ISSN : 0976-8491 (Online) | ISSN : 2229-4333(Print)

# Supervised Machine Learning Approaches for Medical Data Set Classification - A Review

[1]S. Anto, [2]Dr.S.Chandramathi
[1]Dept. of CSE, VLB Janakiammal College of Engineering and Tech, Coimbatore, India
[2]Dept. of ECE, Sri Krishna College of Engineering and Technology, Coimbatore, India

**Abstract**
Clinical decision making, using medical expert systems, is a complex task as it requires more accuracy. Hence the design of such medical expert systems requires relevant and the most suitable machine learning algorithm. This paper reviews the various supervised machine learning classification approaches available along with their functional use in medical field. A number of classification algorithms are considered and reviewed for their relative performances and practical usefulness on different types of health care datasets. This review gives an inference that the performance of the classification technique will depend on the features of the dataset that is analyzed with more emphasis on the health care dataset. While keeping the classification accuracy and speed as major criteria of this study, it is inferred that the SVMs and Neural Networks are more suitable for medical dataset classification with higher performance.

**Keywords**
Machine learning algorithms, Supervised learning, Classifiers, Intelligent Data Analysis

## I. Introduction

The simplest AI applications can be divided into two types: classifiers and controllers. Controllers do however also classify conditions before inferring actions, and therefore classification forms a central part of many AI systems. Classifiers are functions that use pattern matching to determine a closest match. They can be tuned according to examples, making them very attractive for use in AI. These examples are known as observations or patterns. In supervised learning, each pattern belongs to a certain predefined class. A class can be seen as a decision that has to be made. All the observations combined with their class labels are known as a data set. When a new observation is received, that observation is classified based on previous experience. Various classification techniques are used in medical field in several areas including prediction of effectiveness of surgical procedures, medical tests, medication, and the discovery of relationships among clinical and diagnosis data [1].

### A. Comparison Problem

At a first glance it may seem that comparing classifiers is as easy as error counting. Even if the issues of computing time are hopped still there are two other factors of importance: For what application(s) will the classifier be used and by whom? It is perfectly clear that performance differences are a function of class distributions and sample sizes and therefore of the application. If one tries to draw conclusions that are application independent and thus distribution free, then only performance bounds can be obtained, e.g. in terms of the Vapnik-Chervonenkis complexity of classifiers [33]. This is not the only issue here. The interest is more on the performance for practical applications. Therefore, an application domain has to be defined. The traditional way to do this is by a diverse collection of medical datasets.

A second issue, causing even more problems, is the dependence of the performance of some classifiers on the skill of the analyst who applies them. Some classifiers are very flexible, with many user-adjustable parameters, others are almost entirely automatic. Almost all of the traditional classification techniques are well defined and can be used without much user interaction. This holds for linear and nonlinear discriminant analysis, the nonparametric methods as well as for decision trees. This has several consequences for the issue of comparing classifiers on a particular application.

## II. Logic Based Algorithms

Logic Based Techniques have been studied extensively in machine learning areas and have been applied successfully to a wide range of domains. This sections details about two important logic based algorithms: decision trees and rule based algorithms.

### A. Decision Trees

Decision trees are predictive models, used to graphically organize information about possible options, consequences and end value. They are used in computing for calculating probabilities and data mining, and the decision trees examples below relate to 'simpler' decision making, so to speak. Decision trees are used for deciding between several courses of action. They create a visual representation of the various risks, rewards and potential values of each option. A summary, importance and effectiveness of the decision trees were presented by Murthy in his work about decision tree [2]. Decision trees classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. Murthy argues that the Decision trees are a way to represent rules underlying data with hierarchical, sequential structures that recursively partition the data [2]. A decision tree can be used for data exploration in one or more of the following ways:

#### 1. Description

To reduce a volume of data by transforming it into a more compact form this preserves the essential characteristics and provides an accurate summary.

#### 2. Classification

Discovering whether the data contains well-separated classes of objects, such that the classes can be interpreted meaningfully in the context of a substantive theory.

#### 3 Generalization

Uncovering a mapping from independent to dependent variables that is useful for predicting the value of the dependent variable

in the future.

Automatic construction of rules in the form of decision trees has been attempted virtually in all disciplines in which data exploration methods have been developed. Decision trees have proved to be valuable tools for the description, classification and generalization of clinical data [3] [46]. Decision-tree learning is one of the most successful learning algorithms, due to its various attractive features: simplicity, comprehensibility, no parameters, and being able to handle mixed-type data. The performance of fully expanded decision tree classifiers and Pruned decision tree classifiers were studied based on various training and test datasets [3]. Similarly, pruned decision tree classifiers were used to examine their differentiation capabilities. In order to build a generalized decision support system for heart sound diagnosis, they divided the problem into sub problems, dealing with either one morphological characteristic of the heart-sound waveform or with difficult to distinguish cases. Relevance analysis on the different heart sound features demonstrated that the most relevant features are the frequency features and the morphological features. Better accuracy was achieved using decision trees with a fraction of the 100 features. Pruned Differentiation decision trees did not significantly change the classification accuracy of the decision trees both in terms of partial classification and overall classification as well.

### B. Rule Based Algorithms

A rule-based classifier is a technique for classifying records using a collection of"if ... then ..." rules. Systems that learn sets of rules have a number of desirable properties. Rule sets are relatively easy for people to understand, and rule learning systems outperform decision tree learners on many problems. Rule sets have a natural and familiar first order version, namely Prolog predicates, and techniques for learning propositional rule sets can often be extended to the first order case. Certain types of prior knowledge can also be easily communicated to rule learning systems. Many of the techniques used in modern rule learners have been adapted from decision tree learning. Most widely-used decision tree learning systems use an over fit-and-simplify learning strategy to handle noisy data: a hypothesis is formed by first growing a complex tree which "overfits" the data, and then simplifying or pruning the complex tree. A variety of methods have been proposed to prune trees, but one active technique is Reduced Error Pruning (REP). REP can be easily adapted to rule learning systems. In REP for rules, the training data is split into a growing set and a pruning set. First, an initial rule set is formed that overfits the growing set, using some heuristic method. This overlarge rule set is then repeatedly simplified by applying one of a set of pruning operators; typical pruning operators would be to delete any single condition or any single rule. At each stage of simplification, the pruning operator chosen is the one that yields the greatest reduction of error on the pruning set. Simplification ends when applying any pruning operator would increase error on the pruning set.

One weakness with rule learning systems is that they often scale relatively poor with the sample size, particularly on noisy data. Given the prevalence of large noisy datasets in real-world applications this problem is of critical importance.

Furnkranz [4,5] provides an analysis of the behavior of separate-and-conquer or covering rule learning algorithms by visualizing their evaluation metrics and their dynamics in PNspace, a variant of ROC-space. There are different pruning techniques for this type of inductive rule learning algorithm. Its main contributions are two new algorithms: Top-Down Pruning (TDP), an approach that combines pre- and post-pruning, and Incremental Reduced Error Pruning (I-REP), a very efficient integration of pre-and post-pruning.

Pre-pruning methods deal with noise during learning. Instead of trying to find a theory that is complete and consistent with the given training data, heuristics (i.e., stopping criteria) are used to relax this constraint by stopping the learning process although some positive examples may not yet be explained and some of the negative examples may still be covered by the current theory. The final theory is learned in one pass. Most separate and conquer rule learners, like CN2 [47] and Fossil [4], use this form of noise handling. Another family of algorithms deals with noise by simplifying a previously learned overfitting theory. These post-pruning algorithms typically first induce a theory that is complete and consistent with the training data.

The concept of machine learning using decision tree has also been widely used in ophthalmology, especially for diagnosis of glaucoma [6]. Recent studies evaluated the application of machine classifiers in visual field interpretation of glaucoma. Ruchika Gupta's work [7] presents the design and development of the Rule based Decision support system for the diagnosis of the glomerular diseases. Each of the glomerular diseases has been represented by one or more possible combinations of 29 clinico pathological features. For 612 real world patient cases it produced 83.2 % accuracy. Later the rule based classifiers evolved further by the introduction of fuzzy logic and fuzzy sets. Angelov implemented the fuzzy rule based classifiers with wine reproduction data set [8]. These fuzzy rule based classifiers have the property of learning and adapting to new data samples as well as to new rules dynamically. For wine reproduction data set with 28 rules it produced an accuracy of 97%.

Presently instead of using purely rule based decision support system, hybrid systems were introduced in clinical diagnosis [9]. So the introduction of another system with the rule based system can provide required knowledge and hence achieves more accuracy in decision making. Ashwin Kumar [9] discussed about the hybrid system which has case based reasoning system and rule based one which is used for making clinical decision in ICU.

## III. Statistical Learning Methods

### A. Naive Bayes Classification

Naive Bayesian networks (NB) are very simple Bayesian networks which are composed of directed acyclic graphs with only one parent representing the unobserved node and several children corresponding to observed nodes with a strong assumption of independence among child nodes in the context of their parent. The basic independent Bayes model has been modified in various ways in attempts to improve its performance. Semi-naive Bayesian classifier is another important attempt to avoid the independence assumption. Naive Bayes (NB) utilizes a probabilistic method for classification by multiplying the individual probabilities of every attribute-value pair. This simple algorithm assumes independence among the attributes and even then provides excellent classification results. A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent

feature model".

The simple Bayesian classifier is known to be optimal when attributes are independent given the class, but the question of whether other sufficient conditions for its optimality exist has so far not been explored [10]. Empirical results showing that it performs surprisingly well in many domains containing clear attribute dependences suggest that the answer to this question may be positive. The major advantage of the naive Bayes classifier is its short computational time for training. In addition, since the model has the form of a product, it can be converted into a sum through the use of logarithms with significant consequent computational advantages. Domingos [37] has verified that the Bayesian classifier performs quite well in practice even when strong attribute dependencies are present. It has also been proved Bayesian classifier does not require attribute independence to be optimal under zero-one loss. Some necessary and some sufficient conditions have been derived for the Bayesian classifier's optimality. In particular, it has been proved that the Bayesian classifier is an optimal learner for conjunctive and disjunctive concepts, even though these violate the independence assumption. As a probability-based statistical classification method, the Naïve Bayesian classifier has gained wide popularity despite its assumption that attributes are conditionally mutually independent given the class label [11]. Improving the predictive accuracy and achieving dimensionality reduction for statistical classifiers has been an active research area in data mining. The results suggest that on an average, with Minimum Description Length (MDL) discretization the Naïve Bayes Classifier seems to be the best performer compared to popular variants of Naïve Bayes as well as some popular non-Naïve Bayesian statistical classifiers. A Hybrid feature selection algorithm (CHI-WSS) that helps in achieving dimensionality reduction by removing irrelevant data, increasing learning accuracy and improving result comprehensibility was proposed [11]. Experimental results of this algorithm suggest that on an average the Hybrid Feature Selector gave best results compared to individual techniques with popular filter as well as wrapper based feature selection methods. The proposed algorithm [11] which is a multi-step process utilizes discretization, filters out irrelevant and least relevant features and finally uses a greedy algorithm such as best first search or wrapper subset selector. For experimental validation of the algorithm, two established measures to compare the performance of statistical classifiers namely; classification accuracy (or error rate) and the area under ROC were used. This work demonstrates that the proposed algorithm using generative Naïve Bayesian classifier on the average is more efficient than using discriminative models namely Logistic Regression and Support Vector Machine. This work based on empirical evaluation on publicly available datasets validates the hypothesis of development of parsimonious models from the generalized approach.

A decision support system (DSS) model was built based on the Adaptive Bayesian Networks (ABN) methodology and a data mining engine [12]. Increasing the maximum ABN depth slightly increases a procedure's accuracy, but it significantly increases the computation build time. Using the robust Bayes classifier could increase the accuracy of outcome prediction for intensive-care patients up to 78-84%.

### B. Bayesian Network Theory

Bayesian networks (BNs), also known as belief networks (or Bayes nets for short), belong to the family of probabilistic graphical models (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables [13-15]. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics. BNs became extremely popular models in the last decade. They have been used for applications in various areas, such as machine learning, text mining, natural language processing, speech recognition, signal processing, bioinformatics, error-control codes, medical diagnosis, weather forecasting, and cellular networks [14].

One of the disadvantages of naïve Bayesian classifier may be sensitivity to correlated features. The work of P. Langley [17] gives detailed over view of using selective Bayesian classifiers. .The naïve Bayesian classifier perform classification and prediction based on the sub set of the attributes .So the selective Bayesian classifier is termed as the variant of the naïve method. These problems have been addressed by performing greedy search through the space of features. Lucas, gives a detailed picture of a model based clinical decision support system [18] .This model deals with the concept of formalism and the uses of it in decision making using Bayesian network for pneumonia by exploiting the causal and other relationships between domain variables. A DSS for ventilator associated pneumonia (VAP) in ICU has also been modeled [18]. This work uses the randomly selected sub set of the real data and the same is used for learning the probability distributions on the basis of the original probability distribution. Interestingly a drug prescription is prepared based on the diagnosis for pneumonia [18-22]. Lucas deals with the same problem using Boolean threshold functions for the diagnosis of VAP patient .This diagnosis helps it to predict optimized treatment for them who are distinguished by different number of colonizing pathogens.

Another promising pediatric diagnosis system for primary health centers of rural India using Bayesian network was proposed [19-20]. It provides an interaction mechanism with the healthcare professional at a rural site and delivers diagnosed disease and suggests treatment plans as per the supplied sign symptoms. Later a more automatic approach for medical decision support content was generated [24].This expert system for hospitalizations involving pregnancy, designed to have four stages: Feature selection, Converting Data into Vectors, learning and estimation of posterior probabilities. For the test set of real data the ROC curves are calculated [23]. The classification and prediction procedure of various medical datasets is handled. It uses the particle swam optimization (PSO) technique for arriving the best classification result followed by feature selection and extraction from the data set, Bayesian PSO hybrid classification [21]. The PSO optimization is done on non linear discriminant Coefficients. This hybrid system provides better classification accuracy over well known Hepatitis, Pima Indian, Thyroid and Appendicitis datasets over naïve Bayesian and nonlinear Bayes.

## IV. Neural Network Based Learning

Logistic regression and artificial neural networks are the models of choice in many medical data classification tasks. This section of the review summarizes the various neural network based

learning algorithms. One layer of hidden neurons is generally sufficient for classifying most data sets. The number of neurons in the hidden layer needs to be set empirically, e.g. by cross-validation or bootstrapping [25]. It is imperative to not over-fit the network during training; this can be achieved either by restricting the topology of the network (i.e., decreasing the number of nodes), by early stopping, or by using weight decay.

### A. Single Layered Neural Networks

A new approach to supervised learning in a single-layer linear feed forward neural network was discussed by D.Sanger. An optimality principle is proposed which is based upon preserving maximal information in the output units [26]. The neural network seems promising in hard tasks as in extraction of useful information from EEG data recorded on the scalp [27]. This is a multidimensional biosignal (ECG data) processing problem in which the signal and the noise have unknown statistics and in which the signal to noise ratio is very small. A particularly difficult problem is the measurement and extraction of neuroelectric patterns related to the preparatory set for human visuomotor performance.

Neural networks can also be used for the diagnosis of hepatobiliary disorders. Hayashia performs this diagnosis by combining the predictions of an ensemble of neural networks [28]. To improve the accuracy of the diagnosis, the second level networks are trained using the outputs of the first level networks as input data. The second level networks achieve an accuracy that is higher than that of the individual networks in the first level. Compared to the simple method which averages the outputs of the first level networks, the second level networks are also more accurate. Multiplicative weight updating algorithms such as Winnow [29] and Weighted Majority variants have been studied extensively in the theoretical machine learning literature, in which a collection of strong properties have been proven. These algorithms could be said to fall into the category of "learning simple things really well." In particular, when the concept being learned is appropriately simple, they have been proven to have exceptionally good behavior in the face of irrelevant features, noise, or a target function changing with time [29]. In particular, implementations using the single layered neural networks give a substantial improvement along a number of lines, including accuracy and speed, over the results of Mitchell et al. [30], which use a decision tree based approach.

### B. Multi Layered Neural Networks

A Multi Layer Perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network. MLP is a modification of the standard linear perceptron, which can distinguish data that is not linearly separable.

The global and local techniques can be applied to the training of neural network classifiers for solving medical diagnosis problems. A methodology is used which involves systematic and exhaustive evaluation of the classifier performance over a neural network architecture space and with respect to training depth for a particular problem [31]. In this methodology, the architecture space is defined over feed-forward, fully connected artificial neural networks (ANNs) which have been widely used in computer-aided decision support systems in medical domain, and for which two popular neural network training methods are explored: conventional back propagation (BP) and particle swarm optimization (PSO). Both training techniques are compared in terms of classification performance over three medical diagnosis problems (breast cancer, heart disease, and diabetes) from Pro-ben1 benchmark dataset and computational and architectural analysis are performed for an extensive assessment.

Differential diagnosis of multiple disorders is a challenging problem in clinical medicine. According to the divide-and-conquer principle, this problem can be handled more effectively through decomposing it into a number of simpler sub-problems, each solved separately. Vapnik demonstrates the advantages of this approach using abductive network classifiers on the 6-class standard dermatology dataset [32].Three problem decomposition scenarios are investigated, including class decomposition and two hierarchical approaches based on clinical practice and class separability properties. Two-stage classification schemes based on hierarchical decomposition boost the classification accuracy from 91% for the single-classifier monolithic approach to 99%, matching the theoretical upper limit reported in the literature for the accuracy of classifying the dataset. Such models are also simpler, achieving up to 47% reduction in the number of input variables required, thus reducing the cost and improving the convenience of performing the medical diagnostic tests required. Automatic selection of only relevant inputs by the simpler abductive network models synthesized provides greater insight into the diagnosis problem and the diagnostic value of various disease markers. The problem decomposition approach helps plan more efficient diagnostic tests and provides improved support for the decision-making process [32].

## V. Support Vector Machines

A Support Vector Machine (SVM) performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. Using a kernel function, SVMs are an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training.

In the parlance of SVM literature, a predictor variable is called an attribute actor e, and a transformed attribute that is used to define the hyperplane is called a feature. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one case (i.e., a row of predictor values) is called a vector. So the goal of SVM modeling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the support vectors. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high or

infinite dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, it often happens that in that space the sets to be discriminated are not linearly separable. For this reason it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. SVM schemes use a mapping into a larger space so that cross products may be computed easily in terms of the variables in the original space, making the computational load reasonable. The cross products in the larger space are defined in terms of a kernel function K(x,y) selected to suit the problem. The hyperplanes in the higher dimensional space are defined as the set of points whose inner product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters αi of images of feature vectors that occur in the data base. With this choice of a hyperplane the points x in the feature space that are mapped into the hyperplane are defined by the relation:

$$\sum_i \alpha_i K(x_i, x) = constant$$

If K(x,y) becomes small as y grows further from x, each element in the sum measures the degree of closeness of the test point x to the corresponding data base point xi. In this way the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. The support vector machines show best accuracy in binary class problems like heart beat / valve classification [34, 35, 36]. Samjin Choi [35] performed classification based on the difference in frequency ranges of VHD Signals and normal sounds. Here the Heart Sound signals are applied to a decomposition structure tree which has mean and standard deviation of the position indices as terminal nodes. In this study [35] only maximum value of WPE was considered to identify between normal and VHD sounds. Kernel based SVMs use a space called kernel space to which the extracted features will be mapped and then classification will be performed [37].This process of mapping into another space and then classifying the data based on the mapped features will help to improve the accuracy of the classification [38].

The least squares support vector machine (LS-SVM) is a least squares version of support vector machine . In this version one finds the solution by solving a set of linear equations instead of a convex quadratic programming (QP) for classical SVMs. Least Squares SVMs (LS-SVMs) classifiers, was proposed in Suykens and Vandewalle [39]. LS-SVM is a class of kernel based learning methods. Primary goals of the LS-SVM models are regression and classification. The LS –SVM is being used widely for medical diagnosis especially for heart diseases [40, 41]. The hybrid system for medical data classification helps to improve the accuracy [42].This work [42] uses a hybrid system of Discriminant Analysis and LS-SVM for diabetic's disease classification. Another article [43] also uses SVM with generalized discriminant analysis for thyroid disease diagnosis. Optimization of the classification result is also possible by using any optimization techniques. Genetic algorithm is a widely used [44].

## VI. Discussion

Supervised machine learning techniques are more suitable for medical data classification. Now, we present our conclusion about the use of each technique. Generally support vector machines and neural networks perform much better with continuous and multi-dimensions features. On the other hand for discrete /categorical features of medical data Logic Based Systems perform better. For neural networks and SVMs, a large sample size is required in order to achieve its maximum prediction accuracy whereas Naïve Bayes may need a relatively small dataset.

With problems that require diagonal partitioning, most decision tree algorithms cannot perform well. The division of the instance space is orthogonal to the axis of one variable and parallel to all other axes. Therefore, the resulting regions after partitioning are all hyper rectangles. The ANNs and the SVMs perform well when multicollinearity is present and a nonlinear relationship exists between the input and output features. But univariate decision trees are presumed to be quite fast at any rate, several orders of magnitude faster than neural networks and SVMs.

Finally the major metric of predication, the accuracy, is to be compared. Invariably SVMs outperform all the techniques in accuracy even for large datasets. Neural networks come next to SVMs achieving better prediction accuracy. Moreover the techniques of combining classifiers [43-45] can produce still more accuracy. Table 1 depicts the comparison of the various features of learning techniques.

Table 1: Comparison of Learning Algorithms

| | Decision Trees | Rule Based | Naive Bayes | Neural Network | SVM |
|---|---|---|---|---|---|
| Accuracy | ** | ** | * | *** | **** |
| Speed (learning) | *** | ** | **** | * | * |
| Speed (Classification) | *** | *** | *** | **** | **** |
| Handling Irrelevant attributes | *** | ** | ** | * | **** |
| Tolerance to redundant attributes | ** | ** | * | ** | *** |
| Over fitting prevention | ** | ** | *** | * | ** |

The kernel based SVMs deal the problem of over fitting far better than neural networks [37-38]. Both neural network and SVM are performing the classification at almost same speed. But both these techniques take more time for learning usually. This is because of the fact that they normally deal with large datsets. On the other hand naïve Bayes can deal with only small datasets. Moreover the SVMs are more tolerant to redundant and irrelevant attributes.

## VII. Conclusion

The main objective of this survey is to analyze the most imperative machine learning techniques and suggest the best suitable technique for medical data set classification. But, it cannot be concluded that one algorithm is always superior to

other. Instead it could be concluded under which conditions a particular method can significantly outperform other methods for a given problem.

There are many metrics available to compare the performance of the discussed methods. Since the application area considered here is clinical diagnosis, the main concern here is on accuracy of the techniques and also the time required for classification. In this point of view, it could be concluded that the SVMs can perform classification and prediction with highest accuracy for large medical data sets in comparison with all other methods and neural network also can perform classification with better accuracy next to SVMs.

## References


[1] Mario R. Guarracino, Salvatore Cuciniello, Davide Feminiano,Gerardo Toraldo,Panos M. Pardalos,"Current Classification Algorithms for BioMedical Applications:American Mathematical Society", CRM proceedings and Lecture Notes,Volume 45, 2008,

[2] Murthy, "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey", Data Mining and Knowledge Discovery 2: pp.345-389. 1998,

[3] Sotiris A Pavlopoulos, Antonis CH Stasis, Euripides N Loukis,"A decision tree – based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds, BioMedical Engineering OnLine". 2004

[4] Furnkranz, J, "Pruning algorithms for rule learning", Machine Learning 27: pp. 139-171. 1997.

[5] Furnkranz, J., Flach, P, "ROC 'n' Rule Learning - Towards a Better Understanding of Covering Algorithms", Machine Learning, Volume 58 (1), pp. 39 – 77. 2005

[6] Mei-Ling Huang, Hsin-Yi Chen, Jian-Cheng Lin, "Investigative Ophthalmology & Visual Science", Vol. 48, No. 1, pp. 244-250. January 2007.

[7] Ruchika Gupta, Alok Sharma, Sompal Singh, Amit K Dinda,"Rule based decision support system in the biopsy diagnosis of glomerular disease", JCP Online. June 2011,

[8] P. Angelov, X. Zhou, D. Filev, E. Lughofer,"Architectures for evolving fuzzy rule based classifiers", IEEE, pp.2050-2055, 2007

[9] K. Ashwin Kumar, Yashwardhan Singh, Sudip Sanyal, "Hybrid approach using case based reasoning and rule-based reasoning for domain independent clinical decision support in ICU, Science Direct", Expert system with Applications,pp. 65–71 2009

[10] Domingos, "On the Optimality of the Simple Bayesian Classifier Under zero-one loss", Machine Learning29: pp. 103-130. 1997

[11] Ranjit Abraham, Jay B. Simha, S. Sitharama Iyengar,"Effective Discretization and Hybrid feature selection using Naïve Bayesian classifier for Medical data mining", International Journal of Computational Intelligence Research, Vol.4, 2008

[12] D. Kopec , S. G. Shagas, D. Reinharth Tamang,"Development of a Clinical Pathway Analysis System with Adaptive Bayesian Nets and Data Mining Techniques".

[13] Friedman, N., Geiger, D.,Goldszmidt M,"Bayesian network classifiers. Machine Learning 29: pp. 131-163, 1997.

[14] Friedman, N.,Koller, D, "Being Bayesian About Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks", Machine Learning 50(1): pp. 95-125, 2003.

[15] Jensen, F, "An Introduction to Bayesian Networks Springer", 1996

[16] Kononenko, I, "Inductive and Bayesian learning in medical diagnosis", Applied Artificial Intelligence 7(4): pp. 317-337, 1993

[17] Pat Langley, Stephanie Sage,"Induction of selective Bayesian classifiers" , Proceeding of the Tenth conference on Uncertainty Artificial Intelligence, pp: 399-406. (1994)

[18] Peter Lucas, "Bayesian Networks in Medicine", A Model-based Approach to Medical Decision Making.

[19] J.Mago, Parvinder S. sandhu, Neeru Chawla,"Bayesian Network Based Intelligent Pediatric System", World Academy of Science, Engineering and Technology73, pp: 756-760. (2011)

[20] Carolina A.M. Schurink, Stefan Visscher, Peter J. F. Lucas, Henk J. van Leeuwen, Erik Buskens, ReinierG. Hoff, Andy I.M. Hoepelman, Marc J. M. Bonten, "A Bayesian decision Support System for diagnosing ventilator-associated pneumonia, Springer", Intensive Care Med 33: pp. 1379–1386, 2007

[21] S. Devi, A. K. Jagadev, S. Dehuri, R. Mall,"Knowledge Discovery from Bio-medical Data Using a Hybrid PSO/ Bayesian Classifier TECHNIA – International Journal of Computing Science and Communication Technologies", VOL. 2, NO. 1, July 2009

[22] Stefan Visscher, Peter J.F. Lucas, Carolina A.M. Schurink, Marc J.M. Bonten, "Modelling treatment effects in a clinical Bayesian network using Boolean threshold functions", Artificial Intelligence in Medicine 46,pp: 251—266, 2009

[23] Jeffrey Klann, Gunther Schadow, Stephen M. Downs, "A Method to Compute Treatment Suggestions from Local Order Entry Data", AMIA Symposium Proceeding, pp: 387-391. 2010

[24] Jyotirmay Gadewadikar, Ognjen Kuljaca, Kwabena Agyepong, Erol Sarigul, Yufeng Zheng, Ping Zhang, "Exploring Bayesian networks for medical decision support in breast cancer detection", African Journal of Mathematics and Computer Science Research Vol. 3(10), pp. 225-231, October 2010

[25] Stephan Dreiseitla,Lucila Ohno-Machadob, "Logistic regression and artificial neural network classification models: a methodology review",Journal of Biomedical Informatics 35,pp: 352–359, 2002

[26] Terence D. Sanger, "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network", Neural Networks, Vol. 2, pp: 459-473, 1989

[27] Nikola Masic,G. Pfurtscheller, "Neural network based classification of single-trial EEG data", Elsevier,Artificial Intelligence in Medicine 5 pp: 503-513, 1993

[28] Yoichi Hayashia, Rudy Setionob, "Combining neural network predictions for medical diagnosis Computers in Biology and Medicine", 32, pp: 237–246, 2002

[29] Littlestone, N., Warmuth, M., "The weighted majority algorithm", Information and Computation 108(2) pp: 212–261, 1994

[30] Mitchell, T., "Machine Learning. McGraw Hill", 1997.

[31] Turker Ince,Serkan Kiranyaz, Jenni Pulkkinen, Moncef Gabbouj," Evaluation of global and local training techniques over feed-forward neural network architecture spaces for

computer-aided medical diagnosis", Expert Systems with Applications 37, pp: 8450–8461, 2010.
[32] Radwan E. Abdel-Aal, Mona R.E. Abdel-Halim, Safa Abdel-Aal, "Improving the classification of multiple disorders with problem decomposition", Elsevier Journal of Biomedical Informatics 39, pp: 612–625, 2006
[33] Vapnik, V. "The Nature of Statistical Learning Theory. Springer Verlag", 1995
[34] Argyro Kampouraki, Christophoros Nikou, George Manis, "Robustness of Support Vector Machine-based Classification of Heart Rate Signals", Proceedings of the 28th IEEE, EMBS Annual International Conference,NewYork,USA,Aug 30-sep 3,2006, 1995
[35] Samjin Choi, "Detection of valvular heart disorders using wavelet packet decomposition and support vector machine, Elsevier", Expert Systems with Applications 35 2008, pp 1679-1687
[36] Ilias Maglogiannis, Euripidis Loukis, Elias Zafiropoulos,Antonis Stasis, "Support vector machine based identification of heart valve diseases using heart sounds", Elsevier,Computer Methods and Programs in Biomedicine ,95, 2009, pp. 47-61
[37] Domenico Conforti, Rosita Guido, "kernel Based Support Vector Machine via semidynamic programming :Application to Medical Diagnosis", Elsevie, Computer and Operations Research, 37, 2010, pp 1389-1394.
[38] Der-Chiang Li, Chiao-Wen Liu, "A class possibility based kernel to increase classification accuracy for small data sets using support vector machines, Elsevier", Expert Systems with Applications,37, 2010 ,pp. 3104-3110.
[39] Suykens J.A.K., Vandewalle J., "Least squares support vector machine classifiers", Neural Processing Letters, vol. 9, no. 3, Jun. 1999, pp. 293-300
[40] Kemal Polat, Bayram Akdemir, Salih G ne, "Computer aided diagnosis of ECG data on the least square support vector machine, Elsevier", Digital Signal Processing,18, 2008, pp. 25-32.
[41] Davut Hanbay, "An expert system based on least square support vector machines for diagnosis of the valvular heart disease,Elsevier", ExpertSystems with Applications ,36, 2009, pp. 4232-4238.
[42] Kemal Polat, Salih Gu, nes, Ahmet Arslan, "A cascade learning system for classification of diabetesdisease: Generalized Discriminant Analysis and LeastSquare Support Vector Machine", Elsevier ,Expert Systems and Applications, 34, 2008, pp. 482-487.
[43] Esin Dogantekin, Akif Dogantekin, "Derya Avci An expert system based on Generalized Discriminant Analysis and Wavelet Support Vector Machine for diagnosis of thyroid diseases", Elsevier, Expert Systems with Applications,38, 2011, pp. 146-150.
[44] E. Avci, "A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier", Elsevier, Expert Systems with Applications, 36, 2009, pp. 10618-10626.
[45] Chang-Ying Ma, Sheng-Yong Yang, Hui Zhang, Ming-Li Xiang, Qi Huang, Yu-Quan Wei, "Prediction models of human plasma protein binding rate and oral bioavailability derived by using GA–CG–SVM method", Elsevier, Journal of pharmaceutical and Bio medical analysis, 47, 2008, pp. 677-682.
[46] D.Shanthi, Dr.G.Sahoo, Dr.N.Saravanan, "Decision Tree Classifiers to Determine the Patient's Postoperative Recovery Decision", International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume(1), Issue(4).
[47] Clark P., Niblett, T., "The CN2 Induction Algorithm Machine Learning", 1989, 3(4): pp. 261-283.

S. Anto, received his B.E. degree in Electrical and Electronics Engineering from Noorul Islam College of Engineering, Thuckaly, India in 1999, the M.E. degree in Computer Science and Engineering from Annamalai University, Chidambaram, India, in 2005 and pursuing PhD in Artificial Intelligence from Anna University of Technology Coimbatore. He was a lecturer with Department of Electrical and Electronics Engineering, V.L.B. Janakiammal College of Engineering and Technology, in 2001. He was a lecturer, senior lecturer with Department of Computer Science and Engineering, V.L.B. Janakiammal College of Engineering and Technology, in 2005 and 2008 respectively. His research interests include Artificial Intelligence, Pattern recognition and Machine learning. At present he is an Assistant professor with Department of Computer Science and Engineering, V.L.B. Janakiammal College of Engineering and Technology, Coimbatore, India.

Dr. S Chandramathi, received her B.E. degree in Electronics and communication Engineering from PSG College of Technology ,Coimbatore ,India in 1981 ,the M.E. degree in Applied Electronics From CIT,Coimbatore ,India ,in 1987 and PhD in Communication Networks from School of ECE, Anna University Chennai, India. She was a project Associate in IIT Madras; Chennai, India in 1983 .She was a senior lecturer, selection grade lecturer, Assistant Professor, Professor with Department of Electronics and Communication Engineering, Pondicherry Engineering College in 1987, 1995, 1999 and 2008 respectively. Her research interests include Computer Communication, Wireless Communication and Digital Communication. At present she is Professor and Head with Department of Electronics and Communication Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, India.