PERFORMANCE EVALUATION OF SOFT MACHINE LEARNING ALGORITHMS USING FEATURE SELECTION TECHNIQUES FOR DISEASE PREDICTION

[1]Anil Kewat, [2]P N Srivastava, [3]Dharamdas Kumhar
[1]Research Scholar, [2] Professors, [3]Assistant Professor
[1]Department of Computer Science & Engineering,
[1]Jagannath University, Jaipur, India

Abstract: Cancer, Diabetes, and Heart disease are the most common diseases which are the challenges and the major problems for the medical society of this real world and prediction of these diseases in the early stages are very important as it is one of the important causes of sudden death. Although specialists use some basic treatment methods and different types of pathology test for such type of disease prediction computer-aided diagnosis systems that use intelligent machine learning methods are used to help the health care professionals in the diagnosis of these diseases among the patients. For this research work, seven machine learning algorithms with the seven feature selection techniques were used and evaluated on the Heart disease, Breast Cancer and Diabetes datasets. The dataset used in this paper is collected from the UCI machine learning repository. We performed the simulation work on these datasets and analyze the results. After analyzing it is concluded that the support vector machine and J-48 machine learning algorithm gives the better results compare to the other machine learning algorithms for disease prediction.

***IndexTerms* - Support vector machine, UCI, Feature Selection, Machine Learning, Dataset**

I. INTRODUCTION

According to the WHO, heart disease is one of the top leading causes of death per year. An estimated 17.5 million deaths occur due to cardiovascular diseases worldwide. Over 80% of deaths in the world are because of Heart disease [1]. Cardiovascular disease includes coronary artery diseases like angina and myocardial infarction narrows the coronary arteries. There is another heart disease, called coronary heart disease, in which a waxy substance called plaque develops inside the coronary arteries. A heart attack is a common cause of death worldwide. The number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014. The global prevalence of diabetes among adults over 18 years of age has risen from 4.7% in 1980 to 8.5% in 2014. Diabetes prevalence has been rising more rapidly in middle- and low-income countries than in high-income countries [2]. Extracting knowledge and patterns for the diagnosis and treatment of disease from the Medical database becomes more important to promote the development of telemedicine and community medicine. Since the desire to achieve better results in health care services has increased the importance of computer-aided systems [3], they have begun to be used in addition to clinical methods. Therefore, data such as patient information, medical diagnostics, and medical images were started to be recorded [4]. Later, machine learning methods processing these data were used to build decision support systems. Some examples of these methods are as follows. The methods generally used in the applications of medical data include artificial neural network, genetic algorithms, fuzzy system, Naive Bayes classification, Decision tree methods such as J-48 algorithm and C4.5 algorithm, Support vector machine, K-Means Clustering Random Forest, NB-Tree, Linear Regression etc. A genetic algorithm is one of the important methods in medical data mining for feature subset selection. Machine learning works on a simple rule – if you put garbage in, you will only get garbage to come out. By garbage here, I mean noise in data. This becomes even more important when the number of features is very large. You need not use every feature at your disposal for creating an algorithm. You can assist your algorithm by feeding in only those features that are really important. Basic reasons to use feature selection are:

- It enables the machine learning algorithm to train faster.
- It reduces the complexity of a model and makes it easier to interpret.
- It improves the accuracy of a model if the right subset is chosen.
- It reduces over fitting.

II. RELATED WORK

Much work is being done in finding effective ways to medical diagnose of various diseases. Our job is to try to provide an effective diagnosis with a limited number of factors that contribute most to heart disease through classification. Sellapan et al. [5] developed an intelligent heart disease prediction system to predict heart disease using three classification decisions from Tree, Naïve Bayes and Neural Network. Asha et al. [6] developed a system for predicting heart disease using decision trees, simple Bayesian and Artificial Neural Networks. They used all 13 characteristics in the forecast. They concluded that Naïve Bayes was doing well with 96.6% chances. Tu et al. [7] used the decision tree from the Cleveland Heart Disease dataset and showed a correct response rate of 78.9%. Anbarasi et al. [8] the performance improvement of decision trees was investigated by integrating genetic algorithm as a method of selecting a subset of diagnostic properties in patients with heart disease. Pudil et al [9] proposed the Sequential Forward Floating Selection (SFFS) algorithm was one of the commonly used algorithms. The main advantage of this method is that it produces a hierarchy of feature subsets with the best selection for each dimension. Heon Gyu Lee et al [10] proposed a novel technique to develop the multi-parametric feature with linear and nonlinear characteristics of Heart Rate Variability. Chen, J et al. [11] used statistical and classification techniques for the development of multi-parametric features of heart rate variation. In addition, they assessed the linear and nonlinear properties of the heart rate variation. They performed extensive tests on the linear and nonlinear properties of the heart rate variation index to evaluate different classification systems, such as Bayesian classifiers. Niti Guru et al. [12] suggested a method for predicting heart disease, blood pressure and sugar using a neural network. Experiments were performed on an example of a patient database with 13 input variables for testing and training. They recommend that you diagnose heart disease using the supervised network. The training was performed in the neural network using an algorithm for backward propagation. Robert Detrano et al. [13] Use logistic regression generated based on defined goals to accurately classify heart diseases and obtain experimental results with an accuracy of about 77%. Colombet et al [14] evaluated the use of CART and artificial neural networks to predict individual heart disease. Rajeswari et al. [15] recommended finding useful information from the Indian Heart Disease Dataset. Therefore, the proposed decision support system helps to determine the risk test for predicting a patient's heart disease in the years to come. The proposed system has 19 functions that have been simplified to the main functions using genetic algorithms. It is expected that data mining will facilitate the development of future events and will help in the selection of treatments that identify individual risk groups that may determine angioplasty or surgery. Kora et al. [16] studied the signal that ECG predicts a heart attack. They use improved BAT algorithms to reduce the number of electrocardiogram signal functions. Then the selected attributes (the best of the 200 attributes) are inserted into the classification unit, SVM, KNN, NN and scalar conjugated gradient neural network (SCGNN). With the aid of LM NN, the best experimental results were obtained with an accuracy of 98.90%. Deepika et al. [17] used association rule mining to classify the type of heart attack. In this study, they used different characteristics, such as the number of chest pains and the age of the patient as patient characteristics. Harleen et al. [18] investigated the availability of data mining methods such as decision trees, induction rule, and artificial neural networks for the diagnosis of diabetic patients.Vijayarani, S., & Dhayanand, S. et al [19] proposed SVM and Naive Bayes algorithms to predict liver disease. Uma et al. [20] used different classification algorithms (SVM, Bagging, Naive Bayes, Regression, J48, etc.) and methods for selecting functions (CfsSubsetEval, Gain of information, Gain Ratio and Wrapper method method, etc.). There are 689 set of HD datasets. Singh et al. [21] suggested a hybrid approach using genetic function selection and naive Bayesian methods. Takci et al. [22] find the best machine learning method and the method of selecting the heart attack prediction function, the SVM combining the linear kernel and the Relief-Based Feature which makes it possible to obtain the best predictive performance of the heart diseases. Jabbar et al. [23] a predictive model of heart disease was suggested by a randomly selected forest and a feature subset selection. The authors propose a new approach to disease prediction using the RF method and features subset selection chi-square method. Chi-square metric was used to filter the features of the dataset. The classification model was analyzed using five measurements of sensitivity, specificity, disease prevalence, negative predictive value and positive predictive value. The experimental results show that the accuracy of their method has improved considerably. Peter J et al [24] has proposed the heart dataset contains large volumes of which consumes more time for classification so by using attribute selection methods the dimensionality of data is reduced. Naïve Bayes classification technique produced enhanced results. This is observed when the performance of four classification algorithms: Naïve Bayes, Decision Tree, K-NN and Neural Network are investigated on complete heart disease dataset and reduced dataset. Mai Shouman et al [25] various techniques have been proposed to combine the extraction of single and unique hybrid data to diagnose heart disease. Recent studies have shown that hybridization of two or more techniques improves diagnosis. In this paper, various machine learning techniques such as MLP, Naive Bayes, decision tree, and nuclear density were applied to different sets of data on heart disease and the accuracy of each technique was measured. The use of hybrid data mining techniques for various sets of data on heart disease shows a different accuracy.

III. FEATURE SELECTION

Feature selection is a process in which a small portion of the features of the original set of features are converted without transformation (to preserve the interpretation) and this is verified for analysis purposes. The selection process can be implemented in different ways, depending on the purpose, the resources available and the desired level of optimization. The feature selection method helps you create accurate forecasting models in your work. They help you choose better features or accuracy and select features that require less data. The complexity of the model is reduced, simple models are easier to understand and explain, and so fewer attributes are desirable. The purposes of the variable selection are to improve predictor performance, provide faster and more convenient predictors, and better understand the underlying data generation process [26].

3.1. General Procedure of Feature Selection

The feature selection process consists of four main points. Subset generation, subset evaluation, stopping criteria, result verification as shown in fig. 1. Fig. 1 illustrates the development of a general methodology to feature selection. The feature space has been simplified into a subset of the features based on a standard criterion. Finally, check these features by verifying measures. The process for selecting general feature is described in the next section. The method for Feature selection can be categorized in different ways. Features selection algorithms prior to the introduction of the concept are divided into two categories. The wrapper method encloses feature selections around induction algorithms for cross-validation and provides the benefits of adding or removing features to the feature subset used. The filtering method is a general preprocessing algorithm that does not depend on the knowledge of the algorithm used. Both methods have strong arguments. The most common are subdivided into filters, wrappers, embedded methods and hybrid methods [27].

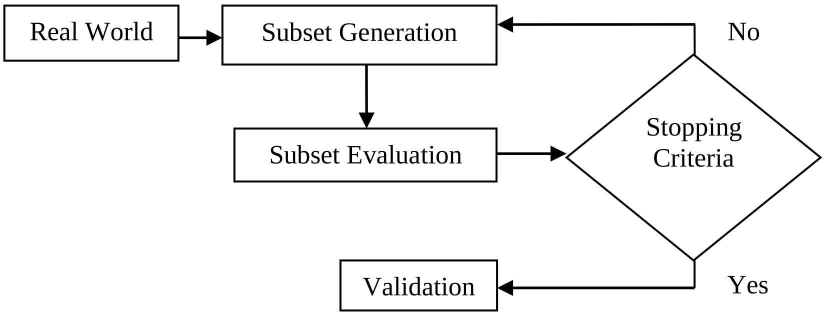

Fig.1 General Feature Selection Procedure

3.2. Filter Methods

Regardless of the algorithm used for data modeling, the filtering method selects a feature based on performance metrics. Model algorithms can only be used after finding the best features. The filtering method can adjust a single function or evaluate a complete subset of functions. The measures developed for feature filtration units can be widely distributed in information, distance, consistency, similarity and statistical measurements. All filtering functions are not available for all category extraction activities. As a result, the filters are classified according to the tasks (classification, regression or clustering). The Feature selection technology must be distinguished from the extraction of the Feature. The Feature extraction retrieves a new Feature from the original Feature and the selection of the Feature provides a subset of this Feature. Feature selection techniques are commonly used in areas with many Feature and relatively few samples [28].

3.3. Wrapper Methods

The wrapper considers subsets of feature based on the performance quality of the modeling algorithm which considered a black box evaluator. For task activities, the wrapper evaluates the subgroups based on classification performance, but the cluster, a wrapper evaluates the subset based on the execution of the clustering algorithm [29]. The evaluation is repeated for each subset and the subset is generated to depend on the search strategy in the same way as the filter. If enough subsets are detected, the Wrappers will be much slower than the filter because it depends on the resource requirements of the modeling algorithm. These feature subsets are also oriented towards modeling algorithms that evaluate them.

3.4. Embedded and Hybrid Methods

The embedding method performs function selection during the execution of the modeling algorithm. Therefore, these methods are embedded into the algorithm as normal functions or extensions. Common embedded methods include various types of decision tree algorithms such as CART, J-48, random forest, and other algorithms such as multiple logistic regression and its variants [30]. A hybrid approach is recommended to combine the forces of filters and wrapper. First, it is possible to obtain several candidate subsets using the filtering method to reduce the space available in the feature space. Then use the wrapper to find the best subset of candidates. The hybrid method generally achieves high accuracy, characteristic of the wrappers and the high efficiency of the filter. The hybrid method can be built using almost any combination of filters and wrappers, but several interesting methods have recently been proposed, such as feature selection based on random fuzzy forest, hybrid genetic algorithm, etc. [31]. The main objective of this research is to contribute to the prediction of medical research of various diseases and to compare general methods of feature selection on the basis of filters in a detailed and comprehensive manner. Filter-based methods do not depend on classifiers and are generally faster and more scalable than feature-based methods. In addition, they have lower mathematical complexity. In this paper, we use the following seven filtering options to select features. Details of these methods can be found in the following section.

A. Information Gain

In probability theory and information theory, information gain (relative entropy or Kullback-Leibler divergence) is a measure of the difference between two probability distributions. Evaluate the X feature by measuring the amount of information gained in relation to the class Y (or group) variable defined as follows.

$$I(X) = H (P(Y)-H (P(Y/X))$$

In particular, it is assumed that the marginal distribution (P (Y)) of the observable Y is assumed to be independent of feature X, and it is assumed that the conditional distribution Y (P (Y / X)) which is dependent on X measured. If there is no differential expression in X, Y becomes independent of X, X has a small-value information gain and vice versa.

B. Relief

Relief-F is an instance-based feature selection method that evaluates feature based on the extent to which values distinguish different sets of samples, but are similar. For each feature, X Relief - F selects random samples and the k nearest neighbors of the same class and different classes. Subsequently, X is evaluated as the sum of the weighted differences between different classes and the same class. If X is expressed differently, it will show a larger difference with samples of different categories, so that it will get a higher score (and vice versa).

C. Symmetric Uncertainty (SU)

Symmetric uncertainty (SU) is one of the best ways of feature selection, and feature selection system based on mutual information use this metric. SU is a measure of correlation between features and classes.

SU= (H (X) + H (Y) – H (X \Y)) / (H (X) + H (Y))

Where H (X) and H (Y) are the common entropies based on the probability associated with each feature and class value, and the joint probability of all combinations of the X value and the Y value is H (X, Y). The symmetric uncertainty value of the high value feature assumes a high value. Due to the correction factor of 2, SU takes the normalized value in the interval [0, 1]. The value of SU = 1 means that the knowledge of the feature is completely predicted, the SU = 0 means that X and Yare are irrelevant. [32].

D. One-R

One-R is a simple algorithm that is suggested by Holte [33]. Create a rule for each feature of the training data, and then select the rule with the least error. They use a simple way to treat all numeric features as continuous functions and divide the values into different intervals. Treat missing values by treating missing values as legal values. This is one of the most primitive ways. It creates simple rules based on a single feature. This is the smallest form of classifier, but it helps to determine basic performance as the basis for other learning scenarios.

E. Gain Ratio

The gain ratio (GR) is a change in information metrics that reduces distortion. When selecting an attribute, the percentage of gain takes into account the number and size of the branches. Change the information gain by taking into account the intrinsic information about the partition. Intrinsic information is the entropy of the distribution of copies in a branch (that is, the amount of information to be specified for the branch to which the agent belongs). As the intrinsic information increases, the value of the attribute decreases [34].

(Gain ratio Attribute) = Gain (Attribute) / intrinsic _ info (Attribute)

F. Principal Component Analysis (PCA)

The goal of the PCA is to reduce the size of a dataset containing a number of associated attributes by converting the original attribute space to a new non-associated attribute space. The algorithm then adjusts the changes between the original dataset and the new dataset. As soon as you save the transform properties that contain most of the variations, we remove the remaining properties. Just as importantly, the PCA does not consider class tags, so it is suitable for unsupervised datasets [35].

G. Chi-Square

Chi square is a statistical test used to determine if there is a significant relationship between two categorical variables. In short, chi-square statistics check whether there is a significant difference between the observation frequency and the prediction frequency of the two variables. The chi-squared statistics are calculated as follows.

$$\chi = \sum \frac{(Observed - Expected)^2}{Expected}$$

IV. PROPOSED METHODOLOGY

4.1. Model Diagram

The proposed methodology is summarized in fig.2 below in the form of a model diagram. The figure displays the flow of the research conducted in constructing the model.

4.2. Dataset Used

The data is normally distributed. For this research work, WEKA tool [36] is used for performing the experiment. It includes a collection of various machine learning methods for data classification, clustering, regression, visualization etc. One of the biggest advantages of using this tool is that it can be personalized according to the requirements. The objective of this study is the prediction of the patients affected by different diseases such as heart attack prediction, breast cancer prediction and diabetes prediction. Following table-1.1 shows a brief description of the dataset.

TABLE 1.1: DATASET

Disease Datasets	No. of Instances	Number of Attributes	No. of Classes
Wisconsin Breast Cancer	699	10	2
Statlog (Heart) Data	270	14	2
PIMA Diabetes	768	9	2

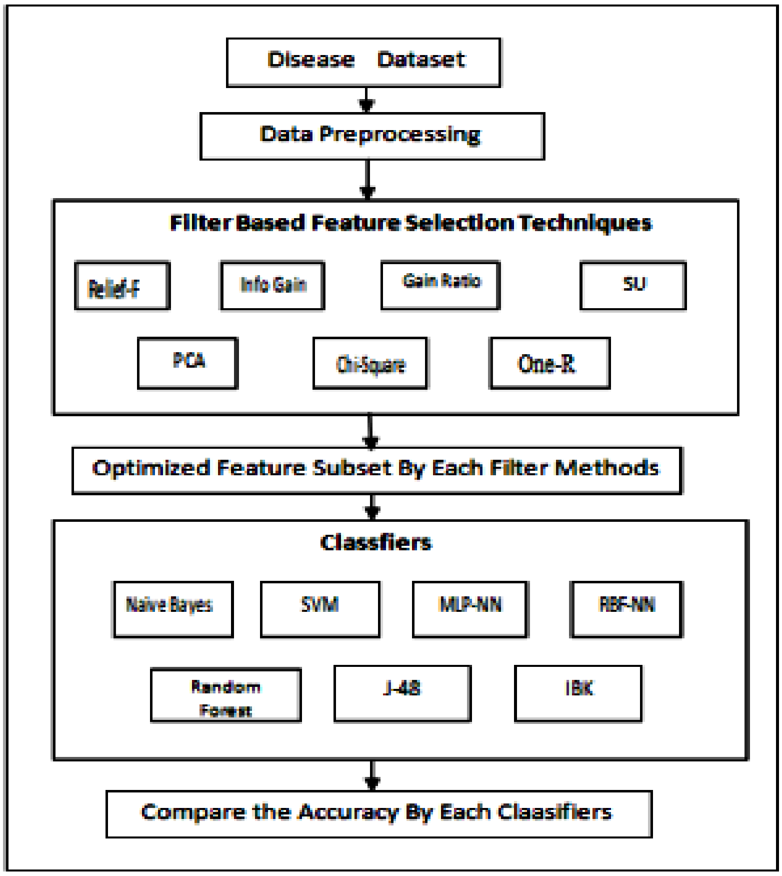

Fig. 2 Proposed Feature Based Classification Procedure

V. EXPERIMENT AND RESULTS

In these experiments we have used seven filter feature selection techniques namely : Relief-F, Information Gain (IG), Symmetric Uncertainty (SU), Gain Ratio (GR), One-R, Chi-Square and Principle Component Analysis (PCA) and for this experiments we have used seven classification algorithms namely : Naïve Bayes, J-48, Random Forest, Radial Basis Function Neural Network (RBF-NN), Multi Layer Perceptron Neural Network (MLP-NN), Support Vector Machine (SVM), and Instance Based Learning algorithm (IBK). For analyzing the performance of classifiers with the feature selection techniques we have used three datasets namely Heart Disease, Wisconsin Breast Cancer and PIMA Diabetes datasets. Following table 1.2 contain the results of classification accuracy for heart disease prediction for the above-mentioned classifiers and feature selection techniques. The following table 1.2, table 1.3 and table 1.4 contain the results of classification accuracy of various classifiers for heart disease prediction, breast cancer and diabetes prediction subsequently. Table 1.2 contains the results of different classifiers for heart disease dataset without feature selection and with feature selection techniques. For heart disease prediction we have used 8 features out of 14 feature set. The result shows that when we applied feature selection techniques on the heart disease dataset the results are much better to compare to the without feature selection approach. The highest classification accuracy 85.5556 is given by Naïve Bayes classifiers with the Symmetric Uncertainty and Gain Ratio feature selection techniques. Although J-48, Radial Basis Function Neural Network (RBF-NN) and SVM classifiers also give good results comparable to the Naïve Bayes classifier.

TABLE 1.2: CLASSIFICATION ACCURACY FOR HEART DISEASE DATASET

Classifier	Accuracy Without Feature-Selection	Relief-F	Info. Gain(IG)	Symmetric Uncertainty (SU)	Gain Ratio (GR)	One-R	Chi-Square	PCA
NAIVE BAYES	76.3021	84.0741	85.1852	85.1852	**85.5556**	**85.5556**	85.1852	84.8148
J-48	73.8281	82.2222	80.7407	80.7407	80.3704	83.3704	80.7407	79.6296
RANDOMFOREST	75.7813	79.6296	80.3704	80.7407	80.7407	79.6296	80.0	81.4815
RBF-NN	75.3906	83.3333	**85.1852.**	**85.1852**	**85.1852**	**85.1852**	**85.1852**	80.3704
MLP –NN	75.3906	82.9630	80.7407	81.1111	79.2593	78.8889	79.6296	80.0
SVM	77.3438	83.7037	**84.8148**	**84.8148**	**84.8148**	**84.8148**	**84.8148**	**84.8148**
IBK	70.1823	75.5556	77.4074	77.4074	77.4074	77.4074	77.4074	76.2963

Where RBF-NN is a Radial Basis Function Neural Network, MLP-NN is Multi Layer Perceptron Neural Network, SVM is a Support Vector Machine and IBK is Instance Based Learning algorithm. The following table-1.3 contains the results of different classifiers along with the various feature selection techniques for the Wisconsin Breast Cancer Dataset.

TABLE 1.3: CLASSIFICATION ACCURACY FOR WISCONSIN BREAST CANCER DATASET

Classifier	Accuracy Without Feature-Selection	Relief-F	Info. Gain(IG)	Symmetric Uncertainty (SU)	Gain Ratio (GR)	One-R	Chi-Square	PCA
NAIVE BAYES	95.9943	96.4235	96.1373	96.1373	96.1373	95.8512	95.1852	94.5637
J-48	**94.5637**	**94.7067**	**94.7067**	**94.8498**	**94.7067**	**95.4220**	**95..4220**	**96.8526**
RANDOMFOREST	96.1373	96.7096	95.8512	95.8512	96.1373	96.1373	96.1373	**97.4249**
RBF-NN	95.8512	96.1373	95.7082	95.7082	95.7082	95.2790	95.2790	94.8498
MLP –NN	**95.2790**	**96.1373**	**95.7082**	**95.4220**	**95.7082**	**95.9943**	**95.8512**	**96.2804**
SVM	96.9957	**96.9957**	96.1373	96.1373	96.1373	95.4220	95.4220	95.9943
IBK	95.1359	95.5651	**94.5637**	**94.5637**	**94.5637**	**94.8498**	**94.8498**	95.4220

For Wisconsin Breast Cancer Dataset J-48 and MLP-NN classifiers performed very well for all the seven feature selection techniques. The highest classification accuracy 97.4249 is given by the Random Forest classifier for the PCA Feature selection technique. The second highest accuracy 96.9957 is given by the SVM classifier. The Performance of Instance Based Learning algorithm IBK is poor compared to other above-mentioned classifiers. The following table-1.4 contains the results of different classifiers along with the various feature selection techniques for the For PIMA Diabetes Dataset.

TABLE 1.4: CLASSIFICATION ACCURACY FOR PIMA INDIAN DIABETES DATASET

Classifier	Accuracy Without Feature-Selection	Relief-F	Info. Gain(IG)	Symmetric Uncertainty (SU)	Gain Ratio (GR)	One-R	Chi-Square	PCA
NAIVE BAYES	76.3021	76.1719	74.8698	76.8698	76.4323	74.8698	74.8698	76.6927
J-48	**73.8281**	**74.7396**	**74.2188**	**74.2188**	**74.3490**	**74.2188**	**74.2188**	71.3542
RANDOM FOREST	75.7813	75.0	73.5677	73.4375	74.6094	73.3073	72.9167	73.5677
RBF-NN	**75.3906**	**75.9115**	**75.3906**	**75.3906**	**76.1719**	**75.3906**	**75.3906**	73.8281
MLP –NN	75.3906	76.8291	74.3490	74.6094	76.0417	74.6094	73.8281	75.6510
SVM	77.3438	77.0833	75.9115	75.9115	76.6927	75.9115	75.9115	**77.3438**
IBK	**70.1823**	**67.5781**	**68.099**	**68.0990**	**67.0573**	**68.099**	**68.0990**	**66.7969**

For PIMA Diabetes Dataset J-48 and RBF-NN classifiers performed very well for all the feature selection techniques except the PCA. The highest classification accuracy 77.3438 is given by the Support Vector Machine (SVM) classifier for the PCA feature selection technique. The Performance of Instance Based Learning algorithm IBK is poor for all the seven feature selection techniques compare to the other classifiers for Diabetes dataset.

VI. CONCLUSION

Diabetes, Heart disease and Breast Cancers are the most common disease which is spreading rapidly in society and the leading cause of death all over the world. Researchers have been investigating applying different machine intelligence techniques to help health care professionals in the diagnosis of these diseases among the patients. From the results, it has been analyzing that for all the three datasets J-48 and SVM classifier gives the best results compare to the other above mentioned machine learning algorithms. Besides this multi layered perceptron neural network (MLP-NN) classifier also give good results except for the diabetes prediction. This paper gives a comparative study for some of the machine learning algorithms for Disease Prediction. This prediction model helps the doctor’s inefficient disease diagnosis process with fewer features. It is observed that the classification accuracy of J-48 decision tree and support vector machine classifier increases better after dimensionality reduction.

REFERENCES

[1]. www.who.int/cardiovascular_diseases/en/

[2]. NCD Risk Factor Collaboration. (2016). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4• 4 million participants. The Lancet, 387(10027), 1513-1530.

[3]. Elmaghraby, A. S., Kantardzic, M. M., & Wachowiak, M. P. (2006). Data mining from multimedia patient records. In Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques (pp. 551-595). Springer, Boston, MA.

[4]. Takci, H. (2018). Improvement of heart attack prediction by the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences, 26(1), 1-10.

[5]. Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on (pp. 108-115). IEEE.

[6]. Rajkumar, A., & Reena, G. S. (2010). Diagnosis of heart disease using data mining algorithm. Global journal of computer science and technology, 10(10), 38-43.

[7]. Tu, M. C., Shin, D., & Shin, D. (2009, October). Effective diagnosis of heart disease through bagging approach. In Biomedical Engineering and Informatics, 2009. BMEI'09. 2nd International Conference on (pp. 1-4). IEEE.

[8]. Anbarasi, M., Anupriya, E., & Iyengar, N. C. S. N. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. International Journal of Engineering Science and Technology, 2(10), 5370-5376.

[9]. Nalavade, J., Gavali, M., Gohil, N., & Jamale, S. (2014). Impelling Heart Attack Prediction System using Data Mining and Artificial Neural Network. International Journal of Current Engineering and Technology, 4(3), 1-5.

[10]. Lee, H. G., Noh, K. Y., & Ryu, K. H. (2007, May). Mining bio signal data: coronary artery disease diagnosis using linear and nonlinear features of HRV. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 218-228). Springer, Berlin, Heidelberg.

[11]. Cheng, J., & Greiner, R. (1999, July). Comparing Bayesian network classifiers. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence (pp. 101-108). Morgan Kaufmann Publishers Inc.

[12]. Guru, N., Dahiya, A., & Rajpal, N. (2007). Decision support system for heart disease diagnosis using neural network. Delhi Business Review, 8(1), 99-101.

[13]. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J. J., Sandhu, S., & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64(5), 304-310.

[14]. Colombet, I., Ruelland, A., Chatellier, G., Gueyffier, F., Degoulet, P., & Jaulent, M. C. (2000). Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. In Proceedings of the AMIA Symposium (p. 156). American Medical Informatics Association.

[15]. Rajeswari, K., Vaithiyanathan, V., & Amirtharaj, P. (2011). Predictions of risk score for heart disease in India using machine Intelligence. In International Conference on Information and Network Technology (Vol. 4).

[16]. Kora, P., & Kalva, S. R. (2015). Improved Bat algorithm for the detection of myocardial infarction. SpringerPlus, 4(1), 666.

[17]. Deepika, N., & Chandrashekar, K. (2011). Association rule for classification of Heart Attack Patients. International Journal of Advanced Engineering Science and Technologies, 11(2), 253-57.

[18]. Kaur, H., & Wasan, S. K. (2006). Empirical study on applications of data mining techniques in healthcare. Journal of Computer science, 2(2), 194-200.

[19]. Vijayarani, S., & Dhayanand, S. (2015). Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), 816-820.

[20]. Uma, K., & Hanumathappa, M. (2016). Heart Disease Prediction Using Classification Techniques with Feature Selection Method. Adarsh Journal of Information Technology, 5(2), 22-29.

[21]. Singh, N., Ferozepur, P., & Jindal, S. (2018). Heart disease prediction using classification and feature selection techniques.

[22]. Takci, H. (2018). Improvement of heart attack prediction by the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences, 26(1), 1-10.

[23]. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2016). Prediction of heart disease using random forest and feature subset selection. In Innovations in Bio-Inspired Computing and Applications (pp. 187-196). Springer, Cham.

[24]. Peter, T. J., & Somasundaram, K. (2012, March). An empirical study on prediction of heart disease using classification data mining techniques. In Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on (pp. 514-518). IEEE.

[25]. Shouman, M., Turner, T., & Stocker, R. (2012, March). Using data mining techniques in heart disease diagnosis and treatment. In Electronics, Communications and Computers (JEC-ECC), 2012 Japan-Egypt Conference on (pp. 173-177). IEEE.

[26]. Korn, F., Pagel, B. U., & Faloutsos, C. (2001). On the" dimensionality curse" and the" self-similarity blessing". IEEE Transactions on Knowledge and Data Engineering, 13(1), 96-111.

[27]. Hoque, N., Bhattacharyya, D. K., & Kalita, J. K. (2014). MIFS-ND: A mutual information-based feature selection method. Expert Systems with Applications, 41(14), 6371-6385.

[28]. Liu, H., & Motoda, H. (Eds.). (1998). Feature extraction, construction and selection: A data mining perspective (Vol. 453). Springer Science & Business Media.

[29]. Kim, Y., Street, W. N., & Menczer, F. (2002). Evolutionary model selection in unsupervised learning. Intelligent data analysis, 6(6), 531-556.

[30]. Cawley, G. C., Talbot, N. L., & Girolami, M. (2007). Sparse multinomial logistic regression via bayesian l1 regularisation. In Advances in neural information processing systems (pp. 209-216).

[31]. Oh, I. S., Lee, J. S., & Moon, B. R. (2004). Hybrid genetic algorithms for feature selection. IEEE Transactions on Pattern Analysis & Machine Intelligence, (11), 1424-1437.

[32]. Hall, M. A., & Smith, L. A. (1998). Practical feature subset selection for machine learning. In Computer science'98 proceedings of the 21st Australasian computer science conference ACSC (Vol. 98, pp. 181-191).

[33]. Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. Machine learning, 11(1), 63-90.

[34]. Han, J., & Kamber, M. (2001). Data Mining: Concepts and Techniques, chapter Mining association rules in large databases.

[35]. Ashraf, M., Chetty, G., & Tran, D. (2013). Feature selection techniques on thyroid, hepatitis, and breast cancer datasets. International Journal on Data Mining and Intelligent Information Technology Applications, 3(1), pp.1.-6

[36]. Garner, S. R. (1995, April). Weka: The waikato environment for knowledge analysis. In Proceedings of the New Zealand computer science research students conference (pp. 57-64).