

A Composite Hybrid Feature Selection Learning-Based Optimization of Genetic Algorithm For Breast Cancer Detection

Ahmed Abdullah Farid^{1,*}, Gamal Ibrahim Selim^{1,2}, and Hatem A. Khater³

¹Department of Computer Engineering, Faculty of Engineering & Technology, Arab Academy for Science Technology and Maritime Transport(AASTMT), Cairo, Egypt

²(email: dgamal55@yahoo.com or dgamalselim@aast.edu)

³ Faculty of Engineering, Horus University in Egypt (email: hkhater@horus.edu.eg)

*Corresponding author, (email: ahmed.abdullah.farid@gmail.com or portal.ahmed@student.aast.edu)

ABSTRACT

Breast cancer is a significant health issue across the world. Breast cancer is the most widely-diagnosed cancer in women; early-stage diagnosis of disease and therapies increase patient safety. This paper proposes a synthetic model set of features focused on the optimization of the genetic algorithm (CHFS-BOGA) to forecast breast cancer. This hybrid feature selection approach combines the advantages of three filter feature selection approaches with an optimize Genetic Algorithm (OGA) to select the best features to improve the performance of the classification process and scalability. We propose OGA by improving the initial population generating and genetic operators using the results of filter approaches as some prior information with using the C4.5 decision tree classifier as a fitness function instead of probability and random selection. The authors collected available updated data from Wisconsin UCI machine learning with a total of 569 rows and 32 columns. The dataset evaluated using an explorer set of weka data mining open-source software for the analysis purpose. The results show that the proposed hybrid feature selection approach significantly outperforms the single filter approaches and principal component analysis (PCA) for optimum feature selection. These characteristics are good indicators for the return prediction. The highest accuracy achieved with the proposed system before (CHFS-BOGA) using the support vector machine (SVM) classifiers was 97.3%. The highest accuracy after (CHFS-BOGA-SVM) was 98.25% on split 70.0% train, remainder test, and 100% on the full training set. Moreover, the receiver operating characteristic (ROC) curve was equal to 1.0. The results showed that the proposed (CHFS-BOGA-SVM) system was able to accurately classify the type of breast tumor, whether malignant or benign.

Keywords: Data Mining; Breast Cancer; Hybrid Feature Selection; Machine learning; Support Vector Machine; Optimize Genetic Algorithm, boosting algorithms

1. INTRODUCTION

The most-reported disease of females is breast cancer. Breast cancer may occur in both men and women, but in women, it is much more severe. Breast cancer is a type of cancer that occurs in breast cells. It is cancer that grows from breast tissue. Signs of breast cancer can include swelling in the breast, a difference in the form of the breast, skin dimpling, blood leaking from the nipple, and then metastasizing to other body regions. Among males and females of all genders, it is the most common 12.3 % of all disease. It is the most common among women in the world, making up 25.4 % of the total number of new cases reported in 2018 (Bray F, Ferlay J, 2018). Breast cancer is one of the most prominent tumors among women today. The number of cancer cases predicted in 2025 to reach 19.3 million, according to the World Health Organization (WHO). While the numbers of women with breast cancer remain increasing enormously, breast cancer death rates have declined recently, because of developments in medical imaging, image processing, and machine learning techniques. Which helped radiologists to detect early cancer. Thus it is essential to identify breast cancer early. A late breast cancer diagnosis enhances patient results and follow-up preparation. It also lowers death rates among women because 96 % of cancers are curable in the initial stages, according to estimates (Ragab & Sharkas, 2019).

Efficient breast cancer detection is critical for effective treatment and reduction of breast cancer deaths. Sometimes, a breast tumor is there, but not cancer; a benign tumor called. Benign or malignant can be a tumor (1) Benign tumor is a mass of cells that can not invade or metastasize the surrounding tissue. It does not spread or strike surrounding tissues (2) Malignant is different tissues. Cells in the blood or lymph system can break away and disperse to lymph glands and distant parts of the body (Bray F, Ferlay J, 2018). Data mining skills involved in biomedical sciences and investigate for providing prediction for help to identify the disease and classify it correctly (Kalló, Gergő & Miklós Emri et al., 2016). In the first phase, we propose a new composite hybrid feature selection approach based on optimizing the genetic algorithm (CHFS-OGA) for better breast cancer disease prediction. In the second stage, we applied the features of the output data set from step one to an architectural classification model to improve classification accuracy. The next section discusses the literature review of other authors who have used data mining and its relative of machine learning algorithm to analyze and diagnose breast cancer and various diseases. Section 3 describes the proposed algorithm used for feature extraction. Section 4 describes the hybrid classification process techniques used to improve the accuracy of breast cancer diagnosis, whereas section 5 describes the experiments and discuss the results. Finally, section 6 presents the paper summary and conclusions.

2. RELATED WORK

The following authors used some popular techniques. The author Iranpour examined the application of SVM, a Radial Base Function (RBF), and obtained 98.1% precision compared to accuracies obtained in other studies such as the linear SVM classifier (94%), the fluorescent classifier (95.8%), the closest neighbor edited with pure filtering (95%) (Iranpour, 2007). The three-dimensionality elimination approaches for Wisconsin Diagnostic Breast Cancer (WDBC), wine, and zoo data sets have contrasted by (Veerabhadrapa et al., 2010). In both approaches adopted, features based on mutual similarities chosen for dimensionality reduction level 1, and selected features for level 2 used to eliminate features using Principal Component Analysis (PCA) and Locality Preserving Projections (LPP) frameworks. The average F-measure for the

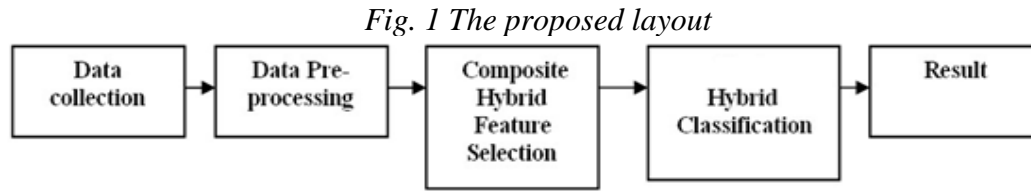
winery, zoos, and breast cancer was 92.950, 85.146, and 87.073, respectively, whereas the average F-measure for LPP was 95.148, 91.898, and 89.752 respectively.

The paper collects numerous studies and scientific papers on the diagnosis and prognosis of breast cancer and reports on current research, which utilizes the data mining tools for enhancing the diagnosis and prognosis of breast cancer. Results in this paper indicate that the decision-tree is the best predictor, 93.62 % Classifier, Neural Network, and Naïve Bayes Classifier, SVM, Bayesian networks, and Logistic Regression and that 93.62 % Classification on a comparison dataset and SEER data sets are the best predictor (Sarvestan Soltani et al., 2010). Besides, the selection of functions (FS) is a fundamental machine learning process. It is widely used in medical applications to reduce data size and growing needless and irrelevant characteristics. Therefore, FS only selects the characteristics that can influence a predictive model's classification results. Many FS strategies for breast cancer screening practices have been studied and proposed. (Fu, J.C, 2005) have used an algorithm of sequential forward scan (SFS) accompanied by the classification of the SVM and general neural regression network (GRNN). The methodology recorded achieved the area under the curve (AUC) of 0.9800 and 0.9780, respectively, for SVM and GRNN after choosing the essential features.

Five new natural-inspired algorithms were introduced by (Osman Hegazy & Omar, 2015) to optimize and train the Least Square Vector Support (LS-SVM) systems. Such algorithms are Flower Pollination Algorithm (FPA), Bat Algorithm (BA), Moded Cuckoo Quest (MCS), Artificial Bee Colony (ABC). The results showed that the models implemented at the early stages of the iteration have a quick convergence rate. For price and pattern estimation, they achieved better precision than comparable approaches. They also solve the problems of relevant and local minima found in ANN and LS-SVM. (Pawar and Talbar, 2016) suggested a genetic fuzzy FS strategy, resulting in 89.47 % precision. (Mohanty et al., 2018) proposed a forest optimization FS algorithm. Classifying normal and abnormal mammographic lesions is performed using various classifiers such as SVM, K-nearest neighbor (k-NN), Naïve Bayes, and C4.5. The maximum accuracy for the C4.5 classification for the DDSM data collection was 99.08 %. The writers have used the MIAS dataset and obtained 97.86 % accuracy for the classification of Naïve Bayes. (Badr et al., 2019) Using modern hybrid approaches such as Support Vector Machine (SVM) and Gray Wolf Optimizer (GWO) to diagnose the type of tumor in the breast, whether malignant or benign. Equate the predictive efficiency of this hybrid model with SVM for the breast cancer data from UCI machine learning accessible from Wisconsin with a minimum of 569 rows, and 32 columns, the maximum accuracy of the GWO- SVM is 97.0%.

3. PROPOSED WORK

In the proposed work, the breast cancer dataset collected from Wisconsin UCI machine learning with a total of 569 rows and 32 columns (Dua & D. & Graff & C., 2019). The breast cancer dataset contains Features that measured from a visual image of a finely suctioned needle (FNA) of a breast mass. The characteristics of the cell nuclei in the picture are defined —the proposed model layout in Fig 1.



The following steps explain the mechanism of the proposed work:

3.1. Data Collection

The author's collected the dataset form UCI and considered a benchmark repository of machine learning and artificial intelligence community. The breast cancer dataset can describe as follow of feature Information by separating the plane described above obtained using Multisurface Method-Tree (MSM-T) with a system of classification that uses linear programming for decision tree creation. A thorough search of 1-4 characteristics and 1-3 splitting planes found the appropriate features (Dua& D. & Graff&C. ,2019):

- ID number
- Class of diagnosis (M = malignant, B = benign)
- 32 Ten real-valued features computed for each cell nucleus:
 - (a) Radius (mean distance between the middle and perimeter points)
 - (b) texture (standard deviation from gray-scale values)
 - (c) perimeter
 - (d) surface area
 - (e) smoothness (local difference from the radius length)
 - (f) compactness ($\text{perimeter}^2 / \text{surface area} - 1.0$)
 - (g) concavity (series of concaved contour portions)
 - (h) concave points (number of the concave portion)

3.2. Data Pre-processing

In the real world, data collected tend to be not wholly complete, noisy, and conflicting, detection missing of data, data irregularity. Prevent the errors and decrease the data to be analyzed leads to massive payouts for decision making(s.r.bahgya,2014) as primary data has to collect in such a way that the above problems not occur. The missing entries in the collected dataset filled up by using the average values, and the author used unsupervised attribute replace missing values filters in the weka data mining tool to solve this. The data is often present in the form of a spreadsheet. However, weka native data storage format is ARFF and transformed from a spreadsheet to CSV format. After that, the CSV file converted to the ARFF file. Thus the data has to be transformed from spreadsheet format to ARFF format (H. Witten & Eibe Frank ,2008) .

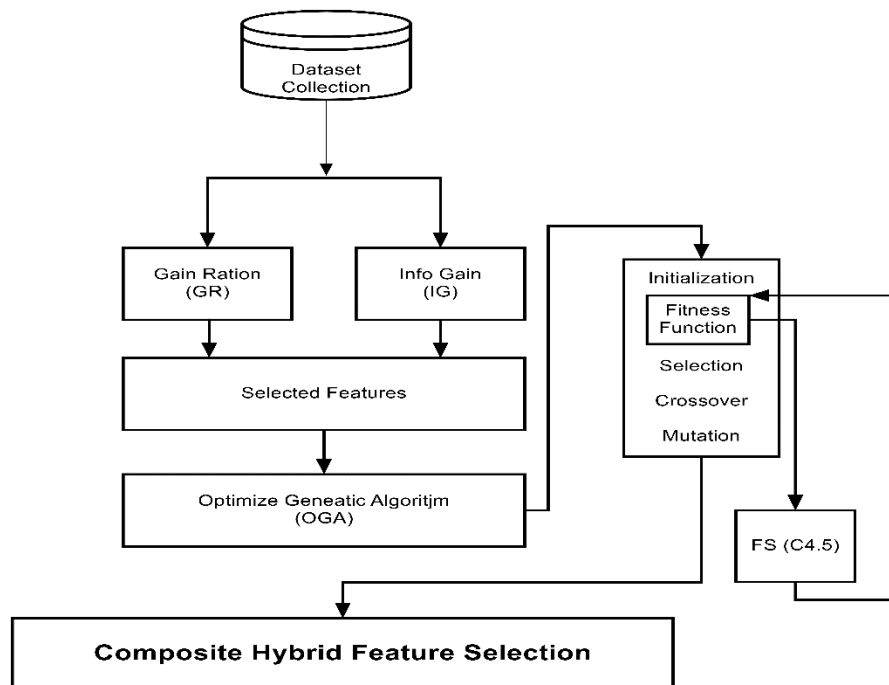
3.3. Proposed Composite Hybrid Feature Selection Model (CHFS)

Feature selection is the approach of taking a subset of relevant features for use in model construction(Chen & ...& F. Li,2010). It combines the advantages of three feature selection approaches (Filter (IG,GR)–Wrapper(improved (Genetic Algorithm)) with Embedded(C4.5)).

3.3.1 Composed Hybrid feature selection architecture

The author combine of three feature extraction technique considered for the optimal selection feature set, and this method is information gain (IG) - gain ratio (GR) and Optimized Genetic Algorithm(Huang & C.,2012) (Yanan Mao & Dingyuan Fan,2016).as shown by fig 2.

Fig. 2 The proposed composite hybrid feature selection architecture



In Figure 2, the feature extraction method using gain ratio- information gain to rank the attributes. The superior ranked features are taken as the optimized subset. The threshold established on the size of the resumption feature subset and amount of attributes in the dataset that appear in the range (7~10), and using classifier subset with C4.5 decision tree method in early-stage to obtain the optimal subset and vote for best feature selection of attributes in an automated process and produced an optimized feature subset with genetic search elevation planning to recognize the features that give the most predictable for the target class (Vemuri & et al.Jack & C. R.,2008) (Devi &Cuckoo ,2013).

3.3.2 Information Gain feature selection

The calculation of the information gain for only one attribute according to the algorithm below (Aouatif Amine & ...& Rziza Driss ,2011) : This gain measure gives the effect of the features, and the following algorithm selects features that are larger than the threshold.

Let S be a part of a training set sample with related labels. The m class in training set has si pattern of class I and s have overall patterns in training set predictable information looked-for to grade using (1)(r.chaves&jami,2012)

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m \left(\frac{s_i}{s} \right) \log_2 \frac{s_i}{s} \quad (1)$$

Feature F with (f1, f2, and f3 to fv) can separate the medical dataset into v subsets {S1, S2... Sv} where Sj and have value fj for the feature F and Sj include sij samples of class i

The entropy of F by"(2)"

$$E(f) = - \sum_{i=1}^m \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{s} I(s_{1j}, s_{2j}, \dots, s_{mj}) \quad (2)$$

And calculate info. Gain by (3)

$$GAIN(F) = I(S_1, S_2, \dots, S_3) - E(F) \quad (3)$$

This gain measure gives the effect of the features, and the following algorithm selects features that are larger than the threshold(Yanan Mao,2016)(William,1990)and shown in fig 3.

Fig. 2 The algorithm of information gain feature selection

```

Procedure Information Gain data mining feature selection Algorithm (IGFS)
var1  sfl /* store selected feature. Initially empty*/
var2  Th  /* hold threshold value*/
var3  f(i) /* contains the ith feature of the data set */
1: begin
2: IGFS.build ();
3: begin
4:   sfl = {}; /* create array of var1*/
5:   for loop i=1 to int of features
6:     INF=compute (IG) for the feature /* store computed features*/
7:     Gain (i) =INF /* compute Gain (i) */
8:   end for
9:   Th= threshold value /* hold threshold value*/
10:  For i= 1 to number of features
11:    If gain (i) > Th then
12:      Sfl=sfl+f {i} /* store and compute every feature in dataset*/
13:    end if
14:  end for
15: end

```

3.3.3. Gain Ratio Feature Selection

A decision tree can be a simple form when non-terminal nodes perform tests on many attributes to the effect of decision outcomes (J.R. Quinlan,1986), as shown in fig 4.

Let Q set of q data and m is a class and can classify data by

$$I(Q) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (4)$$

Qij is many samples of class Ci in a subset Qj. Qj contains those samples in Q that have value aj of A. The predictable information(J.R. Quinlan,1986):

$$E(A) = - \sum_{i=1}^m \frac{i(q)(q1i + q2i + \dots + qmi)}{q} \quad (5)$$

The training information gained by

$$Gain(A) = I(Q) - E(A) \quad (6)$$

$$SplitInfoA(Q) = - \sum_{i=1}^v \left(\frac{|Qi|}{|Q|} \right) \log_2 \left(\frac{|Qi|}{|Q|} \right) \quad (7)$$

The shown value of splitting data into dataset Q into v partitions consequent to v outcomes the test on attribute A (J.R. Quinlan, 1986). The gain ratio is

$$Gain\ Ratio(A) = Gain\ A / SplitInfoA(Q) \quad (8)$$

```

Input: D where D = pruned decision tree
Output: pruned decision tree
1: procedure prune tree
2:   for all nodes n in D do
3:     for all children c of n do
4:       if replacing n with c does not lower accuracy of D then
5:         replace n with c
6:         reclassify nodes of n
7:       end if
8:     end for
9:   end for
10 end procedure

```

Figure 3. Decision tree algorithm

3.3.4. Optimized Genetic Algorithm (OGA)

The authors propose a method to modify a general genetic algorithm to evaluate specified attributes on training data or a separate testing set and uses a decision tree (J.R. Quinlan, 1986) to estimate the 'merit' of a set of attributes to produce an optimized feature subset with genetic search elevation strategy to recognize the features. All feature selection technique should use an evaluation function together with a search strategy to achieve the optimal feature set (Huang & C., 2012). It is unable to be realized to search all subsets to find out an optimal subset and need much effort to indicate whether a particular feature is present or not in the chromosome, one, and zero used. One in a gene position refers to feature and zero to absent (Yanan Mao & Dingyuan Fan, 2016). The number of features and what are the features that are to be present in a chromosome are guided by information gain (IG) and gain ratio (GR). The initial population created using input values of IG and GR of the values present in the chromosome. After Generated the population, the individuals evaluated using a fitness function. There is no general approach to find the fitness function for a genetic algorithm. It is a heuristic approach and depends on the used application. So the authors nominate a C4.5 classifier to be used as a fitness function because C4.5 has some utility of handling both continuous and discrete attributes and training data with missing attribute values, pruning trees after creation - C4.5 goes back through the tree once it has been created and try to eject branches that do not help by replacing them with leaf nodes (Dash & H. Liu, 1997). (J.R. Quinlan, 1986). The following algorithm selects a

feature from the set of features that are gained by OGA, gain ratio, and Information gain, as shown in fig 3.

Figure.4. Proposed CSFS improved of a genetic algorithm with the c4.5 decision tree

```

Proposed feature selection using a Genetic Algorithm with a C4.5 decision tree
var1 sf1 /* set of features used as input to this algorithm*/
var2 sf2 /*contains selected features at the end of the algorithm*/
var3 p-size /*population size*/
var4 a /* an array is used to store the populations*/
1: begin
2: gen =0 /*number of generation*/
2: Design chromosome
3: while gen (count) <= gen
4: Randomly generate a p-size population and store it into the array a.
5: For i=1 to p-size
6: fitness function (a (i)) /* using C4.5 decision tree algorithm*/
7: Store the fitness value /* value of false-positive (fp) and false negative (fn)*/
8: end for
9: /* Select two chromosome say ch1 and ch2 which has minimum Fp and fn values*/
10: crossover (ch1, ch2) /*cr1, cr2 are outputs of crossover*/
11: randomly mutate one bit in cr1 /*mutation operation*/
12: randomly mutate one bit in cr2
13: gencount=gencount+1
14: end while
15: end

```

4. Proposed Stack Hybrid Classification Model Based On Composite Hybrid Feature Selection (CHFS)

A weka software tool (Weka online open-source accessed,2018) shows the list of black-box classifiers. These algorithms, in general, are used to classify the medical dataset.

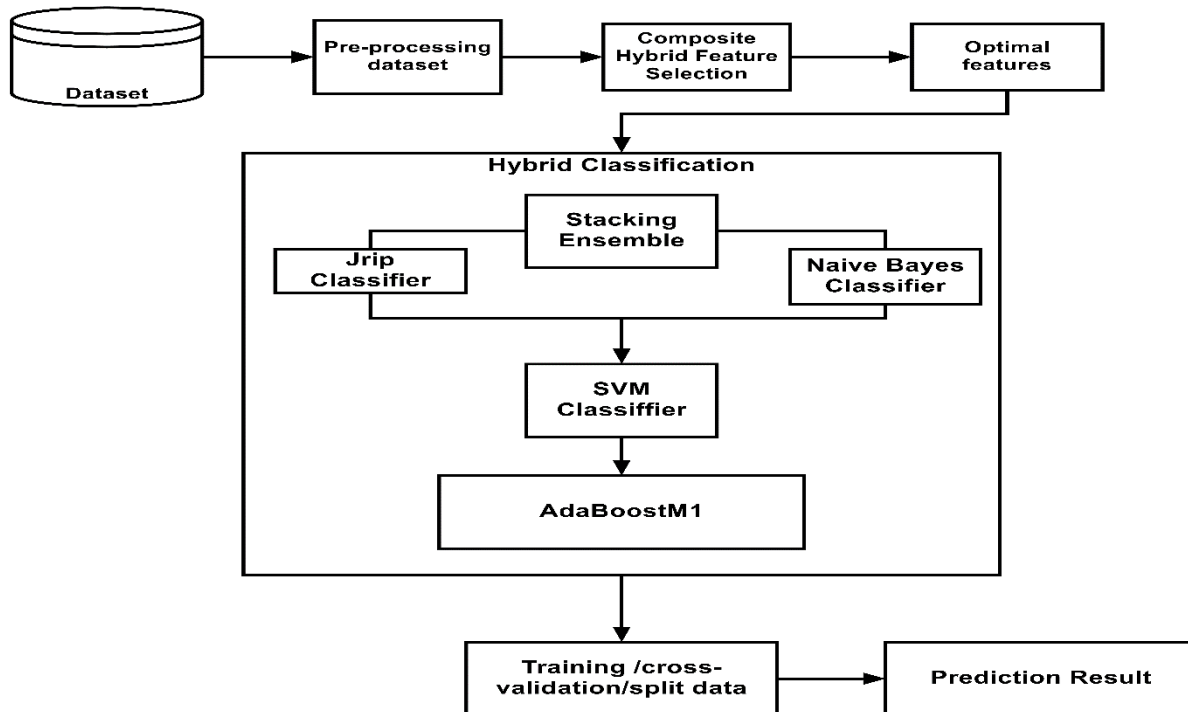
4.1. Two learning evaluators can be used to evaluate the dataset

- Training set: the classifier separates a dataset to test and training data. The result of each model can be saved and can visualize.
- Cross-validation: in case of 10 fold cross-validation, weka develops ten models, when it displays the result it uses the average performance of those ten models. It deletes the remaining models. From the observations, the authors conclude that the model saved with cross-validation and the training set is the same. (Divya Jain, Vijendra Singh,2018).

4.2. Stacking technique

Ensemble methods are learning methods that contain a set of classifiers for classifying data by taking a weighted point of their predictions (Leo Breiman,1996) . The authors combine multiple classifiers to get the maximum efficiency of classification accuracy and overcome the weakness of individual classifiers in the classification process on potential patients. Classifiers, as shown in fig 4.

Figure5. the proposed framework of hybrid classification based on the CHFS model



The author chooses permanent (Jrip, Naïve Bayes) classifiers on stack ensemble hybrid classification. Then combines Naïve Bayes with Jrip classifiers by using SVM classifier as Meta-classifier. Finally, move to the Adaboost algorithm for boosting a trivial classifier using the Adaboost M1 method. Just minimal class issues may tackle and all previous steps based on a result from the table2, which achieves a good indicator of a fitness function problem.

Boosting is a general ensemble method that uses several weak classifiers to create a robust classifier—after that, Done by building a model from the training data—then creating a second model to try to correct the first model's errors. Models introduced before the training set is correctly estimated or until a sufficient number of models added. AdaBoost was the first successful binary classification boost algorithm developed. It is the best starting point for improving awareness. Modern methods of boosting build on AdaBoost, most notably stochastic machines that boost gradient(Schapire R.E.,2013).

Predictions made using AdaBoost Predictions by calculating the weighted average of the sparse classifiers. Per weak learner measures, a predicted value as either + 1.0 or -1.0 for a new input instance. The predicted values weighted by the stage quality of each slow learner. The Ensemble model prediction takes as the sum of the weighted predictions. If the number positive, then the first class is predicted if the second class is estimated to be negative(Singer,1999).

Naive Bayesian classifier is an eclectic classifier and can calculate a group set of probabilities by counting the value and frequency in a given dataset (O. Elsayed, K. Mahar, M. Kholief, and H. A. Khater,2015). It assumes that all variables which contribute towards classification are mutually independent(Duarte Ferreira et al.,2012). Naive Bayesian classifier based on Baye's

notion and theorem of total probabilities. equation 9 is the probability of a document with a vector((S. R Bhagya,2014),(Tina R. Patil,2013))

$x = \{x_1, x_2, x_n\}$ belongs to hypotheses 'h' is given by,

$$P(h_1|dx) = \frac{P(x_i|h_1)P(h_1)}{P(x_i|h_1)P(h_1) + P(x_i|h_2)P(h_2)} \quad (9)$$

$$P(h_1|x_i) = \frac{P(x_i|h_1)P(h_1)}{P(x_1)} \quad (10)$$

$$P(h_1|x_i) = \sum_{j=1}^n P(x_i|h_j)P(h_j) \quad (11)$$

JRip Rules Classifiers (Divya Jain,2018): Classes are a measure of the rising size and generated a group of rules for the class to reduce the error gradually. JRip (RIPPER) continues by treating every one of the instances of an exacting decision in the training data as a class and discovery group of rules that included in the same class.

Support Vector Machines (SVM)(Divya Jain,2018): is a supervised learning algorithm based on statistical learning theory proposed by Vapnik

$$K(X_i, Y_i) = (\gamma X_i^T Y_i + r)^m, \gamma > 0 \quad (12)$$

In the interpretation of SVM, edge. The optimum hyperplane separation, therefore, maximizes the margin of the training data. Provided the training dataset identified with:

$(x_1, y_1), \dots, (x_n, y_n)$, $x_i \in \mathbb{R}^d$ and $Y_i \in \{-1, +1\}$, where x_i is a vector representation function and Y_i a training compound's classmark (negative or positive)

The optimal hyperplane can then defined as:

$$wx^T + b = 0 \quad (13)$$

Where w is the weight vector, x is the input feature vector, and b is the bias. The w and b would satisfy the following inequalities for all elements of the training set

$$wx_i^T + b \geq +1 \text{ if } y_i = 1 \quad (14)$$

$$wx_i^T + b \leq -1 \text{ if } y_i = -1 \quad (15)$$

The SVM algorithm originally proposed to construct a linear classifier by (Vapnik 1964). An alternative use for SVM is the kernel method, which enables us to model higher dimensional, non-linear models (Aizerman,1964).

And the Radial basis kernel function (RBF) (Huang,2012):

$$K(x_i, y_i) = \exp\left(-\gamma \|x_i - y_j\|^2\right), \gamma > 0 \quad (16)$$

4.3. Metrics used in health check systems for evaluation

The different performance metrics generally used to explore the performance of the various models like sensitivity, accuracy, precision, and f-measure (Divya Jain & Vijendra ,2017) . Accuracy: can be calculated by divide number of accurate predictions by the total number of all predications (Klöppe & ...&Frackowiak & R. S,2008)

$$Accuracy = \frac{TN+TP}{fP+tP+fN+tN} \quad (17)$$

Inside equation

(FN = False Negative, FP = False Positive, TN = True Negative, TP = True Positive)

5. Experimental Results and Evaluation

The CSV file of the breast cancer Dataset collected from Wisconsin UCI machine learning with a total of 569 rows and 32 columns loaded to the weka tool. All experiments evaluated by the receiver operating characteristic (ROC) curve, accuracy, F-measure (Dina A. Ragab & Omneya Attallah,2019). All experiments conducted on a laptop with an Intel Core i7- 8750H CPU at 2.2 GHz and 16 G of Ram running on windows 10 (64 bits).

5.1.Result from proposed (CHFS) feature selection model

The model implementation from the three feature selection methods,IG, GR, was applied for input to our optimization GA as initialization instead of randomly. Also, The population size is 100, with several generations is 20. The crossover takes place at the middle position, and mutation is prepared at one point randomly to reduce features of a dataset. The extracted feature subset result was consisting of 10 features terms, as shown in table 1.

Table 1: Number of feature extraction from three feature selection

| | Method | No.of Feature |
|--------------------------------|----------|---------------|
| IG, GR | Filter | 32 |
| GA | Wrapper | 26 |
| Optimized GA Fitness Fun(C4.5) | Embedded | 10 |

Source: (Weka open source software, 2018)

In table 2 below, we calculate the fitness function of j48 and calculate the accuracy of traditional classifiers. This procedure repeated with another classifier in a fitness function calculation to obtain the accuracy among them. The summarizing results shown that the j48 is performing a good indicator as a bigger fitness function value among all classifiers. Furthermore, measure false-positive and false-negative for each chromosome. The chromosome, which has the lowest value, is considered an elite one(Divya Jain, Vijendra Singh,2018).

Table 2 Calculate the fitness function of optimizing Genetic Algorithm

| | FS(J48) | FS(NB) | FS(JR) |
|-------------|---------|--------|--------|
| J48 | 73.06% | 69.41% | 72.50% |
| NB | 75.91% | 80.79% | 72.41% |
| JRIP | 72.90% | 70.13% | 74.85% |

The threshold value equal to 0.02 chosen for information gain lead to reduce feature term to 32 attributes, and a value greater than or equal to 0.04 selected for gain ratio method to rearrange the feature term as priority and preparation for classification method to obtain the maximum efficiency. Table 3 discusses the results of all the techniques. If 70% trained, reminder tests applied. For summarized results in the table (3) from the weka software of the analytics dataset, we used the (ROC) curve to evaluate the output result from our proposed architecture model, as shown in fig 12,13.

Table 3: Experimental result of Accuracy pre-post CHFS model with various classifiers

| Classifier | Pre- (CHFS) Feature Selection Accuracy | Post – (CHFS) Feature Selection Accuracy |
|--------------------|--|--|
| J48 | 61.98% | 95.32% |
| Naïve Bayes | 43.27% | 92.98% |
| JRIP | 61.09% | 97.07% |

Fig.5 Summarized accuracy results of pre-post (CHFS) feature selection

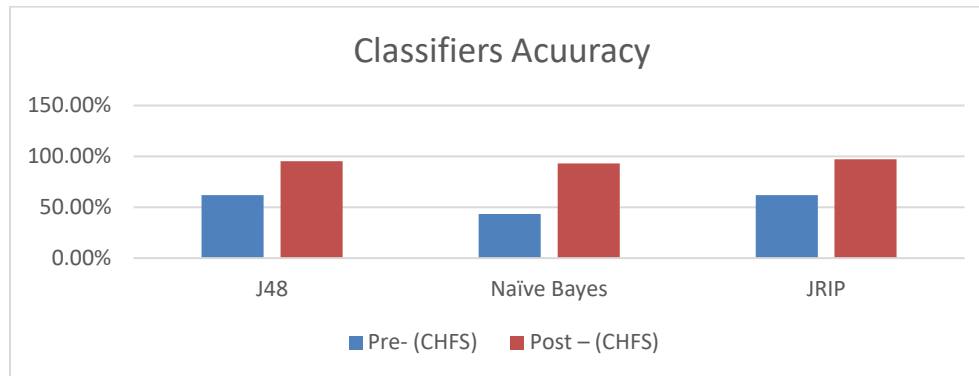
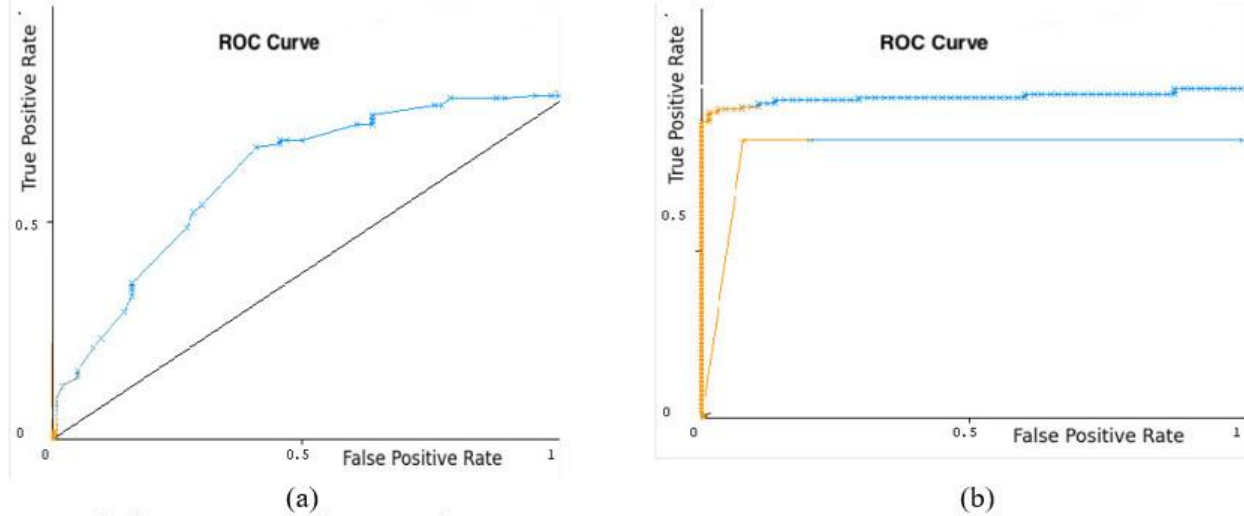


Fig.6 ROC curve for all J48-Naive Bayes-Jrip classifiers

(a) Pre-CHFS model

(b) Post-CHFS model



Source: (Weka open source software, 2018)

5.2 Comparison Between Proposed CHFS Model and PCA Feature Reduction

As an essential point, the proposed (CHFS) model compared against a popular dimensionality reduction technique, the Principal Component Analysis (PCA) (Joliffe IT,2002) . We have used PCA to reduce the number of variables , and the result was 13 features PCA component selected at 97.1% accuracy of variance from 32 attributes that download from UCI. The result of this comparison between PCA and our proposed CHFS feature selection model among three

Traditional classifiers, as shown in Table 4.Furthermore, summarized In fig. 8, Fig 9, explained the ROC curve to evaluate the result of this comparison.

Figure 7: The PCA component accuracy of variance result

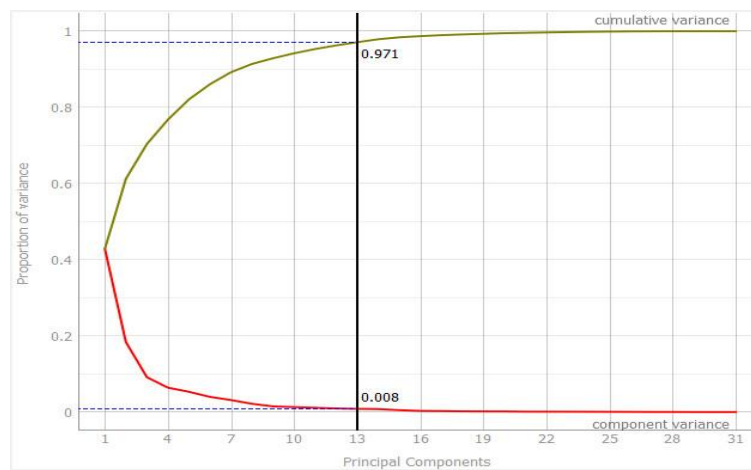


Table 4: Summary of (CHFS)Model Accuracy Compared Against PCA at 70% training, remainder test

| Classifier | PCA Accuracy | (CHFS) Model | |
|-------------|----------------------------|-------------------|----------|
| | (13) Components at (97.1%) | Feature Selection | Accuracy |
| J48 | 61.98% | | 95.32% |
| Naïve Bayes | 38.01% | | 92.98% |
| JRIP | 61.98% | | 97.07% |

Fig.8 Summarized accuracy results of PCA versus (CHFS) feature selection

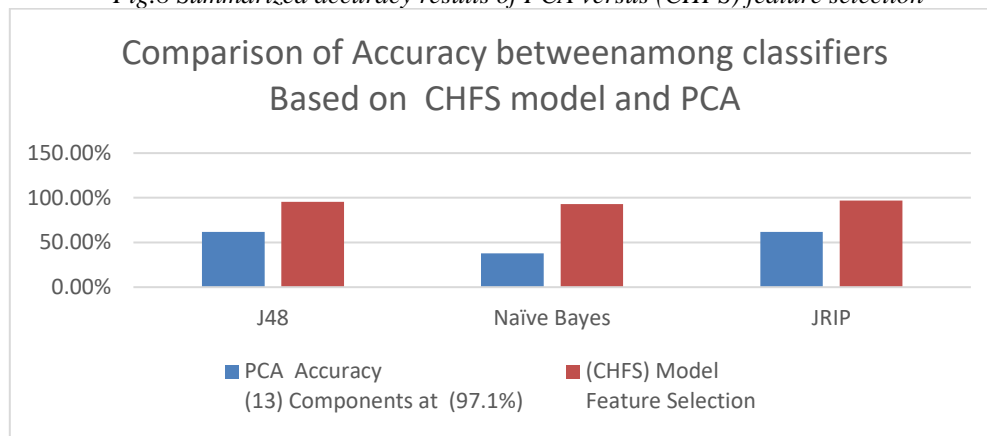
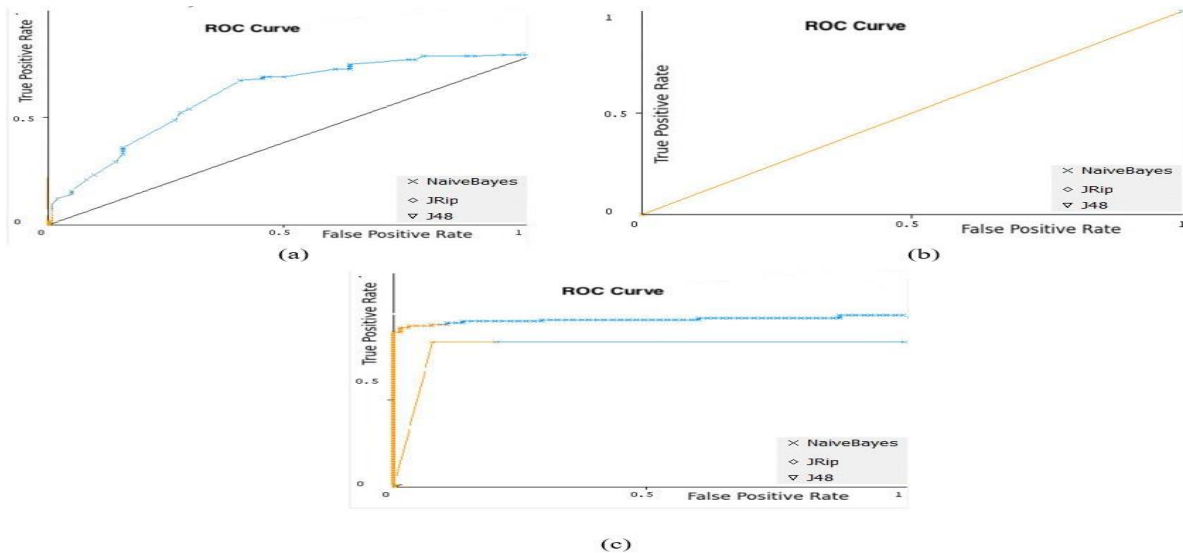


Fig.9 ROC curve for all Three classifiers on raw features

(a) Roc of traditional classifiers(naive Bayes-jrip-j48)

(b) ROC of 13- component –PCA

(c) Post-CHFS model



Source: (Weka open source software, 2018)

5.3 Result of proposed AdaBoost of Stack Hybrid Classification Model Based on (CHFS) Feature Selection

The classification techniques applied to medical dataset. (Yasser Fouad &...& Hatem Khater,2015) . Can be achieved by

$$\text{ClassificationAccuracy} = \frac{\text{CorrectlyClassifiedSamples}}{\text{TotalNumberofSamples}} * 100 \quad (18)$$

The authors use the output dataset from CHFS feature selection to the proposed classification model, which designed by flow knowledge in the weka tool and begins the combination process of Naive Bayes- Jrip classifiers with SVM classifiers as a meta-classifier in comparison with previous traditional classifiers.

The author implements a model of AdaBoost on the Stack Hybrid Classification process of (Naive Bayes Jrip) with SVM as a Meta classifier on full training mode and 70% split training, remainder test. The result showed predicted unseen samples infected with breast cancer disease. The proposed model (CHFS-BOGA-SVM) achieves high overall accuracy (98.25%) on 70% split training, remainder test mode. Also, achieve (100%) on full training at 50 iteration mode compared with the individual j48 classifier (61.98%) and naive Bayes of (43.27%), Jrip classification accuracy of(61.09%), as shown below in table 5 and ROC curve in figure 10.

Table 5: Summary of (CHFS)Model Accuracy Compared Against PCA at 70% training, remainder test

| Classifier | PRE-CHFS/AdaBoost- Classification Accuracy | Post- CHFS-BOGA- SVM | |
|-------------|---|------------------------------|---------------|
| | | 70% training, remainder test | Full Training |
| J48 | 61.98% | 98.25% | 100% |
| Naïve Bayes | 43.27% | | |
| JRIP | 61.09% | | |

Fig.10 ROC curve for AdaBoost-Stacked-CHFS-BOGA-SVM classification model and classification error curve

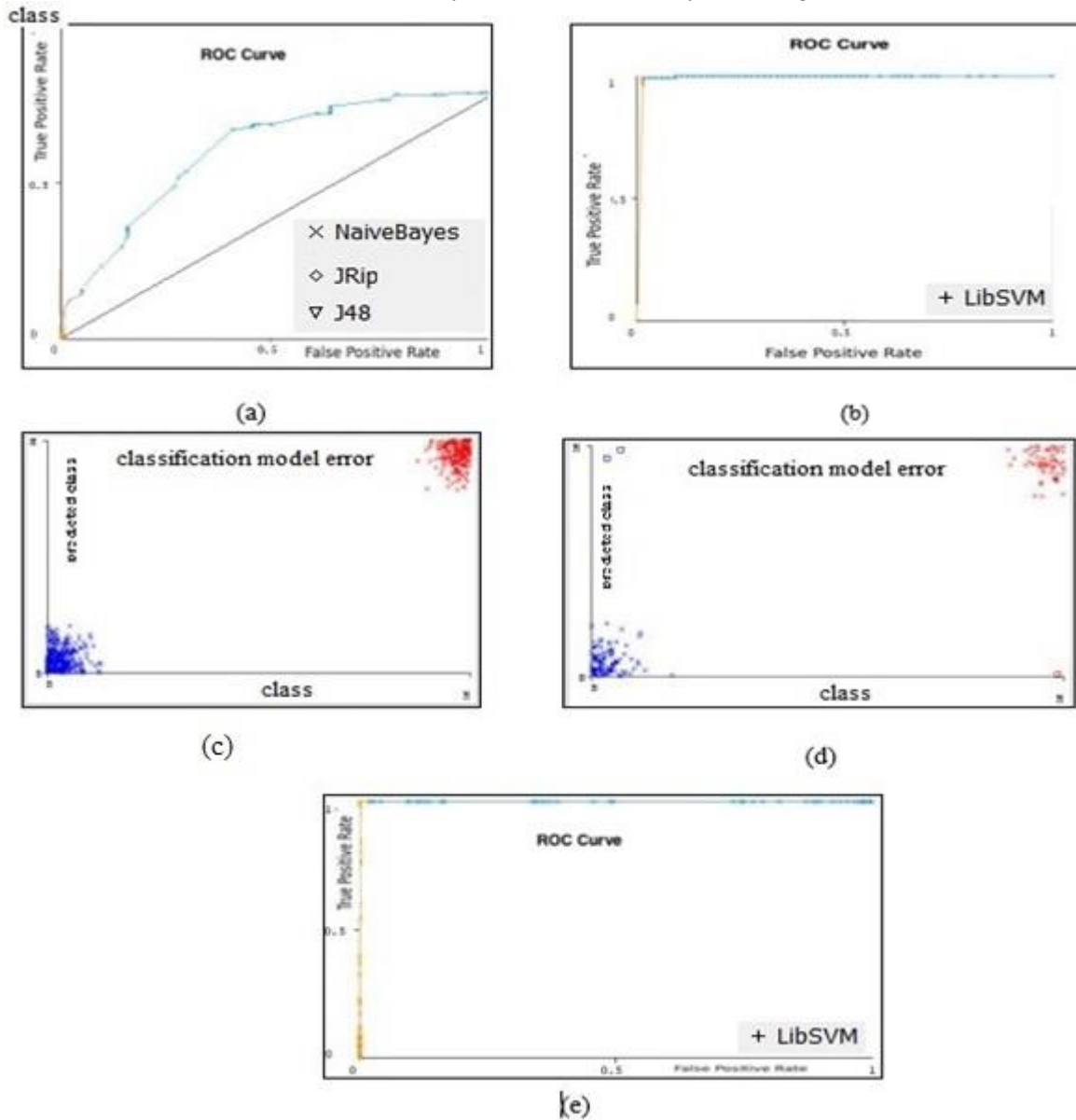
(a) ROC curve pre- AdaBoost-Stacked-CHFS-BOGA-SVM classification modeling.

(b) ROC curve post- AdaBoost-Stacked-CHFS-BOGA-SVM Hybrid classification modeling on 70% training.

(c) ROC curve post- AdaBoost-Stacked-CHFS-BOGA-SVM Hybrid classification modeling on full training.

(d) classification error curve-on 70% training

(e) classification error curve-on full training



Source: (Weka open source software, 2018)

Table 5. The detailed accuracy for AdaBoost -Stacked (Naïve Bayes, Jrip) with the SVM classifier model

| class | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | PRC Area |
|-----------|---------|---------|-----------|--------|-----------|----------|----------|
| Benign | 0.981 | 0.015 | 0.990 | 0.981 | 0.986 | 0.991 | 0.990 |
| Malignant | 0.985 | 0.019 | 0.970 | 0.985 | 0.977 | 0.992 | 0.981 |
| Avg. | 0.982 | 0.017 | 0.983 | 0.982 | 0.982 | 0.992 | 0.986 |

6. Discussion

The authors compare the results of different classification methods obtained with the results of our proposed stack hybrid classification based on (CHFS) feature selection on the breast cancer medical dataset of 569 potential patient samples. We used 400 (70 %) instances of total data for training in traditional classifiers as well as CHFS-BOGA- SVM. The remaining 169 (30 %) instances used to test both in CHFS-BOGA-SVM and J48-Naive Bayes-JRIP- SVM individually. After many attempts to classify the data with our proposed model, which obtains the best accuracy for this algorithm with this data, where the maximum classification accuracy is 98.25% among all classifiers. Besides, Reach to 100% accuracy in full training mode.

These comparisons, according to our proposed model presented in this study, were reduced a false negative rate and showed a relatively high overall accuracy with more accurate results, as shown in fig 14, 15, and table 6.

Table 6: Summary pre-post AdaBoost Stack Hybrid Classification Based on (CHFS) In Weka3.9

| Traditional Classifier | Accuracy of traditional pre-(CHFS-BOGA) | Accuracy of traditional post-(CHFS-BOGA) | AdaBoostM1 Stacked hybrid Classification Based on CHFS-BOGA | | Proposed Prediction model result | |
|------------------------|---|--|---|-----|----------------------------------|---------------|
| | | | | | 70% training, remainder test | Full Training |
| J48 | 61.98% | 95.32% | Naive bayes , JRIP | SVM | 98.25% | 100% |
| Naive Bayes | 43.27% | 92.98% | | | | |
| JRIP | 61.09% | 97.07% | | | | |
| SVM | 61.98% | 62.74% | | | | |

Fig.14 Summarized results of pre-post AdaBoost -stack classification based on (CHFS)

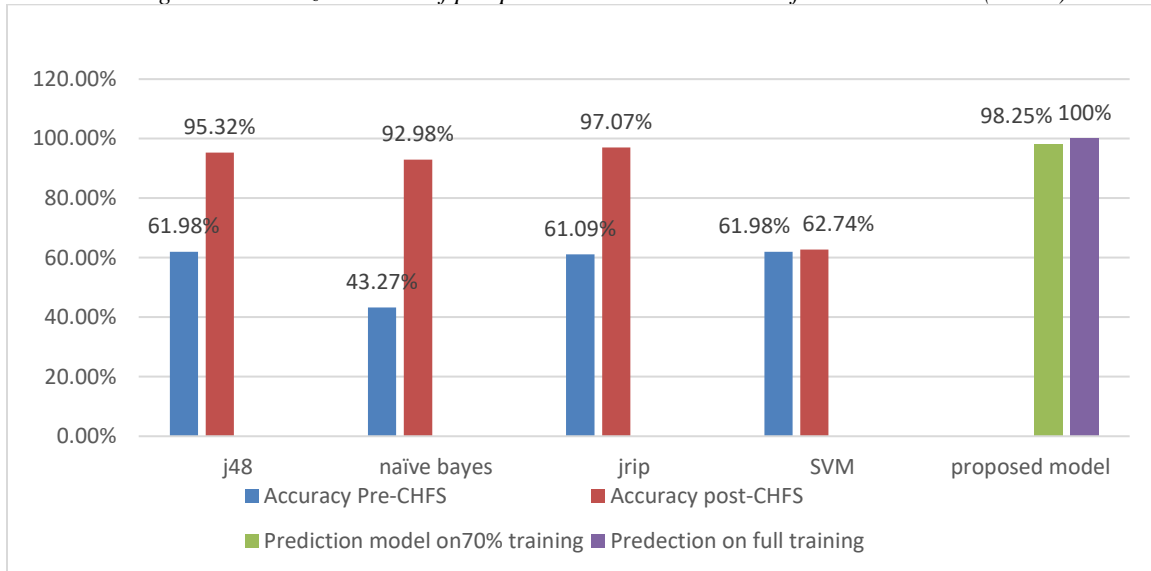


Table 6: Comparison of our proposed model with the last state of the art result

| Author | Contribution | Accuracy |
|--------------------------------------|---|-----------------------|
| Iranpoer,2007 | RBF-SVM | 94% |
| Sarvestan,2010 | Neural Network-NaiveBayes | 93.62% |
| Osman Hegazy,2015 | Optimization svm-with FLP-natural algorithm | High convergence rate |
| Power,2016 | Genetic Fuzzy FS. | 89.47% |
| Mohanty,2018 | Optimize FS,KNN on MIAS dataset | 97.86% |
| E.M.Badr,2019 | Gray wolf optimizer with SVM | 97% |
| Proposed Model CHFS+SHC (SVM) | Novel Hybrid F.S. with optimization of GA on AdaBoost Stacked-Hybrid Classification -with SVM model | 98.25% |

7. Conclusion

In this work, the author aims to early diagnosis of Breast cancer disease by using a benchmark dataset on our proposed composed hybrid feature selection (CHFS) model. Also, combine the advantages of three filter feature selection approaches and optimize Genetic Algorithm (OGA)

by improving the initial population generating and genetic operator's using the results of filter approaches as some prior information by using the J48 decision tree classifier as a fitness function instead of probability and random selection, This lead to speed up convergence and select the best features.

Besides, We are using the selected feature in the AdaBoost algorithm on a stack hybrid classification, which combines three classifiers with improving the prediction and accuracy. The proposed model performs better than the traditional classification approaches for optimum feature selection and improvement of the classification process and effectively reduced the false-negative rate with high accuracy when using a support vector machine (SVM) as meta-classifier in a hybrid classification method with 98.25% compared to 62.74% of usage individually and the last state of the art result shown above in Table 6. The results of the proposed CHFS-BOGA-AdaBoost-Stack Hybrid (naive Bayes-Jrip) with SVM model show an accurate classify breast cancer and classify the Tumor of Breast is (benign or malignant). Hence, it inferred confidently that the built intelligent system would act as a viable alternative decision support system on the breast tumor is cancer or not.

REFERENCES

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, "Global Cancer Statistics 2018," GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, in press.
- [2] Badr, E. M., Salam, M. A., & Ahmed, H. (2019). Optimizing Support Vector Machine using the Gray Wolf Optimizer Algorithm for Breast Cancer Detection. November.
- [3] Kalló, Gergő, Miklós Emri, et al. "Changes in the Chemical Barrier Composition of Tears in Alzheimer's Disease Reveal Potential Tear Diagnostic Biomarkers." *PLoS One*, vol. 11, no. 6, Public Library of Science, June 2016, p. e0158000.
- [4] Sarvestan Soltani A, Safavi A A, Parandeh M N, and Salehi M, "Predicting Breast Cancer Survivability using Data Mining Techniques," *Software Technology and Engineering (ICSTE)*, 2nd International Conference, Vol. 2, pp. 227-231, 2010.
- [5] Iranpour M, Almassi S and Analoui M, "Breast Cancer Detection from SVM and RBF Classifier," In 1st Joint Congress on Fuzzy and Intelligent Systems, 2007.
- [6] Veerabhadrapa, Lalitha Rangarajan, "Bi-level dimensionality reduction methods using feature Selection and feature extraction," *International Journal of Computer Applications (0975 – 8887)* vol. 4, No.2 July 2010.
- [7] Fu, J.C.; Lee, S.K.; Wong, Wang, A.H.; Wu, H.K.; introduce Image segmentation feature selection and pattern classification for microcalcifications. *Comput. Med. Imaging Graph.* 2005, 29, 419–429.
- [8] Pawar, M.M.; Talbar, S.N. Genetic Fuzzy System (GFS) based wavelet co-occurrence feature selection in mammogram classification for breast cancer diagnosis. *Perspect. Sci.* 2016, 8, 247–250.
- [9] Mohanty, F.; Rup, S.; Dash, B.; Majhi, B.; Swamy, M.N.S. Mammogram classification using contourlet features with forest optimization-based feature selection approach. *Multimed. Tools Appl.* 2018, 1–30.
- [10] Osman Hegazy, Omar S. Soliman, and Mustafa Abdul Salam, "Comparative Study between FPA, BA, MCS, ABC, and PSO Algorithms in Training and Optimizing of LS-SVM for Stock

Market Prediction,” International Journal of Advanced Computer Research vol.5, Issue. 18, pp.35-45, March-2015.

[11] Dua, D. and Graff, C. (2019). [<http://archive.ics.uci.edu/ml>]. UIC Machine Learning Repository.

[12] Weka open source software for data mining .availble online (<https://www.cs.wakito.ac.nz/>) accessed 26 Dec 2019.

[13] Chen, F. and F. Li, A combination of feature selection approach with SVM in credit scoring. Expert Systems with Applications, 2010. 37(7): p. 4902-4909.

[14] Chen, Y. -W., and Lin, C. -J., Combining SVMs with various feature selection Strategies, Available from <http://www.csie.ntu.edu.tw/~cjlin/papers/features.pdf>, 2005.

[15] Dash, M., and H. Liu, Feature Selection for Classification. Intelligent Data Analysis, 1997: p. 131C156.

[16] Devi, K.N., V.M. Bhaskaran and G.P. Kumar, Cuckoo ,Optimized SVM for Stock Market Prediction, IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded, and Communication Systems (ICJJECS)2015, 2015.

[17] Dai, W., Y.E. Shao, and C. Lu, Incorporating feature selection method into support vector regression for stock index forecasting. Neural Computing and Applications, 2013. 23(6): p. 1551-1561.

[18] Yanan Mao, Zuoquan Zhang, Dingyuan Fan, “Hybrid feature selection based on improved genetic algorithm for stock prediction,” 2016 6th International Conference on Digital Home,2016.

[19] Dina A. Ragab, Maha Sharks and Omneya Attallah,” Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers,” Diagnostics 2019, 9, 165; DOI:10.3390/diagnostics9040165.

[20] David P Salmon and Mark W. Bondi“Neuropsychological Assessment of Dementia" Access NIH public, PubMed central, US national library of medicine Institutes of Health, May 2010.

[21] Divya Jain, Vijendra Singh,” Feature selection and classification systems for chronic disease prediction: A review,” Egyptian Informatics Journal 19 (2018) 179–189,2018.

[22] R. Chaves, J. Ramírez, J.M., "Integrating discretization and association rule-based classification for Alzheimer’s disease diagnosis," science direct, Expert Systems with Applications, Volume 40, Issue 5, April 2013, Pages 1571-1578, <https://doi.org/10.1016/j.eswa.2012.09.003>.

[23] S. R. Bhagya Shree, Dr. H. S. Sheshadri, "An initial investigation in the diagnosis of Alzheimer’s disease using various classification techniques," 2014 IEEE International Conference on Computational Intelligence and Computing Research.

[24] Data mining: machine learning tools and techniques by Ian H Witten and Eibe Frank, published by Elsevier, second edition 2008.

[25]Aouatif Amine Ali Elakad Mohammed Rziza Driss “GA-SVM and Mutual Information based Frequency Feature Selection for Face Recognition” GSCM-LRIT, Faculty of Sciences, Mohammed V University, B.P. 1014 Rabat, Morocco(2011).

[26] J.R. Quinlan, Induction of Decision Trees, Machine Learning 1: pp.81-106, Kluwer Academic Publishers, Boston, (1986).

[27] Leo Breiman. "Stacked Regressions" Machine Learning, 24, 49-64 (1996)

[28] Duarte Ferreira et al. “Applying data mining techniques to improve diagnosis in neonatal jaundice" BMC Medical Informatics and Decision Making, 0-5, 2012

- [29] O. Elsayed, K. Mahar, M. Kholief and H. A. Khater, "Automatic detection of the pulmonary nodules from CT images," 2015 SAI Intelligent Systems Conference (IntelliSys), IEEE, London, pp.742-746, 2015, DOI: 10.1109/IntelliSys.2015.7361223.
- [30] S. R Bhagya Shree and Dr. H. S. Sheshadri "An initial investigation in the diagnosis of Alzheimer's disease using various classification techniques" IEEE ICCIC, 2014.
- [31] Tina R. Patil and Mrs. S. S. Sherekar "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification" International Journal of Computer Science and Applications Vol. 6, No.2, Apr 2013
- [32] S. R Bhagya Shree and Dr. H. S. Sheshadri "An initial investigation in the diagnosis of Alzheimer's disease using various classification techniques" IEEE ICCIC, 2014.
- [33] Tina R. Patil and Mrs. S. S. Sherekar "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification" International Journal of Computer Science and Applications Vol. 6, No.2, Apr 2013
- [34] Vapnik V., "Pattern recognition using generalized portrait method," Control, vol. 24, pp. 774–780, 1963.
- [35] Aizerman MA, Braverman EM, Rozner LI, "Theoretical foundations of the potential function method in pattern recognition learning," vol. 25, pp. 821–837, 1964.
- [36] Huang, C., A hybrid stock selection model using genetic algorithms and support vector regression. Applied Soft Computing, 2012. 12(2): p.807-818.
- [37] O. L. Mangasarian: "Cancer diagnosis via linear programming," SIAM News, Volume 23, Number 5, September 1990, pp 1 & 18.
- [38] William H. Wolberg and O.L. Mangasarian: "Multisurface method of pattern separation for medical diagnosis applied to," Proceedings of the National Academy of Sciences, U.S.A., Volume 87, December 1990, pp 9193-9196.
- [39] O. L. Mangasarian, R. Setiono, and W.H. Wolberg: "Pattern recognition via linear programming: Thomas F. Coleman and Yuying Li, editors, SIAM Publications, Philadelphia 1990, pp 22-30.
- [40] K. P. Bennett & O. L. Mangasarian: "Robust linear programming discrimination of two linearly inseparable sets," Optimization Software 1, 1992, 23-34 (Gordon & Breach Science Publishers).
- [41] Schapire R.E. (2013) Explaining AdaBoost. In: Schölkopf B., Luo Z., Vovk V. (eds) Empirical Inference. Springer, Berlin, Heidelberg.
- [42] Schapire, R.E., Singer, Y. Improved Boosting Algorithms Using Confidence-rated Predictions. Machine Learning 37, 297–336 (1999). <https://doi.org/10.1023/A:1007614523901>.