International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 13 (2015)
 http://www.ripublication.com

Machine Learning For Medical Decision Support Systems (MDSS): A review

Sunila Godara[1], Rishipal Singh[2]

Deptt. of CSE, GJUS&T, Hisar, India
sunilagodara@gmail.com[1], sandeep1642@gmail.com[2]

Abstract
It has been claimed that Technology will replace 80% of what physician do today. Medical science industry has huge amount of data, but unfortunately most of this data is not mined to find out hidden information in data. Advanced Machine Learning techniques can be used to discover hidden pattern in data. Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain. This paper formalizes Medical Decision Support System (MDSS) and describes the methods that have been developed within machine learning area for addressing these problems. These machine learning techniques comprise Clustering, K-Nearest Neighbors KNN, Decision Tree, Multi-layer perceptron (MLP), Radial Basis Function (RBF) neural network, Support Vector Machine (SVM). In this paper Core contributions by researchers in this area which had discussed or made use of machine learning ideas is also presented

Keywords –MDSS, clustering, K-nearest neighbors (KNN), decision tree, artificial neural networks, radial basis function (RBF) neural network and support vector machine.

I. INTRODUCTION

Diagnosis of diseases is an important and difficult task in medicine. Detecting a disease from several factors or symptoms is a many-layered problem that also may lead to false assumptions with often unpredictable effects. In this way, the endeavor of utilizing the learning and experience of numerous specialists gathered in databases to backing the conclusion procedure appears to be sensible [1][2]. Symptomatic choice backing is still all that much a craftsmanship for doctors in their practices today because of absence of quantitative tools. A therapeutic analytic MDSS is a computer program that contains all significant restorative area information around a certain medicinal space and produces a differential conclusion on the premise of individual patient discoveries. A therapeutic demonstrative MDSS may be amazingly helpful in light of the fact that it has the capacity enhance the availability of specialists,information and patient data, bringing about quality change of the analytic procedure, increment of productivity and decrease of expenses. Machine learning techniques have great potential to become core part of MDSS[5].

There are two fundamental sorts of MDSS[2]:

Knowledge Based

Most CDSS consist of three parts, the knowledge base, inference engine, and mechanism to communicate. The knowledge base contains the rules and associations of compiled data which most often take the form of IF-THEN rules. If this was a system for determining drug interactions, then a rule might be that IF drug X is taken AND drug Y is taken THEN alert user[1][2].

Non-Knowledge-Based CDSS

CDSS's that do not use a knowledge base use a form of artificial intelligence called machine learning, which allow computers to learn from past experiences and/or find patterns in clinical data.

In order to understand what processes are performed in such a system it is essential to know its architecture (Figure 1). It consists of three main parts: an input component, a processing component and an output component [3]. The vital element of the system is a Decision Maker which utilizes computer technology to access domain knowledge. The client has control over results of the framework which is helpful in get ready choice choices. They are then assessed and the best one is picked. Thusly new learning is made which can be utilized as an extra info to the framework later on.

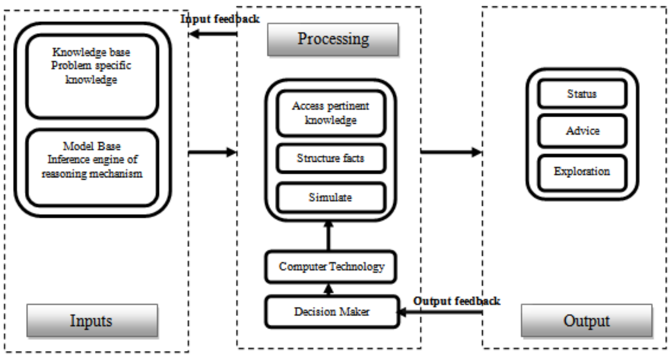

Figure1:MDSS[3]

Examples of MDSS :

De Dombal is a decision support tool that uses naive Bayesian approach to provide automated reasoning under uncertainty for diagnosis of acute abdominal pain and the need for surgery. It was developed at the Leeds University in 1972.

MYCIN is a rule-based expert system developed at Stanford University in 1976.It was used to diagnose and recommend treatment for certain blood infections and other infectious diseases.

The DXplain is a Medical Decision Support System which was developed in Massachusetts General Hospital in 1987. At the beginning the DXplain contained information about 2000 diseases and 4700 symptoms. This information was interconnected by 65000 associations. Nowadays the system has been improved. Today the DXplain stores information about 4900 symptoms and 2200 unique diseases. The number of associations among them is estimated to be about 23000.

The Early Referrals Application (ERA) is one of the most up to date and most encouraging Medical Decision Support Systems [75]. This arrangement is devoted to identification of distinctive sorts of growths in their initial stage. The application has been produced in Great Britain by GP's connected with the college healing facilities of Leicester NHS Trust subsequent to 2001.

TherapyEdge HIV is Web-enabled decision support system for the treatment of HIV infection was developed in 2001.

One of the most popular and advanced Medical Decision Support System is called HELP. It is an information based doctor's facility data framework. The framework is furnished with a decision support component. It helps the clinicians in deciphering restorative data, diagnosing patients, keeping up clinical conventions and different errands. The advancement of medical data frameworks and processing innovation brought about a change of the framework. In 2003 another form was discharged, called HELP II[32].

First production version of the CRS was from 2000 to 2002. The implementation took place in February 2002 in the Medical Ambulatory Care Clinic (MACC) at theWestern Pennsylvania Hospital. TheMACC is a primary care clinic that serves as a rotation site for the hospital's residents. 2004–present A new, web-enabled version of the CRS was developed from 2004 to 2005,based on the research findings and experiences learned from the earlier MACC implementation. The new system was installed in August 2005 in theWest Penn Medical Associates clinic, which replaced the MACC in the Western Pennsylvania Hospital during an organization restructuring.

Simul Consult, which began in the area of neurogenetics. By the end of 2010 it covered ~2,600 diseases in neurology and genetics, or roughly 25% of known diagnoses. SimulConsult can add a disease with less than a total of four hours of clinician time. It is widely used by pediatric neurologists today in the US and in 85 countries around the world [32].

This paper is organized as follows: A brief overview of Machine Learning Techniques will be given in section 2. The Core contributions by researchers in MDSS using Machine Learning this area is presented in section 3. Section 4 contains conclusion and future scope of our study.

2. MACHINE LEARNING TECHNIQUES FOR MDSS

In this section we will discuss various Machine Learning Techniques.

2.1 CLUSTERING

Cluster analysis of gene expression data has proved to be a helpful tool for identifying co-expressed genes. DNA microarrays are emerged as the leading technology to measure gene expression levels primarily, because of their high throughput. Results from these experiments are usually presented in the form of a data matrix in which rows represent genes and columns represent conditions [9]. Each entry in the matrix is a measure of the expression level of a particular gene under a specific condition. Analysis of these data sets reveals genes of unknown functions and the discovery of functional relationships between genes. The most popular clustering algorithms in microarray gene expression analysis are Hierarchical clustering and K-Means clustering [9.] K-means is a prototype based clustering technique which performs one level partition of the data objects. In this we first choose k initial centroids, where k represents the number of clusters desired. Each point is then assigned to the closest centroid, and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is then updated based on the points assigned to the cluster. This assignment and update steps will continue until no point changes in cluster, equivalently, or centroids remain the same [9]. Below Fig 2[9] represents flow chart of K-Means algorithm.

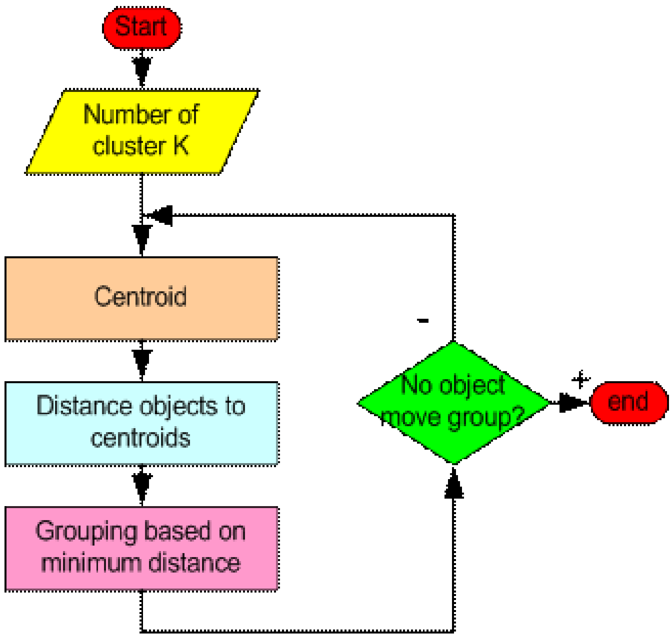

Fig 2. Flow chart of K-Means algorithm

The basic steps of k-means clustering are :

1. Determine the centeroid coordinates.
2. Determine the distance of each object to the centeroids.

3. Group the objects based on minimum distance.
4. Update the centeroids.

For each cluster purity is measured as follows:

$$entropy(D_i) = -\sum_{j=1}^{k} \Pr_i(c_j)\log_2 \Pr_i(c_j),$$

Where $\Pr_i(c_j)$ is proportion of class c_j in cluster I or D_i..Total entropy of whole clustering is given below:

$$entropy_{total}(D) = \sum_{i=1}^{k} \frac{|D_i|}{|D|} \times entropy(D_i)$$

2.2 K-Nearest Neighbors (KNN)

K-nearest neighbors is a supervised learning algorithm used to classify the new sample data based on the similarity measures. K -Nearest Neighbors consists the training data sets and when new case is given as input classification take place on those stored training data sets. Where the output of the process depends on whether KNN used for classification or regression.

KNN in Classification: - When the KNN process take place in the classification the input data is classified based on the majority voting of the neighbors of input data. When K=1 the input case is assigned directly to the single nearest neighbor. When the value of k increase the class to which input case is assigned can change and different result can take place [5][35].

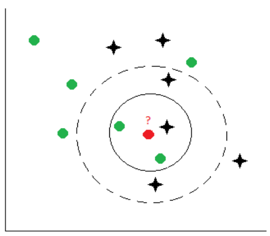

Fig. 3. KNN Used for Classification

Process of KNN can be easily explained by above diagram.In Fig3[9] red input case needs to be classified. Then according to KNN process if K is taken 1 then the data will be assigned to class of star which is nearest. When the neighbors are increased or value of k is increased to 3 the class of data becomes to green circle. And when it comes to k=5 then class change to star but the process gain the good accuracy when k take place between 3-10 [5].

KNN in Regression: - When KNN process in regression takes place then the output value is the average value of k-nearest neighbors. If the K is taken 1 in regression the outcome contains the value of nearest neighbors. Example if there is an input case(X) to process and its nearest neighbor's value is 2Y. When the k is taken 1 in that case

X = 2Y

When there is another second nearest neighbor with value 2Z and K is taken to in that case

X = (2Y+2Z)/2

Euclidean distance or city block distance can be used as distance measures Euclidean distance D can be calculated like

$$\mathbf{D} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

While KNN take place for the predictions then average of the k nearest neighbor is the output. As it compared to regression KNN Approach which uses voting of nearest neighbors.

2.3 DECISION TREE

Decision trees are intense order calculations. As the name infers, this procedure recursively isolates perceptions in branches to develop a tree with the end goal of enhancing the expectation precision. In doing as such, they utilize numerical calculations to distinguish a variable and relating limit for the variable that parts the info perception into two or more subgroups. This stride is rehashed at every leaf hub until the complete tree is developed. The target of the part calculation is to locate a variable-edge match that expands the homogeneity of the subsequent two or more subgroups of tests. Below Fig 4[4]. demonstrates an illustration of choice tree on patient determination. Here non-terminal hubs speak to tests on one or more traits and terminal hubs reflect choice results. Choice tree sums up taking after information: If a patient has swollen organs, the determination is strep throat. In the event that a patient does not have swollen organs and has fever, the determination is cold. In the event that a patient does not have swollen organs and does not have fever, the finding is allergy.

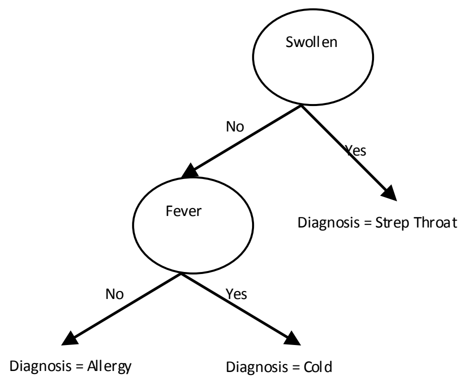

Fig 4: Decision Tree

When the construction of decision tree take place the attribute having largest information gain is selected as splitting attribute[8]. Let see how the information gain is calculated when node N represents the tuple of partitions then the entropy of D is calculated as:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i),$$

Then we need to calculate the InfoA (D) is the expected information required to classify a tuple from D.

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j).$$

Now the information gain can be achived by the above two difference like

$$Gain(A) = Info(D) - Info_A(D).$$

2.4 Multi-layer perceptron (MLP)

ANNs are analytic techniques modeled after the processes of learning in the cognitive system and the neurological functions of the brain and capable of predicting new observations from other observations. One of popular ANN architecture is called multi-layer perceptron (MLP) with back-propagation (a supervised learning algorithm). Fig 5[17] shows MLP feed forward Neural Network Where Input layer provide the functioning of input the data or values and from where data is feed forwarded to the next hidden layer. Hidden layer accept the data from the Input layer along the path from which they are connected. Hidden layer process the data There may be a number of hidden layers and a number of neurons in each layer for solving the particular task. From where, data after processing is fed forward to the output layer. After processing, output is matched to the target. If the match is not found the weight is adjusted at each processing element and process is repeated until there is a match or error reduces to the preferred limit.

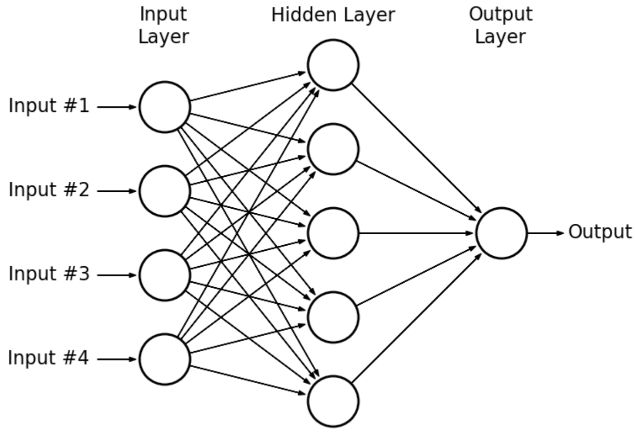

Fig5:MLP

The back-propagation algorithm consists of four step [32]:

1. Compute how quick the error changes with respect to value of output unit. This error derivative (EA) is the difference between the actual and the desired value.

$$EA_j = \frac{\partial E}{\partial yj} = y_j - dj$$

2. Compute how quick the error changes as the total input received by an output unit. This quantity (EI) is the respond from step 1 multiplied by the rate at which the output of a unit changes as its total input is changed.
3. Compute how quick the error changes as a weight on the connection into an output unit is changed. This quantity (EW) is the answer from step 2 multiplied by the value of the unit from which the connection emanates.

$$EI_j = \frac{\partial E}{\partial xJ} = \frac{\partial E}{\partial yj} \times \frac{\partial yj}{\partial xj} = EA_j y_j (1-y_j)$$

$$EW_{ij} = \frac{\partial E}{\partial Wij} = \frac{\partial E}{\partial xJ} \times \frac{\partial xj}{\partial Wij} = EI_j y_i$$

4. Compute how quick the error changes as the activity of a unit in the previous layer is changed. This vital step allows back propagation to be applied to multilayer networks. When the value of a unit in the previous layer changes, it affects the values of all the output units to which it is connected. So to calculate the overall effect on the error, we add together all these separate effects on output units. But each effect is easy to compute. It is the answer in step 2 multiplied by the weight on the connection to that output unit.

$$EA_i = \frac{\partial E}{\partial yi} = \sum_j \frac{\partial E}{\partial xj} \times \frac{\partial xj}{\partial yi} = \sum_j EIjWij$$

By using steps 2 and 4, we can alter the EAs of one layer of units into EAs for the previous layer. This procedure can be continual to get the EAs for as many previous layers as desired. Once we know the EA of a unit, we can use steps 2 and 3 to decide the EWs on its incoming connections.

2.5 RBF NEURAL NETWORK

A Radial Basis Function (RBF) neural network has neurons in the hidden layer contain Gaussian transfer functions whose outputs are inversely proportional to the distance from the center of the neuron [5][6]. The essential idea is that a predicted target value of an item is expected to be about the same as other items that have close values of the predictor variables RBF network positions one or more neurons in the space described by the predictor variables. This space has as many dimensions as there are predictor variables. The Euclidean distance is calculated from the point being evaluated to the center of each neuron, and a radial basis function (RBF) (also called a kernel function) is applied to the distance to compute the weight for each neuron. The radial basis function is so called because the radius or distance is the argument to the function[10][12].

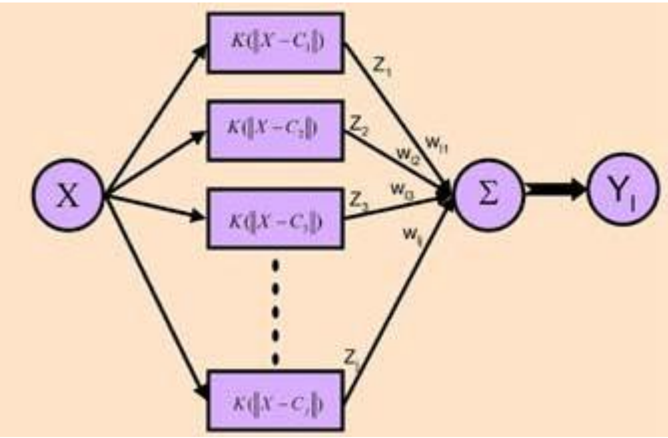

Fig 6: RBF Architecture

Fig 6[35] shows architecture of RBF neural network. The radial-basis functions technique recommends designing of interpolation functions F of the subsequent form[12]:

$$F(\mathbf{x}) = \Sigma_{i=1}^{N} w_i \, \varphi(\|\mathbf{x} - \mathbf{x}_i\|)$$

where $\varphi(\ \|\mathbf{x} - \mathbf{x}_i\|\ )$ is a set of nonlinear radial-basis functions, xi are the centers of these functions, and ||.|| is the Euclidean norm. The unknown weights can be found by solving the following linear matrix equation:

$$\mathbf{w} = (\,\mathbf{\Phi}^T\mathbf{\Phi}\,)^{-1}\,\mathbf{\Phi}^T\,\mathbf{y}$$

where $\mathbf{w}$ is the weight vector, $\mathbf{y}$ is the response vector, and $\mathbf{\Phi}$ is the design matrix

$$\mathbf{\Phi} = \{\ \varphi_{ei} \mid \varphi_{ei} = \varphi(\|\mathbf{x}_e - \mathbf{x}_i\|),\ (e,i)=1,2,...,N\ \}$$

$\mathbf{\Phi}$ is the n × n design matrix of basis functions $\varphi = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2 / 2\sigma_i^2)$

$\mathbf{\Phi} = \varphi_{11}\ \varphi_{12} \ldots \varphi_{1n}$

$\varphi_{21}\ \varphi_{22} \ldots \varphi_{2n}$

...

$\varphi_{N1}\ \varphi_{N2} \ldots \varphi_{Nn}$

If there is more than one predictor variable, then the RBF function has as many dimensions as there are variables. The following picture demonstrates three neurons in a space with two predictor variables, X and Y. Z is the value coming out of the RBF functions. Fig.7[35] shows RBF having three neurons in a space with two predictor variables. Fig 8[35] shows weighted sum of RBF Functions.

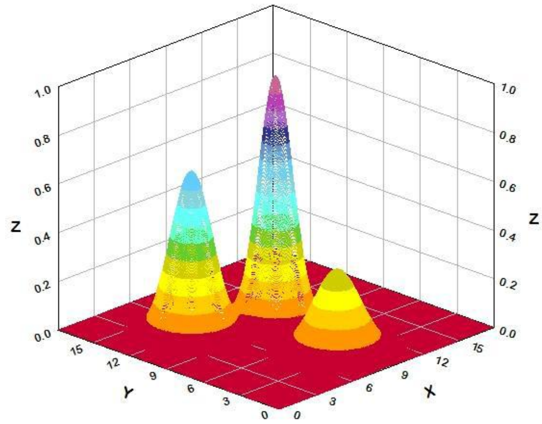

Fig 7: RBF having three neurons in a space with two predictor variables

The best predicted value for the new point is originated by summing the output values of the RBF functions multiplied by weights calculated for each neuron.

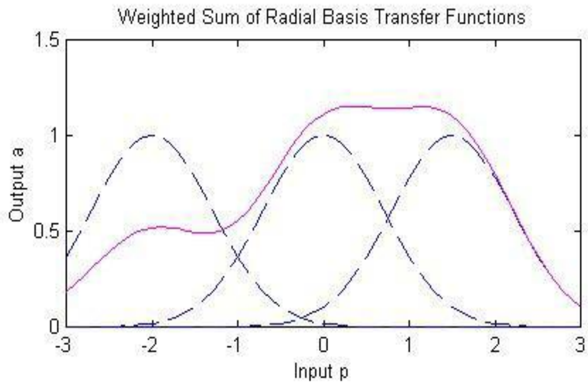

Fig 8: Weighted sum of RBF Functions

2.6 SUPPORT VECTOR MACHINE

SVM is technique for classification of both linear and non-linear data. It uses a non-linear mapping to transform the original training data into a higher dimension. A Support Vector Machine (SVM) [5][35] achieves classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVM models are closely associated to neural networks. SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural networks.

Using a kernel function, SVM's are an substitute training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are initiated by solving a quadratic programming problem with linear constraints[13]. So the goal of SVM modeling is to find the optimal hyper plane 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 close to the hyper plane are the support vectors. The Fig. 9[35] below presents general idea of the SVM process.

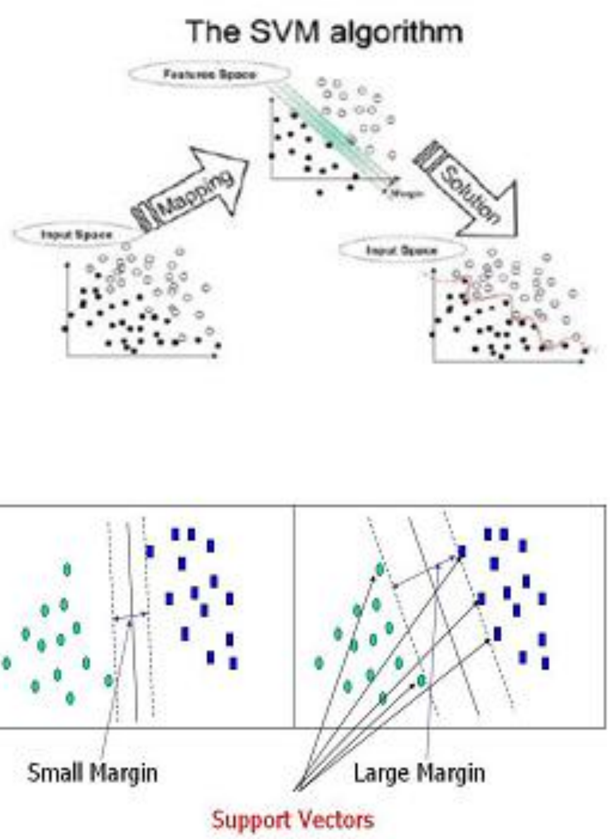

Fig 9: SVM topology

To construct an optimal hyper plane, SVM employs an iterative training algorithm, which is used to minimize an error function. According to the type of the error function, SVM models can be classified into four separate groups:

SVM Type 1[34]:
In this type of SVM, training involves the minimization of the error function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

Subject to the constraints:

$$y_i \left(w^T \phi(x_i) + b \right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \ldots, N$$

Where C is the capacity constant, w is the vector of coefficients, b is a constant, and ξ_i represents parameters for using nonseparable data (inputs). The index i, labels the N training cases. $y \in \pm 1$, signifies the class labels and xi represents the independent variables. The kernel ϕ is used to transform data from the input to the feature space.

SVM Type 1[34]:
The SVM Type 2 model minimizes the error function:

$$\frac{1}{2} w^T w - \nu\rho + \frac{1}{N} \sum_{i=1}^{N} \xi_i$$

Subject to the constraints:

$$y_i \left(w^T \phi(x_i) + b \right) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, \ldots, N \text{ and } \rho \geq 0$$

Regression SVM Type 1[34]:
In this type of SVM the error function is:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i + C \sum_{i=1}^{N} \xi^{\bullet}{}_i$$

Which can be minimized subject to:

$$\begin{aligned} & w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^{\bullet} \\ & y_i - w^T \phi(x_i) - b_i \leq \varepsilon + \xi_i \\ & \xi_i, \xi_i^{\bullet} \geq 0, i = 1, \ldots, N \end{aligned}$$

Regression SVM Type 2[34]:
For this SVM model, the error function is given by:

$$\frac{1}{2} w^T w - C \left(\nu\varepsilon + \frac{1}{N} \sum_{i=1}^{N} \left(\xi_i + \xi_i^{\bullet} \right) \right)$$

Which can be minimized subject to:

$$\begin{aligned} & \left(w^T \phi(x_i) + b \right) - y_i \leq \varepsilon + \xi_i \\ & y_i - \left(w^T \phi(x_i) + b_i \right) \leq \varepsilon + \xi^{\bullet}{}_i \\ & \xi_i, \xi^{\bullet}{}_i \geq 0, i = 1, \ldots, N, \varepsilon \geq 0 \end{aligned}$$

The RBF is by far the most accepted choice of kernel types used in Support Vector Machines. This is due to their localized and finite responses across the entire range of the real x-axis.

3 MDSS USING MACHINE LEARNING TECHNIQUES

Many decision support systems have been proposed by researchers to assist physicians in effectively making decisions for diagnosing diseases. The decision support system employing data mining and artificial intelligence techniques is one of the leading research areas. In this section, we discuss selected important contributions from the existing literature.

The diagnosis of hepatitis disease was conducted with a machine learning system by Kemal Polat et al. [15] proposed machine learning approach having three phase. The first phase, the feature number of hepatitis disease dataset was reduced to 10 from 19 in the feature selection (FS) subprogram by means of C 4.5 decision tree algorithm. In second phase, hepatitis disease dataset was normalized in the range of [0, 1] and weighted with fuzzy weighted pre-processing. In third phase weighted input values obtained from fuzzy weighted pre-processing was classified by using AIRS classifier system. Dataset used were collected from the UCI machine learning database. The obtained classification accuracy of the system was 94.12%. Mohammad Taha et al.[11] presented a prototype model for the breast cancer as well as heart disease prediction using C4.5, C5.0 techniques. Total eight rules have been generated by using C4.5 and C5.0 from cancer data set after pruning at the Confidence level 50. Running the C5.0 on heart disease and on breast cancer data sets seven rules have been generated. It was concluded that C5.0 handles missing values more efficiently

A medical diagnosis decision system (MDSS) based on SVM has been established to diagnose 4 types of acid-base disturbance by Lei Guo et al. [16]. SVM was initially developed for two-class classification. It was unmitigated to solve multi-class classification problem named hierarchical SVM with clustering algorithm based on stepwise decomposition. Compared with other classical classification techniques, SVM not only has more solid theoretical basis, it has better generalization ability as the experiment showed. A multi-layer perceptron based decision support system is developed by Hongmei Yana et al. [17] to support the diagnosis of heart diseases. The input layer of the system contains 40 input variables, categorized into four groups and then encoded using the proposed coding schemes. The numbers of nodes in the hidden layer were calculated through a cascade learning process. Each of the 5 nodes in the output layer corresponds to one heart disease of interest. In their system, the missing data of a patient are handled using the substituting mean method. Furthermore, an improved back propagation algorithm was used to train the system. A total of 352 medical records collected from the patients suffering from five heart diseases have been used to train and test the system., Three assessment methods, cross validation, holdout and bootstrapping, were applied to assess the generalization of the system. The results showed that the MLP-based decision support system can achieve very high diagnosis accuracy (>90%) and comparably small intervals (<5%), proving its utility in support of clinic decision process of heart diseases. Niti Guru et al. [18] have used the Neural Network for prediction of Heart disease, Proposed work recommended a supervised network for diagnosis of heart diseases and trained it using Back Propagation algorithm. The System was trained for 78 patients' records. On the origin of the trained data, when unknown data was entered by Doctor, the system finds list of possible diseases from which patient can suffer. Error made by human being can be avoided in the system. Hence the system was more reliable and helps the doctor to take correct decision. P. Jeatrakul et al. [19] presented an assessment of neural network techniques for binary classification problems. The classification performance obtained by five different types of neural networks for comparison are Back Propagation Neural Network (BPNN), Radial Basis Function Neural Network (RBFNN), General Regression Neural Network (GRNN), Probabilistic Neural Network (PNN), and Complementary Neural Network (CMTNN). The comparison is done based on three yardstick data sets achieved from UCI machine learning repository. The results explained that CMTNN typically provide better classification results applied to binary classification problems. Resul Das et al. [20] presented a methodology which uses SAS base software 9.1.3 for diagnosing of the heart disease. A neural networks ensemble method was in the centre of the proposed system. This ensemble based methods creates new models by combining the posterior probabilities or the predicted values from multiple predecessor models. So, more efficient models can be created. Experiments were carried out with the proposed method using SAS tool and obtained 89.01% classification accuracy on the data taken from Cleveland heart disease database. Experiment achieved 80.95% sensitivity and 95.91% specificity for heart disease diagnosis.

Francisco Fernández-Navarroa et al. [21] described a hybrid approach, Hybrid Algorithm (HA), which combines evolutionary and gradient based learning methods to approximate the architecture, weights and node topology of GRBFNN classifiers. The feasibility and benefits of the approach were established by means of six gene microarray classification problems taken from bioinformatics and biomedical domains. Three filters were applied: Fast Correlation-Based Filter (FCBF), Best Incremental Ranked Subset (BIRS), and Best Agglomerative Ranked Subset (BARS); this was done in order to recognize salient expression genes from among the thousands of genes in microarray data that can directly give the class membership of each pattern. After different gene subsets were obtained, the proposed methodology was performed using the selected gene subsets as new input variables. The results established that the GRBFNN classifier leads to a promising improvement in accuracy. Orhan Er et al. [23] described a comparative chest diseases diagnosis realized by multilayer, probabilistic, learning vector quantization, and generalized regression neural network. Results were prepared by using patient's epicrisis reports from a chest diseases hospital's database. O. Yavuz et al. [24]. presented, R5X4 type of HIV viruses classification using Auto Regressive (AR) model through Artificial Neural Networks (ANNs). The statistical data of R5X4, R5 and X4 viruses was analyzed by using signal processing methods and ANNs. Accessible residues of these virus sequences were obtained and modeled by AR model since the dimension of residues is large and different from each other. Finally the pre-processed data was used to evolve various ANN structures for determining R5X4 viruses. ROC analysis was applied to ANNs to show their real enhancement.

Seena Mary Augusty et al. [22] presented various methods applicable at data level as well as ensemble methods at algorithmic level. Cluster based under sampling, over sampling along with data based methods were evaluated under Data level. Ensemble classifiers were evaluated at the algorithmic level. Unstable base classifiers such as SVM and ANN can be employed for ensemble classifiers such as Bagging, Ad boost, Decorate and Random forest can improve the ensemble classification in dealing with imbalanced data problem. Rahmat Zolfaghair et al. [28] presented an approach to predict the presence of diabetes based on ensemble of SVM and BP NN. The predictive accuracy was 88.04 which was the best accuracy and it was very capable classification system. Paul Mangiameli et al. [30] compared 23 single models and bootstrap aggregating (i.e., bagging) models for their predictive abilities across five diverse medical data sets. Ensembles were more accurate than single models in their predictive ability. The best ensemble model achieved an error level significantly lower than the error of the best single model for four of the five medical applications analyzed. The magnitude of the error reduction ranged from 6.4% to 17.5%. Also, when designing an ensemble for an MDSS, the decision to diversify the model selection should be conducted by the relationship between model instability and generalization error for the population of models under concern

Guang-Bin et al. [25]. This paper presented extreme learning machine (ELM) as emergent technology which overcomes some challenges faced by other techniques.ELM has newly attracted the attention researchers. ELM works for generalized single-hidden layer feed forward networks (SLFNs). The essence of ELM is that the hidden layer of SLFNs need not be tuned. Compared with those traditional computational intelligence techniques, ELM provides better generalization performance at a much faster learning speed and with least human intervenes. Survey on ELM and its variants, especially on (1) batch learning mode of ELM, (2) fully complex ELM, (3) online sequential ELM, (4) incremental ELM, and (5) ensemble of ELM was presented. Jae-Hong Eom et al. [26] proposed a classifier ensemble-based method for supporting the diagnosis of cardiovascular disease (CVD) based on aptamer chips. This AptaCDSS-E system overcomes conventional performance limitations by utilizing ensembles of different classifiers. For CVD diagnosis, system combines a set of four different classifiers with ensembles. Support vector machines,neural networks, Decision trees and Bayesian networks are adopted. The experimental results showed that system achieved high diagnosis accuracy (>94%) and comparably small prediction difference intervals (<6%), proving its usefulness in the clinical decision process of disease diagnosis.

Yuehui Chen et al. [27]. presented hierarchical RBF network was to detect the breast cancer. For developing a hierarchical RBF network model, Extended Compact Genetic Programming (ECGP), a tree-structure based evolutionary algorithm and the Differential Evolution (DE) were used. An optimal detection mode was constructedl. The performance of proposed method was then compared with Flexible Neural Tree (FNT), Neural Network (NN), and RBF Neural Network (RBF-NN) by using the same breast cancer data set. Simulation results showed that the obtained hierarchical RBF network model has a fewer number of variables with reduced number of input features and with the high detection accuracy.

Manaswini Pradhan et al.[29] presented an Artificial Neural Network (ANN) based classification model The GA was used for optimally finding out the number of neurons in the single hidden layered model.For training and testing 10-fold cross validation method was adopted for Pima Indian Diabetes.For Pima dataset the ANN gives the best accuracy with 5 neurons in the hidden layer. Bestaccuracy being 72% with average accuracy of 72.2%. The designed model was compared with the Functional Link ANN (FLANN) and several classification systems like NN (nearest neighbor), kNN(k-nearest neighbor), BSS(nearest neighbor with backward sequential selection of feature, MFS1(multiple feature subset), MFS2(multiple feature subset) for Data classification accuracies.

Rajiv Gandhi et al.[35] constructed classification rules using the Particle Swarm Optimization Algorithm for breast cancer datasets. Feature subset selection as a pre-processing step was used which learns fuzzy rules bases using GA implementing the Pittsburgh approach. It was used to produce a smaller fuzzy rule bases system with higher accuracy. The resulted datasets after feature selection were used for classification using particle swarm optimization algorithm. The rules developed were with good accuracy defining the essential attributes effectively.

4 CONCLUSION

Our aim was to produce a critical review of the key ideas, rather than a simple list of all publications which had discussed those ideas.Well-known machine learning techniques such as Clustering,KNN, Decision Tree, Multi-layer perceptron (MLP), Radial Basis Function (RBF) neural network, Support Vector Machine (SVM) were discussed. Significant contributions by researchers in this area was discussed. Therefore, a more realistic goal is to build an intelligent machine learning technique as an effective Medical Decision Support System (MDSS). The role of such a data analysis tool is not to replace human experts, but only to assist human experts to make decisions more reliable.

5 REFERENCES

1. Berner, Eta S., ed. Clinical Decision Support Systems. New York, NY: Springer, 2007.
2. Finlay PN, "Introducing decision support systems", Cambridge, MA: Blackwell Publishers, 1994.
3. Mora M., Forgionne G.A., Jupta J., Decision Making Support Systems: achievements, trends and challenges for the next decade. Idea-Group: Hershey, P.A, 2002.
4. Z. Yao, P.L., L. Lei, and J. Yin, "R-C4.5 Decision tree model and its applications to health care dataset", in Proc of International Conference on Services Systems and Services Management, pp. 1099-1103, 2005.
5. Herron P., "Machine Learning for Medical Decision Support: Evaluating Diagnostic Performance of Machine Learning Classification Algorithms", INLS 110, Data Mining, 2004.
6. I. Colombet, A. Ruelland, G. Chatellier, F. Gueyffier, "Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression" in Proc of AMIA Symp, pp. 156-160, 2000.
7. E. Avci, I.Turkoglu, "An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases", in Proc of Expert Systems with Application, Vol. 2, pp. 2873-2878, 2009.
8. Imran Kurt, Mevlut Ture, A. Turhan Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease", in Proc of Expert Systems with Application, Vol. 3, pp. 366-374, 2008.
9. K.R De and A. Bhattacharya, "Divisive Correlation Clustering Algorithm (DCCA) for grouping of genes: detecting varying Patterns in expression profiles," bioinformatics, Vol. 24, pp.1359- 1366, 2008.S
10. Cheng, C.F.N cowan and P.M. grant "Orthogonal Least squares Learning Algorithm for Radial basis Function Networks". IEEE Transaction of Neural Network 2001, p 1-10, ol. 2 No.2.
11. Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin," A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012)
12. Park, J. and I. Sandberg, Approximation and radial-basis-function networks, Neural Computation 5 (1993), 305-316.
13. My Chau Tu, Dongil Shin,"A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms", in Proc of IEEE International Conference on Dependable, Autonomic and Secure Computing, pp. 183-187, December 2009.
14. Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", in Proc of IEEE/ACS International Conference on Computer Systems and Applications, pp.108-115, April 2008.
15. Kemal Polat, Salih Günes, "Medical decision support system based on artificial immune recognition system (AIRS), fuzzy weighted pre-processing and feature selection", An International Journal of Expert Systems with Applications Vol:33,No:2,pp:484-490, 2007
16. Lei Guo, Youxi Wu, Weili Yan, Xueqin Shen, Ying Li, "Research on Medical Diagnosis Decision Support System for Acid-base Disturbance Based on Support Vector Machine", proc. of the IEEE 27th Annual International Conference of the Engineering in Medicine and Biology Society, pp: 2413-2416, 2006.
17. Hongmei Yana, Yingtao Jiangb, Jun Zhenge, Chenglin Pengc, and Qinghui Lid, "A multilayer perceptron-based medical decision support system for heart disease diagnosis", Expert Systems with Applications, Vol 30, No: 2, pp: 272-281, 2006.
18. Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System For Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, Issue 1, 2007.
19. P. Jeatrakul and K.W. Wong, "Comparing the Performance of Different Neural Networks for Binary Classification Problems", IEEE Eighth International Symposium on Natural Language Processing, pp. 111-115,2009.
20. Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur, "Effective diagnosis of heart disease through neural networks ensembles", An international journal of expert system with applications, vol. 36,no. 4,pp. 7675-7680,May 2009.
21. Francisco Fernández-Navarroa, César Hervás-Martíneza, Roberto Ruizc, Jose C. Riquelme," Evolutionary Generalized Radial Basis Function neural networks for improving prediction accuracy in gene classification using feature selection", International journal of Applied Soft Computing, pp. 1787–1800, Jan. 2012.
22. Seena Mary Augusty, Sminu Izudheen, "A Survey: Evaluation of Ensemble Classifiers and Data Level Methods to Deal with Imbalanced Data Problem in Protein- Protein Interactions", Review of Bioinformatics and Biometrics (RBB), vol. 2 Issue 1,pp. 1-9, Mar. 2013
23. Orhan Er, Nejat Yumusak, Feyzullah Temurtas, "Chest diseases diagnosis using artificial neural networks", Expert Systems with Applications, pp. 7648–7655,2010.
24. O. Yavuz, and L. Ozyilmaz, "Analysis and Classification of Hiv-1 Sub-Type Viruses by AR Model through Artificial Neural Networks", World

Academy of Science, Engineering and Technology, pp.826-831,2009.
25. Guang-Bin Huang, Dian Hui Wang,Yuan Lan, " Extreme learning machines: a survey", Springer Int. J. Mach. Learn. & Cyber. pp. 107–122,May 2011.
26. Jae-Hong Eom, Sung-Chun Kim, Byoung-Tak Zhang, "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction", Expert Systems with Applications, vol. 34,pp. 2465–2479.2008.
27. Yuehui Chen, Yan Wang, and Bo Yang, "Evolving Hierarchical RBF Neural Networks for Breast Cancer Detection", Springer-Verlag Berlin Heidelberg, pp. 137-144, 2006.
28. Rahmat Zolfaghari, "Diagnosis of Diabetes in Female Population of Pima Indian Heritage with Ensemble of BP Neural Network and SVM", International Journal of Computational Engineering & Management, vol. 15, Issue 4,pp. 115-121, Jul. 2012.
29. Manaswini Pradhan and Dr. Ranjit Kumar Sahu, "Predict the onset of diabetes disease using Artificial Neural Network (ANN)", International Journal of Computer Science & Emerging Technologies (E-ISSN: 2044-6004), pp.303 -311, vol. 2, iss. 2, 2011.
30. Paul Mangiameli, David West, Rohit Rampal, "Model selection for medical diagnosis decision support systems", Decision Support Systems, vol. 36, pp. 247– 259,2004.
31. Gandhi Rajiv K., Karnan Marcus and Kannan S., "Classification rule construction using Particle Swarm Optimization algorithm for breast cancer datasets," Signal Acquisition and Processing.ICSAP, International Conference, 2010, pp. 233 – 237.
32. Sunila,Prabhat Panday, Nirmal Godara," Decision Support System for Cardiovascular Heart Disease Diagnosis using Improved Multilayer Perceptron", International Journal of Computer Applications (0975 – 8887), Volume 45– No.8, May 2012.
33. http://www.openclinical.org/aisp_era.html
34. http://www.statsoft.com/Textbook/Support-Vector-Machines.
35. http://www.dtreg.com

About Authors

Sunila Godara received MSc and MTech degree in Computer Science & Engg. She is working as Assistant Professor in Deptt. of Computer Sc. & Engg, Guru Jambheshwar University of Science & Technology, HISAR. She has published more than 20 papers in national/ international journals and conferences. Her research areas are Data Mining, Big Data and Database Management Systems.

Dr. Rishipal Singh received PhD degree in Computer Science from J.N University Delhi.He is working as Prof. in Deptt. of Computer Sc. & Engg, Guru Jambheshwar University of Science & Technology, HISAR. He has published more than 35 papers in national/ international journals and conferences.