patterns using rbfnn and svm," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5284–5290, 2009.

[25]. D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli, and I. O. Ellis, "A âĂŸnon-parametricâĂ Ź version of the naive bayes classifier," *Knowledge-Based Systems*, vol. 24, no. 6, pp. 775–784, 2011.

[26]. B. Haouari, N. B. Amor, Z. Elouedi, and K. Mellouli, "Naïve possibilistic network classifiers," *Fuzzy Sets and Systems*, vol. 160, no. 22, pp. 3224– 3238, 2009.

[27]. B. Durgalakshmi and V. Vijayakumar, "Progonosis and modelling of breast cancer and its growth novel naive bayes," *Procedia Computer Science*, vol. 50, pp. 551–553, 2015.

[28]. D. M. Farid, L. Zhang, C. M. Rahman, M. A. Hossain, and R. Stra- chan, "Hybrid decision tree and naïve bayes classifiers for multi-class classification tasks," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1937–1946, 2014.

[29]. A. Chaudhary, S. Kolhe, and R. Kamal, "An improved random forest classifier for multi-class classification," *Information Processing in Agri- culture*, vol. 3, no. 4, pp. 215–222, 2016.

[30]. E. E. Tripoliti, D. I. Fotiadis, and G. Manis, "Modifications of the construction and voting mechanisms of the random forests algorithm," *Data & Knowledge Engineering*, vol. 87, pp. 41–65, 2013.

[31]. T. Mu and A. K. Nandi, "Breast cancer detection from fna using svm with different parameter tuning systems and som–rbf classifier," *Journal of the Franklin Institute*, vol. 344, no. 3, pp. 285–311, 2007.

[32]. A. Alharbi and F. Tchier, "Using a genetic-fuzzy algorithm as a computer aided diagnosis tool on saudi arabian breast cancer database," *Mathemat- ical Biosciences*, vol. 286, pp. 39–48, 2017.

[33]. A. Büyükavcu, Y. E. Albayrak, and N. Göker, "A fuzzy information- based approach for breast cancer risk factors assessment," *Applied Soft Computing*, vol. 38, pp. 437–452, 2016.

[34]. P. Sivagami, "Supervised learning approach for breast cancer classi- fication," *International Journal of Emerging Trends & Technology in Computer Science*, vol. 1, no. 4, 2012.

[35]. E. Venkatesan and T. Velmurugan, "Performance analysis of decision tree algorithms for breast cancer classification," *Indian Journal of Science and Technology*, vol. 8, no. 29, 2015.

[36]. K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8–17, 2015.

[37]. R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "An ant colony based system for data mining: applications to medical data," in *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, pp. 791–797, Morgan Kaufmann Publishers Inc., 2001.

[38]. A. Majid, S. Ali, M. Iqbal, and N. Kausar, "Prediction of human breast and colon cancers from imbalanced data using nearest neighbor and support vector machines," *Computer methods and programs in biomedicine*, vol. 113, no. 3, pp. 792–808, 2014.

[39]. S. Kulkarni and M. Bhagwat, "Predicting breast cancer recurrence using data mining techniques," *International Journal of Computer Applications*, vol. 122, no. 23, 2015.

[40]. K. Gra¸bczewski and W. Duch, "Heterogeneous forests of decision trees," in *International Conference on Artificial Neural Networks*, pp. 504–509, Springer, 2002.

[41]. X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Advances in neural information processing systems*, pp. 507–514, 2006.

[42]. Z. Zhao and H. Liu, "Spectral feature selection for supervised and unsupervised learning," in *Proceedings of the 24th international conference on Machine learning*, pp. 1151–1157, ACM, 2007.

[43]. A. K. Farahat, A. Ghodsi, and M. S. Kamel, "Efficient greedy feature se- lection for unsupervised learning," *Knowledge and information systems*, vol. 35, no. 2, pp. 285–310, 2013.

[44]. P. Mitra, C. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 301–312, 2002.