Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis

Sunila Godara* and Rishipal Singh

Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Techonology, Hisar - 12500, Haryana, India; sunilagodara@gmail.com, sandeep1642@gmail.com

Abstract

Background/Objectives: Medical science industry has immense measure of information; however a large portion of this information is not mined .Machine Learning takes analytics to the extreme by exploring hidden information in data. Disease diagnosis is major intention of medical decision support system which will assist the physicians to obtain valuable decision. **Methods/Statistical Analysis:** This research work Machine Learning techniques: K-Nearest Neighbors, Decision Tree, Artificial neural networks, Radial Basis Function neural networks and Support Vector Machine are analyzed. **Findings:** Performance of these techniques is compared through various performance measures such as sensitivity, specificity, accuracy, F measure, and Kappa statistics, True Positive Rate, False Positive Rate and ROC on Breast Cancer Wisconsin, Liver Disorder, Hepatitis and cardiovascular Cleveland Heart disease datasets. Research work consists of 10V fold cross validation method to evaluate the fair estimate of prediction techniques. **Application/Improvements:** Evaluation of these techniques on diverse medical datasets gave an insight into predictive ability of Machine Learning in medical diagnosis and there is a wide space of improvement.

Keywords: Artificial Neural Networks, Decision Tree, K-Nearest Neighbors, Medical Diagnosis, Performance Measures, Radial Basis Function Neural Networks and Support Vector Machine

1. Introduction

*Author for correspondence

The past belief that no one has made-up an effective doctoring algorithm and no one will for a extremely long time; is not true now a day although it is very complicated to grant a machine the intuition required to know how to keep on. Intuition is more essential for medical diagnosis as compared to lab data or some hard evidence. Now, it has been claimed that decision support systems will begin to replace physician in common tasks in the future¹. Medical science industry has vast measure of information, however unfortunately the greater part of this information is not mined to find out concealed data in information. Information mining strategies can be utilized to find concealed example in information. Medicinal information digging has incredible potential for investigating the shrouded examples in the informa-

tion sets of the restorative area. These samples can be utilized for clinical examination. In any case, the available unprocessed data is extensively passed on, heterogeneous in nature, and voluminous. The data removed from the broad databases is obliged to be as exact as could sensibly be normal. In any case, on account of multifaceted design and enormous sizes of databases, it is troublesome or even hard to find 100% accurate learning².

Cognitive restrictions prevent doctor's remembering 10000+ diseases human can get. Technology compensates for human deficiencies. Technology is better at organizing and recalling complex medical information than a hotshot Harvard MD. So a good Medical Decision Support System (MDSS) can be considered as a better tool for improving the efficiency of physicians. It is very imperative for clinician as well as patients to know the future holds of a disease for arranging the better treatment. The significance of making a accurate medical diagnosis cannot be over pushed. There are passionate and lawful results if a patient is told ill when, they are not. The patient bear keen distress; Physician may be legitimately responsible of this distress .Really the more major results is declaring up patient as disease free ,while they are not.

Proper cure is withheld because of misdiagnosis. Patient will pass on due to this misjudgment. Machine Learning Technology can enhance the capacity of correct diagnosis. The acknowledgement of utilization of Machine Learning Technologies in medical diagnosis framework is expanding rapidly. This is basically on the grounds that the adequacy of these ways to deal with order and expectation frameworks has enhanced, especially in connection to helping medical specialists in their decision making^{1, 3}. Tradition of current century is "first do no harm" means if techniques hurts ten patents a year and saves a thousand lives we must reject it. So research in field of medical diagnosis using Machine Learning is rapidly increasing for improving the decision making capabilities with greater degree of confidence by lowering the degree of uncertainty.

2. Previous Work

Paper⁴ proposed Adaptive Neuro-Fuzzy Inference System (ANFIS) with Kalman filter for fetal ECG exclusion from abdominal areas .The abdominal ECG is complex and Kalman filter calculated the maternal part from abdominal ECG. The maternal part in the abdominal ECG contained nonlinear transformed description of maternal ECG. Adaptive neuro-fuzzy inference systems have been used to distinguish these nonlinear complex relationships. Validated results proposed effectiveness of this research work.

Paper⁵ presented an interactive image segmentation algorithm. Algorithm was designed to remove undesirable bias and noise from medical images. The research work concealed the robustness and effectiveness in the delineation in cardiac MRI images.

Paper⁶ proposed Medical Decision Support System based on Improved Multilayer perceptron algorithm which divides datasets into multiple subsets. Then MLP algorithm was called independently for each subset and results obtained from different subsets were combined using voted combiner with majority probability rule. Performance was measured through sensitivity, specificity, accuracy and ROC. Research work was done using WEKA tool and concluded that Improved MLP approach significantly outperforms MLP approach.

Paper⁷ proposed Extreme learning machine as growing technology which overcomes various disputes faced by other machine learning techniques. ELM works for single-hidden layer feed forward networks (SLFNs). Non tuned hidden layer is strength of ELM. Evaluated with other conventional computational intelligence techniques, ELM gave superior generalization performance at faster learning speed.

Paper⁸ presented an intelligent heart valve diseases diagnosis system based on Principle Component Analysis (PCA) and Adaptive Neuro-Fuzzy Inference System (ANFIS).

Paper⁹ proposed SAS based methodology used for heart disease diagnosis. A neural networks ensemble method was proposed by combining the posterior probabilities. Experiments were carried out using SAS on Cleveland heart disease dataset, attained 80.95% sensitivity, 95.91% specificity and 89.01% classification accuracy for heart disease diagnosis.

Paper¹⁰ proposed a new model R-C4.5 based on C4.5 and enhanced the competence of attribute selection and partitioning models. Paper¹¹ presented KNN and Artificial Neural Network to categorize bloggers of Kohkiloye and Boyer Ahmad province in Iran. The simulation results have proposed improvements over earlier methods with better bloggers classification.

Paper¹⁸ implemented Image Fusion Algorithms for Clinical Diagnosis.

Paper¹⁹ proposed Image Fusion process by combining common features of a set of images to get an output image with superior content in terms of subjective as well as objective analysis. Review on some of the Image Fusion Techniques was done. A comparison of all techniques was also done by fusing MRI and CT images and analysis was done by using quality measures such as entropy, MSE, PSNR, SSIM and BRISQUE. CT modalities provide information on denser tissues whereas MRI provides information on soft tissues. For medical applications CT and MRI images are fused for diagnostic purposes.

In paper²⁰, clustering of images was done using integration of hyperbolic tangent kernels and Gaussian kernels. The proposed algorithm was tested on MRI image data base. Paper²² demonstrates creation of expert system for diabetic mellitus diagnosis using clustering and classification techniques of data mining. Paper²³ introduces competent MRI brain image investigation method, classified as normal or non cancerous brain tumour and cancerous brain tumour. Steps used by proposed method: 1. Pre-processing, 2. Segmentation, 3. Textural and shape feature extraction and 4. Classification. In this proposed MRI image analysis segmentation was done using Active Contour Method (ACM) and Artificial Neural Network (ANN) based on Levenberg-Marquardt (LM) algorithm was used for classification process. Paper²⁴ employed a fuzzy logic approach to propose a model to diagnose some cases of Anemia. Paper²⁵ proposed Tubercle Bacilli Diagnosis System with TSK-type Neuro Fuzzy Controllers by combining different partial solutions.

3. Prediction Models

3.1 K-Nearest Neighbors (KNN)

K-Nearest neighbor is supervised machine learning algorithm used for classification of objects based on closest training examples. An object is classified by a majority of its neighbors i.e K having always positive value. The K nearest neighbors' algorithm steps are as given^{12,13}:

- Value of K is decided which is used to specify number of nearest neighbors.
- Distance between the input instance and all the training samples is calculated using different distance measures.
- Distances for all the training samples are sorted and minimum distance based on nearest neighbor is determined.
- Training data categories for the sorted values are located.
- The output is calculated by using the majority of nearest neighbors.

3.2 Decision Tree

Decision trees procedure recursively isolates perceptions in branches to develop a tree with the end goal of enhancing the expectation precision. They utilize numerical calculations to differentiate a variable and relating boundary for the variable that divides the information perception into two or more subparts. This step is rehashed at every node until the complete tree is developed. The target is to locate a variable-edge match that

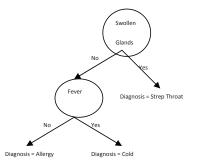


Figure 1. Decision Tree.

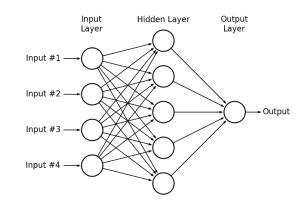


Figure 2. MLP.

expands the homogeneity of the subsequent two or more subparts of tests. Figure 1¹⁰ demonstrates an illustration of choice tree on patient determination. Here non-terminal hubs speak to tests on one or more traits and terminal hubs reproduce choice results. Choice tree sums up taking subsequent information: If a an enduring has swollen organs, the finding is strep throat and enduring does not have swollen organs and has fever, the finding is cold. In the happening if enduring have no swollen organs and have no fever, the finding is allergy.

3.3 Artificial Neural Network

ANNs are diagnostic cognitive system having popular architecture called Multi-Layer Perceptron (MLP) with back-propagation algorithm. Figure 2^{13, 14} shows MLP feed forward Neural Network. Input layer takes input and data is forwarded to the next hidden layer. Hidden layer accept the data from the Input layer along the path they are connected and process the data. Number of hidden layers and a number of neurons in each layer many vary

for solving the particular task. After processing at hidden layer data is fed forward to the output layer. Then output is matched to the target. If the match is not found than the weight is adjusted at each processing element and process is repeated until there is a match or error reduces to the preferred limit. For each input variable there is one neuron in the input layer. N categories are represented by N-1 neurons.

Input Layer: Predictor variable is given to the input layer. This layer standardizes these values in range of-1 to 1 and transfers these values to the hidden layer. Bias input of 1.0 is given to each hidden layer; the bias is then multiplied by weight and added to the sum supplied to the neuron².

Hidden Layer: At this layer value from each input neuron is multiplied by a weight, these weighted values are added and weighted sum is obtained which is further given to a transfer function known as σ .

Output Layer: The value obtained from each hidden layer neuron is multiplied by a weight, these weighted values are added collectively and new value is generated. The back-propagation algorithm adjusts the network weights and biasing values. Gradient decent method is used to decrease the square sum of the error E between the known output (Y) and predicted output (Y').

$$E = \frac{1}{2} N\Sigma (Y - Y')^2$$

Where N is the number of data points and E is square sum of the error.

3.4 RBF Neural Network

A Radial Basis Function (RBF) neural network contains only one hidden layer and neurons in the hidden layer may vary depending upon a particular task. At this layer Gaussian transfer functions having outputs inversely proportional to the distance from the center of the neurons¹⁷ are used. The predicted target value of an item is same as other items, close to the predictor variables. Positions one or more neurons can be positioned by RBF network in the space described by the predictor variables. A dimension of this space is same as the number of predictor variables. The Euclidean distance is calculated from the point being evaluated to the center of each neuron, and a Radial Basis Kernel Function (RBF), is applied to this distance and weight for each neuron is calculated. Radius is used as one of the parameter so it is called Radial Basis Kernel Function (RBF)^{15,16}. Figure 3² shows architecture of RBF neural network. The radial-basis functions technique recommends designing of interpolation functions F of the subsequent form¹⁷:

$$F(x) = \sum_{i=1, Nw} \phi(||x - x_i||)$$

where $\varphi(\|\mathbf{x} - \mathbf{x}_i\|)$ is a set of nonlinear radial-basis functions known as kernels²⁰, \mathbf{x}_i are the centers of these kernels, and $\|.\|$ is the Euclidean norm.

3.5 Support Vector Machine

SVM can be used for classification of both linear and nonlinear data and known as maximum margin classification algorithms. Training data is transformed into a higher dimension space by non-linear mapping. Classification is done by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVM employs kernel function as training method and multi-layer perceptron employs the weights adjustment¹⁶. Predictor variable is known as an attribute, and transformed attribute used to describe the hyperplane is called a feature in SVM. Most suitable feature depiction process is known as feature selection. A set of features that

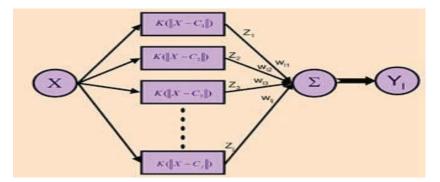


Figure 3. RBF Architecture.

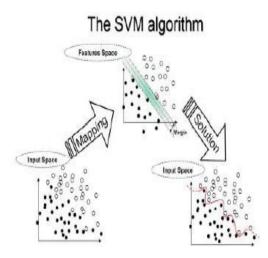


Figure 4. SVM topology.

explains one case is called a vector². Therefore SVM modeling provides optimal hyperplane that separates clusters of vector with one category of the target variable on one side of the plane and remaning are on other side of the plane. The vectors close to the hyperplane are known as support vectors. The Figure 4¹⁶ shows the SVM process.

4. Data Source and Tool Used

4.1 Data Source

Breast Cancer Wisconsin data set: This is downloaded from archive.ics.uci.edu/ml/datasets/ and has 11 attributes and 699 instances.

Liver Disorder data set: This is downloaded from archive.ics.uci.edu/ml/datasets/ and contains 7 attributes and 345 instances.

Hepatitis data set: This is downloaded from archive. ics.uci.edu/ml/datasets/ and contains 20 attributes and 155 instances.

Cleveland Heart data set: This is downloaded from archive.ics.uci.edu/ml/datasets/ and contains 14 attributes and 303 instances. Above four data sets are unrelated and will signify a fair test of diagnostic accuracy for various machine learning techniques. Characteristics of data sets differ with number of records, number of variables, percentage of categorical variables and balance. The percent of categorical variables measure relative mix between real and ordinal variables. Balance measures the ratio of records in largest classification group to records in smallest group. As balance value increases data set becomes

	No of records	No of variables	Balance
Brest cancer	699	11	3.4
Liver	345	7	1.4
Heart	303	14	2.7
Hepatitis	155	20	2.1

highly unbalanced. Properties of each data set in this research are given in Table 1.

4.2 Tool Used

For this research work Machine Learning models are developed using Weka version 3.6. WEKA is known as landmark in the narration of machine learning research. SIGKDD, 2005 award for outstanding work in KDD and machine learning was given to WEKA by ACM. It is freely available for download. WEKA presents many influential features and now it has befallen one of the most extensively used machine learning tool. The WEKA has been downloaded 300,000 epochs since it was placed on Source Forge in April 2000, and at present speed of download is 10,000/month. The WEKA mailing list has 1400 subscribers in its credits in more than 50 countries. More than 15 extensive projects extending WEKA are in market till now. The WEKA GUI Chooser consists of four buttons and these buttons present the following applications:

- Explorer: It gives an environment for exploring data.
- Experimenter: It gives an environment for performing experiments based on various Machine learning techniques.
- Knowledge Flow: Facility of drag-and-drop and incremental learning is provided.
- Simple CLI: Facility of execution of WEKA commands for operating systems is provided.

WEKA supports execution various Machine Learning techniques in Java. Releasing WEKA as open source software and as its implementation in Java has huge participation in its success.

5. Predictive Model Performance Measures

Evaluation of Machine Learning Techniques can be compared according to number of measures. They are not only used as the criteria to evaluate learning techniques, but also used as the heuristics to make learning models. But it is not easy to state which measure is most suitable to evaluate the performance. If we consider two measure accuracy and ROC than ROC is more reliable than accuracy, which indicates that ROC should be favored over accuracy in evaluating learning techniques. We also compare Precision, Recall, F Measure and Kappa coefficient to give a preference order in comparing performance of these techniques.

5.1 Sensitivity, Specificity and Accuracy

Prediction results of a classifier can be represented by Confusion Matrix. Confusion matrix was obtained to calculate sensitivity, specificity and accuracy. Table 2 demonstrates confusion matrix.

The upper left cell confirms the number of samples classified as true while they were true i.e TP, and the lower right cell confirms the number of samples classified as false while they were actually false i.e TN. The other two cells of confusion matrix confirm misclassified samples. FN confirming the number of samples classified as false as they actually were true, and the lower left cell and FP confirming the number of samples classified as true as they actually were false.

- Sensitivity: The part of sample with the disease and correctly identified as true.
- Specificity: The part of sample without the disease and correctly identified as false.

Underneath formulae are used to compute sensitivity, specificity and accuracy^{2.}

- Sensitivity = TP / (TP + FN)
- Specificity = TN / (TN + FP)
- Accuracy = (TP + TN) / (TP + FP + TN + FN)

Table 2.	Confusion	Matrix
----------	-----------	--------

	Classified as True	Classified as not True	
Actually True	ТР	FN	TP+FN=L2
Actually not True	FP	TN	FP+TN=L1
	TP+FP=C2	FN+TN=C1	

5.2 Precision, Recall and F Measure

Precision is the number of True Positives divided by the sum of True Positives and False Positives .Recall is the number of True Positives divided by sum of True Positives and the number of False Negatives. Precision and recall are calculated² as:

• Precision= TP / (TP + FP)

• Recall= TP / (TP + FN)

F Measure is used to combine precision and recall and known as Harmonic mean of precision and recall.

F=2(Precision* Recall)/(Precision+ Recall)

5.3 Kappa Coefficient

The Kappa coefficient is used to measure the level of agreement or disagreement of a group of sample. Kappa coefficient k tells the proportion of observed agreement Po between the actual class and the predicted class. Pa is portion of agreement expected by chance. Below equations demonstrate calculation of Kappa coefficient². For C1,C2, 11,l2 calculation confusion matrix given in Table 2 is used.

$$k = \frac{P_o - P_a}{1 - P_2}$$
$$P_o = \frac{TN + TP}{N}$$

$$P_a = ((C1 * L1) + (C2 * L2)) / N$$

Kappa coefficient k defines closeness of the actual and predicted values. Values of k close to zero identifies that agreement between the two classes is very low and values of k close to 1 describes the agreement is very high between the two classes.

5.4 Receiver Operating Characteristic (ROC)

A Receiver Operating Characteristic is defined as a space by False Positive Rate and True Positive Rate at several cut-off points.

- True Positive Rate = TP / (TP + FN)
- False Positive Rate = FP / (FP + TN)

Let us consider two parts of a dataset. One part with a disease, the other part without the disease, there is no

perfect separation between two groups. Subsequently, the distribution of the test results will overlap, as depicted in the Figure 5a. ROC curve having area closer to one defines clear cutoff about this separation²¹ Figure 5a: ROC curve Analysis. Consider sensitivity, specificity, True Positive Rate and False Positive Rate at three cutpoints given in Figure 5b. Its ROC curve is depicted in Figure 5b²¹. TPR is depicted on X axis. FPR(1- specificity) is on Y axis.

6. Experimental Setup and Results

We will analyze the performance of Machine Learning techniques; K-Nearest Neighbors, Decision Tree, Artificial neural networks, Radial Basis Function neural networks and Support Vector Machine on Breast Cancer Wisconsin, Liver Disorder, Hepatitis and Cardiovascular Cleveland Heart disease datasets using WEKA tool. Several parameters are required to define for these learning techniques. For KNN 3, 5 and 7 is good range of Neighbors to examine. For J48 confidence factor is .25 and seed is taken as

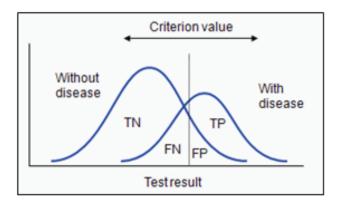


Figure 5a. ROC curve Analysis.

	Sensitivit	Specificit	True	False
Cutpoint	у	у	Positives	Positives
5	0.56	0.99	0.56	0.01
7	0.78	0.81	0.78	0.19
9	0.91	0.42	0.91	0.58

one. Again for MLP number of hidden layers and number of neurons in each hidden layer are automatically adjusted at optimum level by WEKA .Learning rate is set as .3 which is a control parameter used to control step size when weights are iteratively adjusted .Momentum is again a control parameter tells how much amount of weight have to change is set as .2. For RBF is trained using K Means clustering which is unsupervised machine learning algorithm and number of cluster is set as 2 and 3. For SVM seed is taken as 1 and RBF kernel function is used.

10V fold Cross Validation is used to validate results. 10Vfold cross validation consists of arbitrarily dividing the existing data into 10 subparts and subsequently training 10 classifiers using all data but one subpart which is always taken different for all 10 classifiers is used for testing the performance of the classifiers. The approximation error of the classifier built from the entire data is the average error of all the subparts. Now we will perform analysis of various Machine Learning Techniques on various Medical Datasets.

6.1 Breast Cancer Dataset

Table 3 given below shows confusion matrix for five classification techniques on Breast Cancer Wisconsin dataset. TP and FP both are high as compared to TN and FN .Which tells that all classifiers have good classification accuracy. Misclassification rate is low.

Figure 6a shows sensitivity, specificity and accuracy for five machine learning classifiers on Breast Cancer Wisconsin dataset .This shows that all classifiers perform well as accuracy is above 94% in all cases. But still SVM outperform others. RBF also compete well with SVM. Figure 6a also shows graphical demonstration of sensitivity, specificity and accuracy for five classification techniques on Breast Cancer Wisconsin dataset.

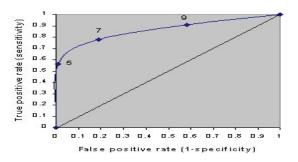


Figure 5b. shows sensitivity, specificity, True Positive Rate and False Positive Rate at three cut points and Area under this curve(ROC).

	Sensitivity	Specificity	Accuracy
KNN	97.80%	90.90%	95.42%
J48	95.60%	92.59%	94.56%
MLP	96.10%	93.80%	95.30%
RBF	95.40%	96.30%	95.71%
SVM	97.40%	96.20%	96.90%

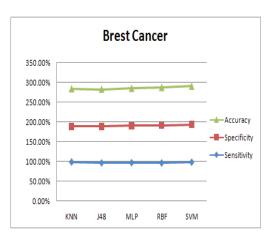


Figure 6a. Models comparison and Graphical representation of comparison on Breast Cancerdataset.

BREAST CANCER				
K	INN confusion matri	x		
а	b	classified as		
448	10	2		
22	219	4		
J	48 Confusion Matrix	X		
a	b	classified as		
438	20	2		
18	223	4		
Ν	/LP confusion Matri	X		
a	b	classified as		
440	18	2		
15	226	4		
F	BF Confusion Matri	X		
a	b	classified as		
437	21	2		
9	232	4		
SVM Confusion Matrix				
a	b	classified as		
446	12	2		
9	232	4		

Table 3. shows confusion matrix for five classification
techniques on Breast Cancer Wisconsin

Figure 6b shows TPR, FPR, ROC, F Measure and Kappa Statistics of various techniques for Breast Cancer. TPR for SVM is high among all and its FPR is low. SVM, F measure and Kappa statistics is again high. Which means that agreement between actual and predicted class is high. But ROC area of SVM is less then KNN, MLP as well as RBF. ROC area, Kappa Statistics and F Measure values are nearly equal to 1 for all models. So above models have good predictive capabilities for Breast Cancer Wisconsin dataset. From here we can see that RBF also perform equally well as SVM. But always there is a space for improvement as misclassification cost is very high so there is further chance of improvement. Figure 6b also shows graphical demonstration of performance measures for Breast Cancer dataset.

6.2 Liver Disorder

TP and FP both are not high on Liver disorder dataset which tells that little less classification accuracy. TN and FN values are more for all classification techniques which shows more misclassification rate on Liver disorder dataset. Figure 7a shows sensitivity, specificity and accuracy for five classification techniques on this dataset and this shows that all classifiers have accuracy with in range of 57.97% to 63.19%. SVM still outperform others but there is good chance of improvement in these techniques so that accuracy can be increased further. Figure 7a also shows their graphical demonstration.

Figure 7b shows TPR, FPR, ROC, F Measure and Kappa Statistics of various techniques for liver disorder dataset .TPR for SVM is high among all and its FPR is low. Low FPR signifies small false rate. SVM F measure, ROC area and Kappa statistics is more among all other techniques. Kappa Statistics and F Measure values are not nearly equal to 1. It reveals that there is very less agreement between actual and predicted class values. So above models have less predictive capabilities for

	BREAST CANCER WISCONSIN				
	True Positive Rate	False Positive Rate	F Measure	ROC	Kappa Statistics
KNN	0.954	0.067	0.954	0.988	0.8975
J48	0.946	0.064	0.946	0.955	0.8799
MLP	0.953	0.054	0.953	0.986	0.8958
RBF	0.957	0.034	0.97	0.986	0.9061
SVM	0.97	0.03	0.97	0.968	0.9337

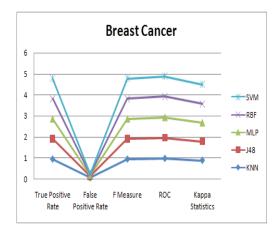


Figure 6b. TPR, FPR, ROC, F Measure and Kappa Statistics and their Graphical representation for various techniques on Breast Cancer.

Table 4. shows confusion matrix for five classification
techniques on Liver Disorder dataset.

liver Disorder				
ŀ	KNN confusion matri	ix		
а	b	classified as		
65	80	1		
57	143	2		
J	48 Confusion Matri	x		
а	b	classified as		
5	130	1		
15	195	2		
MLP confusion Matrix				
а	b	classified as		
25	120	1		
25	175	2		
F	RBF Confusion Matri	x		
a	b	classified as		
81	64	1		
70	130	2		
SVM Confusion Matrix				
a	b	classified as		
79	66	1		
61	139	2		

Liver disorder dataset. Thus there is a scope to enhance the performance of these techniques as balance factor is 1.4 which signifies dataset has good balance and classes are equally distributed. Figure 7b also provides graphical representation of performance measures for liver disorder.

6.3 Hepatitis Dataset

TP and FP both are high which tells that good classification accuracy on Hepatitis dataset. Misclassification rate is low as TN and FN are low. Figure 8a shows sensitivity, specificity and accuracy for five classification techniques on Hepatitis dataset .This shows that all classifiers perform well as accuracy is above 80% in all cases. But SVM and RBF have same accuracy and they outperform others.Figure8a also shows graphical demonstration of sensitivity, specificity and accuracy for five classification techniques on Hepatitis dataset.

Figure 8b shows TPR, FPR, ROC, F Measure and Kappa Statistics of various techniques for Hepatitis dataset .TPR for RBF and SVM is high among all and its FPR for RBF is low among all. Low FPR signify small false rate. RBF F measure, ROC area and Kappa statistics is again high for RBF among all. ROC values are nearly equal to 1. So above models have good predictive capabilities for Hepatitis dataset. From here we can see that RBF outperforms SVM. Figure 8b also shows graphical demonstration of TPR, FPR, ROC, F Measure and Kappa Statistics of various techniques for Hepatitis dataset.

6.4 Cleveland Heart Data set

Figure 9a shows sensitivity, specificity and accuracy for five classification techniques on Cleveland Heart dataset. TP and FP both are high which tells that good

	Sensitivity	Specificity	Accuracy
KNN	59.25%	71.50%	60.20%
J48	3.70%	92.59%	57.97%
MLP	17.24%	87.50%	57.97%
RBF	55.86%	65.00%	61.15%
SVM	54.48%	69.50%	63.19%

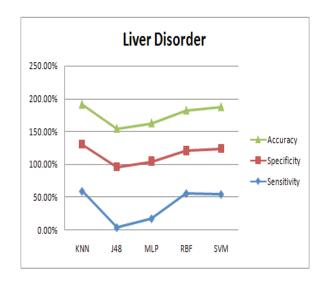


Figure 7a. Models comparison and Graphical representation of comparison on Liver Disorder dataset.

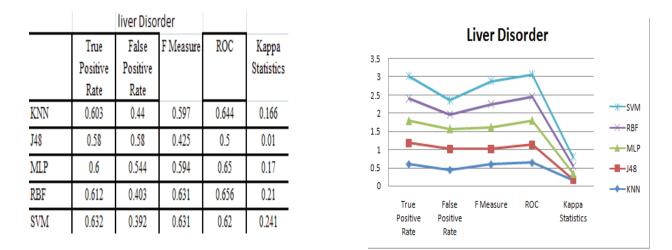


Figure 7b. TPR, FPR, ROC, F Measure and Kappa Statistics and their Graphical representation for various techniques on Liver Disorder.

classification accuracy. Misclassification rate is low as TN and FN are low for Cleveland Heart dataset. Balance factor of Cleveland Heart dataset is 2.7 which implies that it is highly unbalanced dataset and classes are not equally distributed. But all models have good predictive capabilities This shows that all classifiers perform well as accuracy is above 74% in all cases. SVM outperform others . As balancing factor of data set is high so data set is highly imbalanced although models have good predictive capabilities. Figure 9a also shows graphical demonstration of sensitivity, specificity and accuracy for five classification techniques on Cleveland Heart dataset. Figure 9b shows TPR, FPR, ROC, F Measure and Kappa Statistics of various techniques for Cleveland heart dataset.

TPR, F Measure and ROC of SVM is high among and its FPR is Low among all. ROC and F Measures values for RBF and SVM are nearly equal to 1. So above models have good predictive capabilities for Hepatitis dataset. From here we can see that SVM outperforms others for Cleveland heart dataset .Figure 9b also shows graphical demonstration of TPR , FPR, ROC, F Measure and Kappa Statistics of various techniques on Cleveland heart dataset. These days if some technique predicts one misclassification corresponding to one thousand correct predictions then technique must be left. This thing tells that in medical diagnosis cost of misclassification is very high. So model

	Hepatitis							
K	NN confusion matri	x						
a	b	classified as						
15	17	2						
11	112	4						
J48 Confusion Matrix								
a	b	classified as						
14	18	2						
7	116	4						
N	MLP confusion Matrix							
a	b	classified as						
18	14	2						
17	106	4						
R	BF Confusion Matri	X						
a	b	classified as						
21	11	2						
12	111	4						
SVM Confusion Matrix								
a	b	classified as						
19	13	2						
10	113	4						

 Table 5. shows confusion matrix for five classification

 techniques on Hepatitis dataset

	Sensitivity	Specificity	Accuracy
KNN	46.87%	91.05%	81.93%
J48	43.75%	94.30%	83.87%
MLP	56.25%	86.17%	80.01%
RBF	65.62%	90.24%	85.16%
SVM	59.37%	91.86%	85.16%

selection and its performance matters a lot. Based upon above results we have observed that single model of same architecture don't perform equally well on all above datasets. SVM and RBF have good performance on medical datasets. Performance further depends a lot on data. So choice of a stable model having less dependency on data with generalized behavior is challenging issue. In prospect an attempt will be made to improve performance of RBF Neural Networks because it performs equally well as SVM.

7. Conclusion

At present, different machine learning techniques can be used for the medical diagnosis and in this research work five predictive techniques evaluated are: K-Nearest Neighbors (KNN), Decision Tree, Artificial Neural Networks (ANNs), Radial Basis Function (RBF) neural networks and Support Vector Machine (SVM). Evaluation is done on the basis of Sensitivity, Specificity, Accuracy, F Measures, True Positive Rate and False Positive Rate, ROC and Kappa statistics. Error is represented by FP and FN which must be minimized to trim down misclassification .Our study illustrated that RBF and SVM model come out to be most excellent classifier for disease classification and prediction. Furthermore our conclusion is based on four datasets. Consequently more research work would be helpful to ensure further generalization of results on additional medical domains. There is a potential to improve performance of KNN, J48, MLP, RBF and SVM techniques by creating their ensemble model and creating diversity in ensemble which can be used

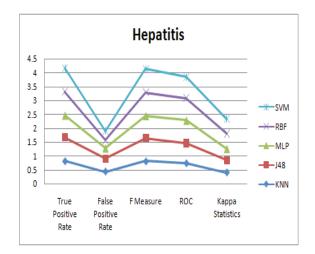
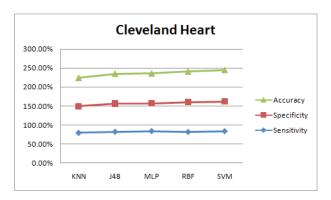


Figure 8a. Models comparison and Graphical representation of comparison on Hepatitis dataset.

	True	False	F Measure	ROC	Kappa						
	Positive Rate	Positive Rate			Statistics		200.00%			Нера	titis
KNN	0.819	0.44	0.812	0.744	0.4076		300.00% 250.00%				
J48	0.839	0.458	0.825	0.708	0.436		200.00%				
MLP	0.8	0.376	0.803	0.823	0.4101		150.00%				
RBF	0.852	0.293	0.852	0.808	0.5523		50.00%	+			
SVM	0.852	0.339	0.849	0.756	0.5309	1	0.00%	KNN	J48	MLP	RB

Figure 8b. TPR, FPR, ROC, F Measure and Kappa Statistics and their Graphical representation for various techniques on Hepatitis dataset.

	Sensitivity	Specificity	Accuracy
KNN	79.26%	69.78%	74.92%
J48	82.31%	74.10%	78.54%
MLP	84.14%	72.66%	78.87%
RBF	81.70%	78.41%	80.19%
SVM	83.83%	78.41%	82.17%



Accuracy Specificity Sensitivity

SVM

Figure 9a. Models comparison and Graphical representation of comparison on Cleveland dataset.

	True Positive Rate	False Positive Rate	F Measure	ROC	Kappa Statistics	Cleveland Heart	
KNN	0.749	0.259	0.748	0.836	0.4927	4.5 4 3.5	
J48	0.785	0.221	0.825	0.785	0.5644	3 2.5	- ж
MLP	0.789	0.221	0.788	0.806	0.5718		-×- -
RBF	0.789	0.251	0.797	0.855	0.6026	0.5 0 True False F.Measure ROC Kappa	-
SVM	0.802	0.201	0.802	0.886	0.6216	Positive Positive Statistics Rate Rate	

Figure 9b. TPR, FPR, ROC, F Measure and Kappa Statistics and their Graphical representation for various techniques on Cleveland Dataset.

Table 6.shows confusion matrix for five classificationtechniques on Cleveland datase

Cleveland Heart								
KNN confusion matrix								
а	b	classified as						
130	34	1						
42	97	2						
J48 Confusion Matrix								
a	b	classified as						
135	29	1						
36	103	2						
N	MLP confusion Matrix							
a	b	classified as						
138	26	1						
38	101	2						
R	RBF Confusion Matrix							
a	b	classified as						
134	30	1						
30	109	2						
SVM Confusion Matrix								
a	b classifie							
140	27	1						
30	30 109 2							

for medical diagnosis. Above analyzed techniques have incredible space for improvement to act as core part of MDSS which will help the physician to take efficient and reliable decision?

8. Acknowledgements

This work was carried out in part by Department of Computer Science and Engg., Guru Jambheshwar University of Science and Technology Hisar, Haryana, India. The authors thankfully acknowledge suggestions and guidance.

9. References

- 1. Mora M, Forgionne GA, Jupta J. Decision Making Support Systems: achievements, trends and challenges for the next decade. Idea-Group: Hershey. 2002.
- 2. Sunila G, Rishipal S. Machine Learning For Medical Decision Support Systems (MDSS): A review International

Journal of Applied Engineering Research. 2015; 10(13):32864–73.

- 3. Finlay PN. Introducing decision support systems Cambridge MA: Blackwell Publishers; 1994.
- 4. Panigrahy D, Sahu PK. Extraction of fetal Electrocardiogram (ECG) by extended state Kalman filtering and adaptive neuro-fuzzy inference system (ANFIS) based on single channel abdominal recording Sadhana c Indian Academy of Sciences 2015; 40(4):1091–1104
- 5. HoonK H, Poltayev R, Woo HB. An Intrinsic Image Representation and its Application to Left Ventricle Segmentation in Cardiac MRI Images. Journal of Medical Imaging and Health Informatics. 2014; 4(4):1–9
- Sunila G, Prabhat SP, Nirmal G. Decision Support System for Cardiovascular Heart Disease Diagnosis using Improved Multilayer. International Journal of Computer Applications. 2012; 45(8):12–20.
- 7. Guang HB, Hui WD, Yuan L. Extreme learning machines: a survey Springer Int. J. Mach. Learn. And Cyber. 2011. p. 107–22.
- Avci E, Turkoglu I. An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. Journal of Expert Systems with Application. 2009; 2(1):2873–78.
- 9. Das R, Abdulkadir S. Effective diagnosis of heart disease through neural networks ensembles. An international journal of expert system with applications. 2008; 36(4): 7675–80.
- Yao Z, Lei L, Yin J. R-C4.5 Decision tree model and its applications to health care dataset. Proceedings of International Conference on Services Systems and Services Management; 2005. p. 1099–1103.
- Farhad Soleimanian G, Seyyed RK, Isa M. New Approach in Bloggers Classification with Hybrid of K-Nearest Neighbor and Artificial Neural Network Algorithms. Indian Journal of Science and Technology. 2015 Feb; 8(3). Doi no: 10.17485/ ijst/2015/v8i3/59570
- Han J, Kamber M. Data Mining Concepts and Techniques .2nd Edition Morgan Kaufmann San Francisco; 2006.
- Hongmei Y, Yingtao J, Jun Z, Chenglin P, QinghuiL. A multilayer perceptron-based medical decision support system for heart disease diagnosis Expert Systems with Applications. 2006 Feb; 30(2):272–81.
- Patil S, Kumaraswamy Y. Intelligent and effective Heart Attack prediction system using data mining and artificial neural networks. European Journal of Scientific Research. 2009; 31: 642–56.
- 15. Navarroa FF, Martíneza CH, Ruizc R, Jose and Riquelme C. Evolutionary Generalized Radial Basis Function neural networks for improving prediction accuracy in gene classification using feature selection. International journal of Applied Soft Computing. 2012.p. 1787–1800

- Garci'a AS, Ferna'ndezLuengo ÆJ, Herrera F. A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Computing. Springer-Verlag; 2008.
- Cheng C, cowan FN, Grant PM. Orthogonal Least squares Learning Algorithm for Radial basis Function Networks. IEEE Transaction of Neural Network. 1991 Mar; 2(2):302–09.
- Mredhula L, Dorairangaswamy MA. Implementation of Image Fusion Algorithms for Clinical Diagnosis. Indian Journal of Science and Technology. 2015 July; 8(15). Doi no:10.17485/ijst/2015/v8i15/74197.
- Hsien-Tse C, Sheng-Fuu L, Yung-Chi H. Design and Identify Tubercle Bacilli Diagnosis System with TSK-type Neuro Fuzzy Controllers. Indian Journal of Science and Technology. 2012 Dec; 5(12). Doi no:10.17485/ijst/2012/ v5i12/30601.
- NookalaVenu, B. Anuradha. Multil-Kernels Integration for FCM Algorithm for Medical Image Segmentation using Histogram Analasis. Indian Journal of Science and Technology. 2015 July; 8(16). Doi no:10.17485/ijst/2015/ v8i16/61974
- 21. Flach PA. The Geometry of ROC Space: Understanding Machine Learning Metrics through ROC Isometrics.

Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003); Washington DC. 2003.

- 22. Srideivanai Nagarajan, Chandrasekaran RM. Design and Implementation of Expert Clinical System for Diagnosing Diabetes Using Data Mining Techniques. Indian Journal of Science and Technology. 2015 Apr; 8(8). Doi no: 10.17485/ ijst/2015/v8i8/69272
- 23. Shenbagarajan, V. Ramalingam, C. Balasubramanian, S. Palanivel.Tumor Diagnosis in MRI Brain Image using ACM Segmentation and ANN-LM Classification Techniques. Indian Journal of Science and Technology,2016 Jan, 9(1), Doi no:10.17485/ijst/2016/v9i1/78766.
- 24. Javad Aramideh, Hamed Jelodar. Application of Fuzzy Logic for Presentation of an Expert Fuzzy System to Diagnose Anemia. Indian Journal of Science and Technology. 2014 Jan; 7(7). Doi no:10.17485/ijst/2014/ v7i7/47973
- 25. Hsien-Tse Chen, Sheng-Fuu Lin, Yung-Chi Hsu. Design and Identify Tubercle Bacilli Diagnosis System with TSKtype Neuro Fuzzy Controllers. Indian Journal of Science and Technology. 2012 Dec; 5(12). Doi no:10.17485/ ijst/2012/v5i12/30601