ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(11), DOI: 10.17485/ijst/2016/v9i11/67151, March 2016* ISSN (Online) : 0974-5645

Analytical Study of Selected Classification Algorithms for Clinical Dataset

V. R. Suma, Shwetha Renjith, Sreeja Ashok* and M. V. Judy

Department of Computer Science and IT, Amrita School of Arts and Sciences, Amrita Vishwa Vidya peetham, Kochi - 682024, Kerala, India; sumavrillam92@gmail.com, shweshweranjith@gmail.com, sreeja.ashok@gmail.com, judy.nair@gmail.com

Abstract

The objective of this paper is to analyze and identify the best classification solution for clinical decision making. Several classification algorithms Like Discriminant Analysis (LDA), Support Vector Machine (SVM), Artificial Neural Network (ANN), Naive Bayes (NB), and Decision Trees are compared to find the optimum diagnostic accuracy. The performance of classification algorithms are compared using benchmark dataset, breast cancer. The effects of normalization using z-score and min-max approaches are also investigated. The results are compared based on different performance parameters like accuracy, sensitivity, specificity and root node error value. Accuracy has been improved for all classifications methods after normalizing the data set. Z-score normalization performs better for all the measures when compared to min-max normalization. The proposed approach shows higher accuracy rate for Naive Bayes algorithm when compared with the other algorithms.

Keywords: Accuracy, Classification, Min-Max Normalization, Sensitivity, Specificity, Z-Score Normalization

1. Introduction

Data mining is an analysis step in KDD process to investigate pre-existing huge database in order to create new information that is understandable for future use. DM has applied in many different areas, i.e., Financial Data Analysis, Retail Industry, Telecommunication Industry, Biological Data analysis, Scientific Applications, Medical Domain, Intrusion Detection etc. Data mining is used in medical field because it is used todeterminefresh trends, significant patterns from data. Classification, Clustering, Association etc., are the functionalities of data mining.Classificationisa supervised learning process that classifies dataand analyses and extracts representation for data classes. These models are called classifiers. These classifiers predicts the truthful class label. Classification can be explained as a two-stepprocedure[1]:

Step 1: Learning Step-Predict a classification model.

Training data set is analyzed by classification algorithms and the classifiers are derived in the form of classification rules.

Step 2: Classification Step-Model is used to predict class labels for new dataset.

The accuracy of the classifiers is identified with the percentage of valid data set that is rightly classified by the classifier. For that we use a Confusion matrix (Contingency Table). It is a square matrix where each feature represents instance of a predicted class and each line represents instance of authentic class. It visually represents the performance of an algorithm[1].

In this paper five efficient classification techniques such as Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Naive Bayes (NB), Artificial Neural Network (ANN), and Decision Tree (DT) are consider for comparing and evaluating the performance. Section 2 deals with explanation of different classification algorithms. Section 3 deals with advantages and disadvantages of classification methods. Section 4 explains the results and discussionandSectionconcludes with final remarks.

* *Author for correspondence*

2. Classification Algorithms

Several studies are done on different classification methods. Many researchers have compared the performance of different classification methods in different domains.

2.1 Decision Tree

A Decision Tree is a flow chart like structure used for decision analysis. A test of an attribute is represented by a non-leaf node. The outcome of the test represented by branches of the tree and leaf nodes represents a class label. A Root node is the top node. X is a given tuple, where the class label associated to it is unidentified. The attribute values are experienced for class labels and derived using decision tree. From source nodes to leaf nodes a path is constructed, that holds the class prediction for that tuple[2].

Decision Tree is easy to recognize. It deals both numeric and nominal dataset. It can represent complex decision boundaries and can handle multidimensional data. Accuracy is superior for decision tree and it follows greedy approach (top-down). Decision tree uses different algorithms like ID3 (Iterative Dichotomizer), C4.5 (successor of ID3), CART (Classification and Regression Tree).

Attribute selection measures**:** For selecting the element that better discriminates a data set is identified using different attribute selection measures[3]. Following are some of the attribute selection measures commonly used.

2.1.1 Information Gain

Information gain of class is calculating using Equation (1).

$$\text{Info}(S) = \sum_{o=1}^{m} po \log 2(po) \quad (1)$$

Information gain of each element is calculated using Equation (2).

$$\text{Info}_m(S) = \sum_{k=1}^{v} \left(\frac{FO}{F}\right) * info(Sk) 2(po) \quad (2)$$

Information gain is the difference between the information gain of a class and information gain of an attribute. It is given in Equation (3).

$$\text{Gain}(C) = \text{Info}(S) - \text{Info}_m(S) \quad (3)$$

The element that has utmost information gain can be chosen as the Splitting Attribute.

2.1.2 Gain Ratio

Using Equation (4) and Equation (5) we can find Gain ratio.

$$\text{Split Info}_c(F) = -\sum_{k=1}^{v} \left(\frac{FO}{F}\right) * log2\left(\frac{FO}{F}\right) \quad (4)$$

$$\text{Gain\_Ratio}(C) = \frac{Gain(C)}{Split\ Infoc(F)} \quad (5)$$

Table 1. Application of Decision Tree

Year	Author	Data set	Method	Result
2013	Athul Kumar Pandey, K L Jaiswal, Ashish Kumar Sen	Heart diesease data from machine learning respositories for UCI.	This applied in three types of decision tree J48 Un-pruned tree, J48 Pruned tree, J48 Reduced Error Pruning.	Applying reduced error pruning provides more compact decision rules and reduces the number of extracted rules with 75.73% accuracy
2008	G William Wong, Guangtao Ge.	Pancreatic cancer peptide mass-spectrometry data applied for Wekamat lab statistics and bioinformatics tool box.	Compare the act of a solitary desicion tree algorithm c4.5 with six different decision trees classifires ensembles.	Ensembles classifiers perform higher than single desicion trees with higher accuracy 99%.
2014	P Deepika, K Thenmozhi.	This uses different dataset with 13, 15 and 19 attributes.	Comparing desicion trees with other categorization methods. Decision tree type is selected using: Information Gain, Gini index and Gain ratio.	Decision tree classifiers are selected best for its effortlessness with accuracy 99.62% saccuracy.
2007	C Tjortjis, J A Keane, M Saraee, B Theodoulidis.	Med_123 and med_newlive	Build T3, a desicion Tree classifiers which builds models based on known classes in order to achieve better accuracy.	T3 builds a desicion Tree with depth 3 and results high accuracy, keeping the tree size small, with 0% classification error.

The element with maximum gain ratio is choosing as the splitting attribute.

Summary of the applications of decision tree on clinical data setfrom the literature survey is given in Table1[4–7].

2.2 Naive Bayes

This is a family of simple probabilistic classifiers based on Bayes theorem with independent assumption involving predictors. NB is well scalable and based on occurrenceofdata. NB is a supervised learning method used in large dataset and in complex situations[8]. This model uses maximum possibility method, estimating the parameter of an arithmetical model.

Using Equatin (6) and Equation (7), it calculates posterior probability.

$$\text{Posterior Probability} = \frac{priori * likelihood}{evidence} \quad (6)$$

$$\text{P(c | x)(Conditional Independence} = \frac{p\left(\frac{x}{c}\right)/_{p(c)}}{p(x)} \quad (7)$$

P (c|x) is the posterior probability of target class (predictor is given). P (x|c) is the probability of predictor (class is given). P (c) is the prior probability of class. P (x) is the prior probability of predictor.It assumes that the outcome of the value of the classifier on a given class c is independent of the values of the other predictors. This is called conditional independence.

First derive an independent feature model called naive Bayes probability models. Using a decision rule the Bayes classifier combine with this model. A MAP (Maximum a Posteriori) is a familiar rule to pick the most probable hypothesis[9].Various applications of Naive Bayes on clinical data set from literature survey isshown in Table 2[10–13].

2.3 Support Vector Machine (SVM)

It is a discriminative classifier and is used for separating hyperplane. It is based on the concept of decision planes that define decision boundaries. This decision plane separates a set of objectives of two classes[14]. SVM performs linear classification and non-linear classification. A classification that is non-linear separates objects into individual groups. In Non-linear classification SVM uses kernel functions. Popular kernel methods are Fisher kernel, RBF kernel, String kernel, Polynomial kernel, Graph kernel. SVMmostly uses RBF kernel[14,15].

The RBF kernel oftwosamples is represented using Equation (8).

$$K(x, x1) = \exp(-|(|x\text{-}x1|)|^2/2\,\sigma^2) \quad (8)$$

Table 2. Application of Naive Bayes classification

Year	Author	Data set	Method	Result
2004	Adam Bucinski, Tomaz Baczek, Jerzy Krysinki, Renata Szoszkiewics, Jerzy Zaluki	A genome-wide dataset on Alzheimer's disease applled to MATLAB.	Bayesian algorithm compared with several learning algorithms like data preprocessing, feature selection, model learning and classification scoring.	Bayesian algorithm is best compare to other methods with 73% accuracy.
2008	Jau-Huei Lin, Peter j Hang	EHR, EDW, IHC	Build a model to find the missing data, NB as the underlying model.	The classification performed better when using explicit missing value treatment.
2012	Bekir Karlik, Emre Oztoprak	TPMT polymorphisms applied to Rapid Miner.	Automatic detection of cancer, cancer risk, and not-risk by using pharmacogenomics data using 2 step that is training step and prediction step.	It provides small amount of preparartion data to approximate the parameters that are used for classification.
2008	Dinora Araceli Morales, Endika Bengoetxea, Pedro Larranga, Miguel Garcia, Yous Franko, Monica Fresnada, Marisa Merino	63 clinical files of the IVF programme in Clinical Pillar in San Sebastian applied to Elvira Software	The best embryos are selected classification problem. Compare with different Bayesian approach. The success is achieved when implanation is obtained (Pregnanacy).	The result is obtained based on the filter technique chosen to select the subset and performance of the Bayesian classification method.

Table 3. Application of SVM

Year	Author	Data set	Method	Result
2000	Terrence S Furey, Nello	A genome-wide dataset on Alzheimer's disease applled to MATLAB.	Bayesian algorithm compared with several learning algorithms like data preprocessing, feature selection, model learning and classification scoring.	Bayesian algorithm is best compare to other methods with 73% accuracy.
2013	Jau-Huei Lin, Peter j Hang	EHR, EDW, IHC	Build a model to find the missing data, NB as the underlying model.	The classification performed better when using explicit missing value treatment.
2014	Bekir Karlik, Emre Oztoprak	TPMT polymorphisms applied to Rapid Miner.	Automatic detection of cancer, cancer risk, and not-risk by using pharmacogenomics data using 2 step that is training step and prediction step.	It provides small amount of preparartion data to approximate the parameters that are used for classification.
2015	Dinora Araceli Morales, Endika Bengoetxea, Pedro Larranga, Miguel Garcia, Yous Franko, Monica Fresnada, Marisa Merino	63 clinical files of the IVF programme in Clinical Pillar in San Sebastian applied to Elvira Software	The best embryos are selected classification problem. Compare with different Bayesian approach. The success is achieved when implanation is obtained (Pregnanacy).	The result is obtained based on the filter technique chosen to select the subset and performance of the Bayesian classification method.

(x-x1)^2 is the Euclidean distance

For linear classification there are two sets of classes x1, x2. If we denote it in a diagram, x1 in x axis and x2 in y axis. One or more hyper plane can separates the data onto two groups. But we choose the best hyper plane that separates the data into twogroups effectively. The best hyper plane leaves utmost margin for both classes. Support Vectors are the pointsadjacentto the hyper plane. This hyper plane is defined by an equation y = ax+b. This equation is also used for regression. It mainly focused on binary classification. Summary of different applications of SVM for clinical dataset is given in Table 3[16–19].

2.4 Linear Discriminant Analysis (LDA)

Basically there are two types of Discriminant Analysis Method. One is Linear Discriminant Analysis and Quadratic Discriminant Analysis. It is used to categorize the dataset with two or more classes.LDA is attributeselection method[20]. It is a global method of Fishers Linear Discriminant. It selects a linear combination of features and classification is done based on this combination called linear classifier. It is also used for regression analysis.

LDA measurement of each class is normally distributed. Maximum separability provided with this method is by maximizing the ratio of between class variance and within class variance. Dividethe data set into preparation set and test data. Find the mean of each data set and find the mean of entire data set by merging the two dataset, Equation (9)[20,21].

$$\mu 3 = \mu 1 * p1 + \mu 2 * p2 \quad (9)$$

Where p1, p2 are the probability factors

Next step is to find inter class and intra class scatter matrices.Intra class scatter is assumed to be the covariance of each class. Scatter matrices are computed using Equation (10).

$$Sw = \sum_j Pj * cov(j) \quad (10)$$

Covariance matrix is computed using Equation (11).

$$covj = (xj - \mu j)(xj - \mu j)^{\wedge T} \quad (11)$$

The inter class scatter matrix is computed using Equation (12).

$$So = (\mu i - mean\ of\ entire\ dataset) * (\mu i - mean\ of\ entire\ dataset)^T \quad (12)$$

Then discover the Eigen values and Eigen vectors of the matrix. To discover Eigen vector we primarily locate the value of λ that satisfy the Equation (13).

Table 4. Application of LDA

Year	Author	Data set	Method	Result
2009	Desheng Huang, Yu Quan, Miao He and Baosen Zhou.	6 different data set	Comparing LDA, with PAM, SDA, SLDA, SDDS and SCRDA.	High accuracy
2011	Chin Lee, Brittany Nkounkou, Chun-His Huang	Clinical dataset	Comparing LDA, with SPRT and MSPRT.	MSPRT is better than LDA
2006	Patricia J Pardo, Apostolos P Georgopoulos, John T Kenny, Truci K Stuve, Robert L Finding, S Charles Schulz	Clinical interview of patients and their parents	Classify the dataset of three classes.	High accuracy
2008	Edmundo Bonilla Huerta, Beatrice Duval, Jin-Kao Hao	Seven well known public dataset.	Combing Genetic algorithm with LDA.	81-100% accuracy with less number of Genes.

$$(A - \lambda I) \tag{13}$$

In the decreasing order of Eigen values, sort the Eigen vectors. Select the first k Eigen vectors with the highest Eigen value. Then transform the samples into fresh space. When the transformations are ended, trace the Euclidean distance or RMS distance to classify the data points. The Applications of LDA based on the literature survey is given in Table 4[22–25].

2.5 Artificial Neural Network

It is a family of arithmetic learning methods.It is an electronic system of Neurons with a set of input values associated with weights and a function that sums up the weights. There are three layers. Input, Hidden, Output. These layers arrange the Neurons. The input layer consists of input values, the functions are taking place in unseen layers and several unseen layers are presented on one neural network. Final layer is the output layer which produces output, one node for which class. Back propagation algorithm is a procedure for training multilayer Artificial Neural Network[26–28].

Back propagation maps the input to the best output. This calculates the slope of a loss function of optimization method. Optimization method means choice of best elements from many essentials. To compute the loss gradient function the output of each input value should be known. Loss function maps an event or values of one or more variable to a real number. The Applications of Neural network is given in Table 5[29–32].

3. Advantages and Disadvantages of Classification Methods

Advantages and Disadvantages of all the five classification methods explained in this paper is given in Table 6.

Table 5. Application of ANN

Year	Author	Data set	Method	Result
2011	Dr. Usharani	Cleveland Database	Comparing LDA, with PAM, SDA, SLDA, SDDS and SCRDA.	94% accuracy
2008	K Mumthaz, S A Sheriff, Dr. K Duraiswamy.	Wisconsin breast Cancer Database	Using SOM, ART, BPN, to produce the result benign or malignant[30].	BPN perform well with other methods.
2015	Patricia J Pardo, Apostolos P Georgopoulos, John T Kenny, Truci K Stuve, Robert L Finding, S Charles Schulz	Hepatitis, Wisconsin breast Cancer and Stat log heart diesease dataset and it applied in MATLAB tool version R2012.	Acquire smooth dataset and classify using BPN[31].	97.3% accuracy
2004	Adam Bucinski, Tomaz Baczek, Jerzy Krysinki, Renata Szoszkiewicz, Jerzy Zaluski	718x11 data matrices and Applied in Statistical Neural Network Software V6	PCA first apply to produce best subset and then apply ANN to produce whether the disease present or absent[32].	High accuracy

Table 6. Advantages and disadvantages of classification methods

Method	Advantages	Disadvantages
Linear Discriminant Analysis	Used for binary classification. Class Label is output. Comparable to regression. Low error rate. Easy to implement.	Sensitive to detect outliers. May overfit the data. LDA is a parametric method.
Support vector Machine	High accuracy. Popular in text classifiaction. Uses kernel functions. It has a regularization parameter.	Depending on the kernel. Limitation of speed and size. Presents of discrete data.
Artificial Neural Network	Easy to use and compare. Non-parametric. Solve complicated problems.	Slow process. Black box approach. Computational problems. Overfit. Cannot be retrained.
Naive Bayes	Simple. Needless training data. Perform well.	The class label and attribute value not occur together. The consistent accuracy of NB only gets with large dataset. Loss of accuracy due to class conditional independence. Dependency between variables.
Decision Tree	Easy to interpret and explain. Non-parametric. Can handle both character and numeric data. Can handle missing values and errors. Working with continuous attribute. Robust.	Overfit. Low performance. Poor resolution on data with complex relationship among variable. Practically limited to classification. Poor resolution with continuous expectation variable.

4. Results and Discussion

We have selected breast cancer data set from UCI with 11 variables and 699 observations for analysis. The performance of four classification methods LDA, SVM, NB, ANN is evaluated using three measures like Accuracy, Sensitivity and Specificity and Decision Trees are evaluated using percentage of root node error value. R (V3.1.2) is used for experiment and analysis. Table 7 describes the features of the dataset.

Table 7. Breast cancer dataset and attributes

Attribute	Type
id	Integer
Clump Thickness	Integer
Uniformity of Cell Size	Integer
Uniformity of Cell Shape	Integer
Marginal Adhesion	Integer
Single Epithelial Cell Size	Integer
Bare Nuclei	Integer
Bland Chromatin	Integer
Normal Nucleoli	Integer
Mitoses	Integer

Performance metrics used for evaluation are

- Accuracy- test set tuple that are appropriately classified by the classifier.

$$Accuracy = \frac{number\ of\ true\ positives + number\ of\ true\ negatives}{no:of true\ positives + false\ positives + false\ negatives + true\ negatives}$$

- Sensitivity- rate of true positives; positive tuples that are acceptably classified.

$$Sensitivity = \frac{number\ of\ true\ positives}{no:of true\ positives + no:of\ false\ negatives}$$

- Specificity- rate of true negatives; negative tuples that are appropriately classified.

$$Specificity = \frac{number\ of\ true\ negatives}{no:of\ false\ positives + no:of\ true\ negatives}$$

Usually clinical dataset follows huge variation in dataset; In order to compare the performance, the analysis is done with normalization and without normalizationbefore classification. Accuracy, sensitivity and specificity of classification methods without normalization is given in Table 8.

Table 8. Analysis done without normalization

Without Normalization			
Classification Method	**Accuracy**	**Sensitivity**	**Specificity**
LDA	54	65	35
ANN	54	67	32
SVM	54	65	34
NB	56	66	36

4.1 With Normalization

Two normalization methods are applied here for comparing the performance.

- Min-max normalization – linearly convertactual data value such that smallest and highest of transformed data take certain values. Frequently 0 or 1.

$$X' = \frac{x - x\min}{x\max - x\min}$$

- Z-Score normalization – linearly changereal data value such that the mean value of the transformed data is 0 while the value of standard deviation is 1.

$$X' = \frac{x - \mu}{\sigma}$$

Table 9 and 10 explains the analysis results on applying both normalization methods.

Table 9. With min-max normalization

With min-max normalization			
Classification Method	**Accuracy**	**Sensitivity**	**Specificity**
LDA	55	63	30
ANN	54	68	34
SVM	56	67	38
NB	55	66	35

Table 10. With Z-Score normalization

With Z-Score normalization			
Classification Method	**Accuracy**	**Sensitivity**	**Specificity**
LDA	55	65	35
ANN	54	68	32
SVM	56	64	33
NB	66	65	34

Table 11shows the results of Decision Treeevaluation with and normalization on the dataset.

Table 11. Decion Tree results

Method	**Without normalization**	**With normalizatoin**	
		Min - max	**Z - Score**
J48	33.5% root node error.	32% root node error.	32.5% root node error.
Classification Tree	34% root node error.	33% root node error.	40% root node error.
Regression Tree	22% root node error.	23% root node error.	23.5% root node error.

Accuracy has improved for all classifications methods with normalization. Accuracy of Naive Bayes improved with Z-score normalization. Out of the two normalization methods, z-score performs better for all the three measures when compared to min-max normalization.

In Decision Tree, Regression Tree has less root node error irrespective of normalization methods.Min max normalization reduces the root node error of J48 and Classification Tree whereas z-score normalization increases the error rate.

5. Conclusion

The breast cancer data set with 11 attributes from UCI is tested with selected classification algorithms. All the algorithms do well, but the result show that Naïve Bayes classifies better than other algorithms with better accuracy. Classification is an important technique of Data Mining and it can be applied to any data set like financial data, market data, biological data for predication and proper labeling. The outcomeis varying depending on the data set and methods applied. Most of the classification provides better outcomeafter normalization. All the classification methods are vitalandwe can use it depends on the requirements.

6. Acknowledgement

This work is supported by the DST Funded Project, (SR/CSI/81/2011) under Cognitive Science Research Initiative

in the Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham University, Kochi.

7. Reference

1. Han J, Pei J, Kamber M. Classification advanced methods. A volume in The Morgan Kaufmann Series in Data Management Systems; 2012. p. 393–442.
2. Chakradeo SN, Abraham RM, Rani BA,Manjula R. Data mining: Building social network. Indian Journal of Science and Technology. 2015 Jan; 8(S2):212–6.
3. Ravindra C, Gummadi A, Yedukondalu G, Raju UNPG. Classification by decision tree induction algorithm to learn decision trees from the class labeled training tuples. International Journal of Advanced Research in Computer Science and Software Engineering. 2012; 2(4):427–34.
4. Pandey AK, Pandey P, Jaiawal KL, Sen AK. A heart disease prediction model using Decision Tree. IOSR Journal of Computer Engineering (IOSR-JCE). 2013 Jul/Aug; 12(6):83–6. e-ISSN: 2278-0661, p- ISSN: 2278-8727.
5. Banfield R, Hall L, Bowyer K, Kegelmeyer W. A comparison of decision tree ensemble creation techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence.2007; 29(1):173–80.
6. Deepika P, Thenmozhi K. A heart disease prediction using classification with different decision tree techniques. International Journal of Engineering Research and General Science. 2014 Oct/Nov; 2(6):6–11.
7. Tjortjis C, Keane JA, Saraee M, Theodoulidis B.Using T3 improved decision tree classifier, for mining stroke related data. Methods Inf Med. 2007; 46(5):523–9.
8. Archana S, Elangovan K.Survey of classification techniques in data mining. International Journal of Computer Science and Mobile Applications. 2014 Feb; 2(2):65–71. ISSN: 2321-8363.
9. Bhavsar H, Ganatra A. A comparative study of training algorithms for supervised machine learning. IJSCE. 2012 Sep; 2(4). ISSN: 2231-2307.
10. Coopper GF, Hennings-Yeomans P, Shyam. An efficient Bayesian method for predicting clinical data outcomes from Genomic wide data. AMIA Annu Symp Proc 2010; 2010. p. 127–31. Published Online 2010 Nov 13.
11. Lin J-H, Haug PJ.Exploiting missing clinical data in Bayesian network modelling for predicting medical problems. Journal of Biomedical Informatics. 2008 Feb; 41(1):1–14.
12. Karlik B, Oztoprak E.Personalized cancer treatmentusing Naive Bayes Classifier.International Journal of Machine Learning and Computing. 2012 Jun; 2(3):339–44.
13. Karthika S, Sairam N. A Naive Bayesian classifier for educational qualification. Indian Journal of Science and Technology. 2015 Jul; 8(16).
14. Guyon I, Weston J, Barnhill S,Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn. 2002; 46:389–422.
15. Chen K-H, Wang K-J, Tsai M-L, Wang K-M, Adrian AM, Cheng W-C, Yang T-S, Teng N-C, Tan K-P, Chang K-S. Gene selection for cancer identification: A decision tree model empowered by particle swarm optimization algorithms. BMC Bioinformatics. 2014; 15:49.
16. Furey TS, Cristianini N, Duffy N, Bednarski DW, Shummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data.Bioinformatics. 2000; 16(10):906–14
17. Kumari VA, Chithra R. Classification of diabetes disease using support vector machine. IJERA. 2013 Mar/Apr; 3(2): 1797–801. ISSN: 2248-9622.
18. Qiu Y, Jiang H, Shimada K, Hiraoka N, Maeshiro K, Ching W-K, Aoki-Kinoshita KF, Furuta K. Towards prediction of pancreatic cancer using SVM study model. JSM Clinical Oncology and Research. 2014; 2(4):1031.
19. Cong G, Tan KL, Furey TS, N. Cristianini, Duffy N, Bednarski DW, Schummer M. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics. 2005; 16(10):906–14.
20. Huang M-L, Hsu Y-Y. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. J Biomedical Science and Engineering. 2012; 5:526–33.
21. Anuradha C,Velmurugan T. A comparative analysis on the evaluation of classification algorithms in the prediction of students' performance. Indian Journal of Science and Technology. 2015 Jul; 8(15).
22. Sharma A,Paliwal KK. Cancer classification by gradient LDA technique using microarray gene expression data. Data Knowl Eng. 2008;66:338–47.
23. Lee C, Nkounkou B, Huang C-H. Comparison of LDA and SPRT on clinical dataset classification. 2011 Apr 19; 4:1–7. DOI:10.4137/BII.S6935.
24. Pardo PJ, Georgopoulos AP, Kenny JT, Stuve TK, Findling RL, Schulz SC.Classification of adolescent psychotic disorders using linear discriminant analysis. 2006 Oct; 87(1-3):297–306.
25. Huerta EB, Duval B, Hao J-K. Gene selection for micro array data by a LDA based genetic algorithm. PRIB, LNBI. Heidelberg: Springer-VerlagBerlin. 2008; 5265:252–63.
26. Gharehchopogh FS, Khaze SR, Maleki I. A 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):237–46.
27. Gorynski K, Safian I, Gradzki V, Marszall MP, Krysinski J, Goryński S, Bitner A, Romaszko J, Bucinski A. Artificial neural networks approach to early lung cancer detection. Cent Eur J Med. 2014; 9(5):632–41. DOI: 10.2478/s11536-013-0327-6.
28. Ghwanmeh S, Mohammad A, Al-Ibrahim A. Innovative artificial neural networks-based decision support system for heart diseases diagnosis. Journal of Intelligent Learning Systems and Applications. 2013; 5:176–83.
29. Vaidehi K, Subashini TS. Breast tissue characterization using combined K-NN classifier. Indian Journal of Science and Technology. 2015 Jan; 8(1):23–6.

30. Mumthaz K, Sheriff SA, Duraiswamy K.Evaluation of neural network based classification systems for clinical cancer data classification. Journal of Computer Applications. 2008; 1(4):9–14.
31. Pardo PJ, Georgopoulos AP, Kenny JT, Stuve TK, Findling RL, Schulz SC. Knowledge mining from clinical dataset using rough data set and back propagation neural network. Computational and Mathematical Methods in Medicine. 2015; 2015:13. Article ID: 460189. Available from: http://dx.doi.org/10.1155/2015/460189
32. Suganya P, Sumathi CP.A novel metaheuristic data mining algorithm for the detection and classification of parkinson disease. Indian Journal of Science and Technology. 2015 Jul; 8(14).