KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS

Dr.M.Rajaiah[1], Dean Academics & HOD, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Ms.EswararajuHemasri[2], UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Ms. Damai Venkamma[3], UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur

Mr. Annapureddy Teja[4], UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

Mr. Bulagakula Gowtham[5], UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur

ABSTRACT_ Huge volumes of healthcare data are gathered by the industry, but regrettably they are not "mined" to reveal hidden information for efficient analysis, diagnosis, and decision-making. Finding hidden patterns and linkages frequently involves idle. Advanced data mining methods can assist and offer a solution to deal with this scenario. The technique of extracting hidden information from a large dataset is known as data mining. Clustering, classification, association analysis, regression, and other data mining techniques summarization, analysis of time series and sequences, etc. Data mining methods are important a crucial part in several fields including text mining, graph mining, medical mining, Web mining and mining of multimedia. The goal of this research project is to forecast renal Utilizing Artificial Neural Networks (ANN) and Support Vector Machine (SVM). This study compares the accuracy and execution times of these two algorithms in order to assess how well they function. According to the experimental findings, the ANN performs better than the alternative approach.

1. INTRODUCTION

Extraction of hidden information from a huge dataset is known as data mining. Data mining is utilised in a variety of fields, including image, text, sequence, web, graph, spatial, and other types of mining. Numerous applications, including fault diagnosis, anomaly detection, medical diagnosis, email filtering, face recognition, and oil spill detection, employ data mining techniques . To extract unknown knowledge from databases, data mining techniques like classification, clustering, association rules, and others are very important.

A data mining method called classification is used to forecast the group membership of data examples. As with clustering, classification divides information retrieved into different segments known as classes. The algorithm analyses a training set made up of a number of qualities and the desired outcome, often known as the objective or Prediction of four different kidney diseases—Chronic Kidney Disease, Acute Renal Failure, Acute Nephritic Syndrome, and Chronic Glomerulonephritis was the main emphasis of this investigation. Data mining methods like the Support Vector Machine (SVM) and Artificial Neural Network are used to forecast kidney disorders (ANN).

The remainder of the essay is structured as follows. Section 2 discusses related works. In Section 3, the suggested technique is presented. The experimental findings are examined in Section 4. The conclusion is given in Section 5.

2. LITERATURE REVIEW

Abhishek et.al [1] have used two neural network techniques, Back Propagation Algorithm (BPA), Radial Basis Function (RBF) and one non-linear classifier Support Vector Machine (SVM) and compared in accordance with their efficiency and accuracy. They used WEKA 3.6.5 tool for implementation to find the best technique among the above three algorithms for Kidney Stone Diagnosis. The main purpose of their thesis work was to propose the best tool for medical diagnosis, like kidney stone identification, to reduce the diagnosis time and improve the efficiency and accuracy. From the experimental results they concluded, the back propagation (BPA) significantly improved the conventional classification technique for use in medical field.

Andrew Kusiak et al [2] have used data preprocessing, data transformations, and data mining approach to elicit knowledge about the interaction between many of measured parameters and patient survival. Two different data mining algorithms were employed for extracting knowledge in the form of decision rules. Those rules were used by a decisionmaking algorithm, which predicts survival of new unseen patients. Important parameters identified by data mining were interpreted for their medical significance. They have introduced a new concept in their research work, it have been applied and tested using collected data at four dialysis sites. The approach presented in their paper reduced the cost and effort of selecting patients for clinical studies. Patients can be chosen based on the prediction results and the most significant parameters discovered.

Ashfaq Ahmed K et.al, [3] have presented a work using machine learning techniques, namely Support Vector Machine [SVM] and Random Forest [RF]. These were used to study, classify and compare cancer, liver and heart disease data sets with varying kernels and kernel parameters. Results of Random Forest and Support Vector Machines were compared for different data sets such as breast cancer disease dataset, liver disease dataset and heart disease dataset. The results with different kernels were tuned with proper parameter selection. Results were better analyzed to

establish better learning techniques for predictions. It is concluded that varying results were observed with SVM classification technique with different kernel functions.

Sadik Kara et.al [13] had concentrated on the diagnosis of optic nerve disease through the analysis of pattern electroretinography (PERG) signals with the help of artificial neural network (ANN). Implemented Multilayer feed forward ANN trained with a LevenbergMarquart (LM) backpropagation algorithm. The end results were classified as healthy and diseased. The stated results shown that the proposed method PERG could make an effective interpretation.

Koushal Kumar et.al [10] used Learning vector quantization (LVQ), two layers feed forward perceptron trained with back propagation training algorithm and Radial basis function (RBF) networks for diagnosis of kidney stone disease. They have compared the performance of all three neural networks on the basis of its accuracy, time taken to build model, and training data set size. They used Waikato Environment for Knowledge Analysis (WEKA) tool for execution. Finally from the experimental results, authors concluded that multilayer perceptron trained with back propagation is best algorithm for kidney stone diagnosis.

3. METHODOLOGY

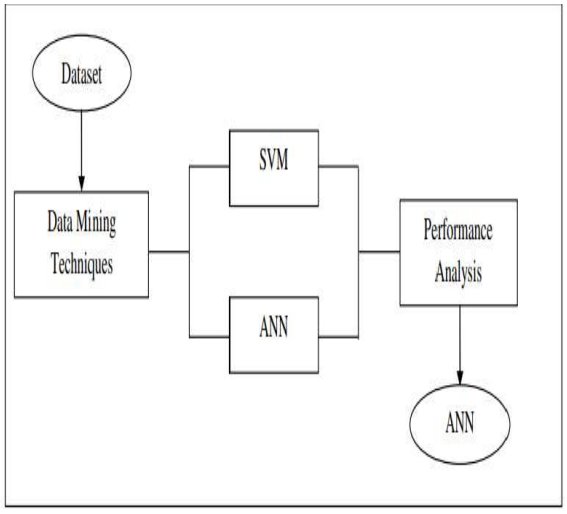

Fig 1. System Architecture

DATASET

The dataset is compiled from a number of hospitals, clinics, and medical labs. A synthetic kidney function test (KFT) dataset has been produced from this for the investigation of renal disease. Five hundred eighty-four cases make up this dataset, which uses six characteristics. comparative evaluation Age, Gender, Urea, and Creatinine are the characteristics in this KFT dataset. , as well as Glomerular Filtration Rate (GFR). The renal diseases included in this dataset include information.

DATA MINING TECHNIQUES

Various techniques are employed to carry out data mining jobs. In addition to requiring particular kinds of data structures, data mining techniques also imply particular kinds of algorithmic approaches. Data mining methods include classification, clustering, prediction, sequential pattern, and others. Parametric and nonparametric approaches are two different categories. A parametric model, in which certain parameters are not stated,
lgebraic equation to explain between
input and output the connection

Data mining applications are better suited to nonparametric approaches. A data-driven model is a nonparametric model. The model is established without the use of explicit equations. Neural networks, decision trees, and evolutionary algorithms are examples of nonparametric approaches.

SUPPORT VECTOR MACHINE (SVM)

A novel technique for classifying both linear and nonlinear data is support vector machines. A support vector machine (SVM) algorithm operates as follows in a case. Ittransforms the special training data into a higher dimension using a nonlinear mapping.

It looks for the linear optimum separation hyperplane in this new dimension.i.e., a "decision border" that distinguishes between tuples belonging to different classes. With an appropriateData from two classes may always be combined via nonlinear mapping to a necessary high dimension a hyperplane divides them apart. Support vectors and margins are used by the SVM to find the hyperplane. Even the fastest SVMs can have extremely long training times, yet despite this, they are incredibly accurate due to their capacity to simulate intricate nonlinear decision boundaries.

Compared to other approaches, they are far less prone to over fitting. A concise description of the learnt model is also provided by the support vectors initialization. SVMs may be used for classification and prediction. They have been used in a variety of contexts, such as benchmark time-series prediction tests, speaker identification, object recognition, and handwritten digit recognition.The pseudo code for the perceptron learning, taken from is given in Table 1

Table 1. Pseudo code for SVM

Require: A linear separable set S, learning rate $\eta \in \Re^+$

Step1 $w_0 = 0; b_0 = 0; k = 0;$

Step 2 $R = \max_{1 \le i \le l} \| x_i \|$

Step3 while at least one mistake is made in the for loop do

Step4 for $i = 1,...,l$ do

Step5 if $y_i(< w_k, x_i > + b_k) \le 0$ then

Step6 $w_{k+1} = w_k + \eta y_i x_i$

Step7 $b_k + \eta y_i \Re^2$ (updating bias [1])

Step8 $k = k + 1$

Step 9 end if

Step10 end for

Step11 end while

Step12 Return w_k, b_k, where k is the number of mistakes

Artificial Neural Network (ANN)

Backpropagation absorbs by repeatedly analysing a collection of training tuples and contrasting each tuple's predicted value with the actual known target value. The desired value could be the training tuple's known class label (for classification issues) or a continuous value (for prediction). The weights are changed for each training session to reduce the mean squared error between the goal value and the forecast made by the network.The term "backpropagation" refers to the process of making changes in a "backwards" fashion, starting with the output layer and moving through each hidden layer to the first hidden layer. Though it is not assured, in most cases the weights will eventually converge, and the learning process will come to an end. Inputs, outputs, and errors are used to describe the process stages.

Although each step is intrinsically straightforward once you've become used to the method. Table2 provides an algorithm summary.

Table 2: Pseudo code for Backpropagation

Input:

- *D*, a data set consisting of the training tuples and their associated target values;
- *l*, the learning rate;
- *network*, a multilayer feed-forward network

Output: A trained neural network.

Methods:

(1) Initialize all weights and biases in *network*;
(2) while terminating condition is not satisfied {
(3) for each training tuple *X* in *D* {
(4) // Propagate the inputs forward:
(5) for each input layer unit *j* {
(6) $Oj = Ij$; // output of an input unit is its actual input value
(7) for each hidden or output layer unit *j* {
(8) $I_j = \sum_i w_{ij} O_i + \theta_j$; //compute the net input of unit *j* with respect to the previous layer, *i*
(9) $O_j = \frac{1}{1+e^{-1}j}$;} // compute the output of each unit *j*
(10) // Backpropagate the errors:
(11) for each unit *j* in the output layer
(12) $Err_j = O_j(1-O_j)(T_j - O_j)$; // compute the error
(13) for each unit *j* in the hidden layers, from the last to the first hidden layer
(14) $Err_j = O_j(1-O_j)\sum_k Err_k w_{jk}$; // compute the error with respect to the next higher layer, *k*
(15) for each weight *wi j* in *network* {
(16) $\Delta w_{ij} = (l) Err_j O_i$; // weight increment
(17) $w_{ij} = w_{ij} + \Delta w_{ij}$;} // weight update
(18) for each bias θ_j in *network* {
(19) $\Delta\theta_j = (l) Err_j$; // bias increment
(20) $\theta_j = \theta_j + \Delta\theta_j$;} // bias update
(21) } }

4. EXPERIMENTAL RESULTS

Classification Accuracy

Table 3 represents the performance of classification accuracy measure of the datasets using classification algorithms such as SVM and ANN.

Table 3: Accuracy Measure

Algorithms	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	TP Rate
SVM	76.32	23.68	0.763
ANN	87.70	12.3	0.877

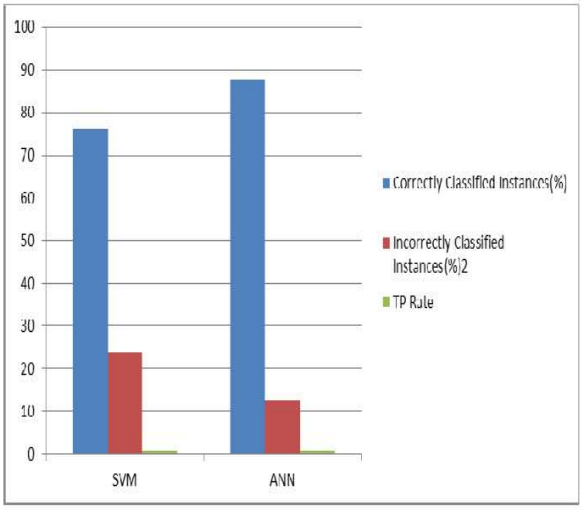

Fig 2. Accuracy Measure

Figure 2 represents the accuracy measure for the data mining algorithms namely SVM and ANN. From the experimental result, ANN performs best in classifying process than SVM algorithm.

EXECUTION TIME

Table 4 represents the execution time required for SVM and ANN algorithms

Table 4: Execution time Analysis

Algorithms	Execution Time in Seconds
SVM	3.22
ANN	7.26

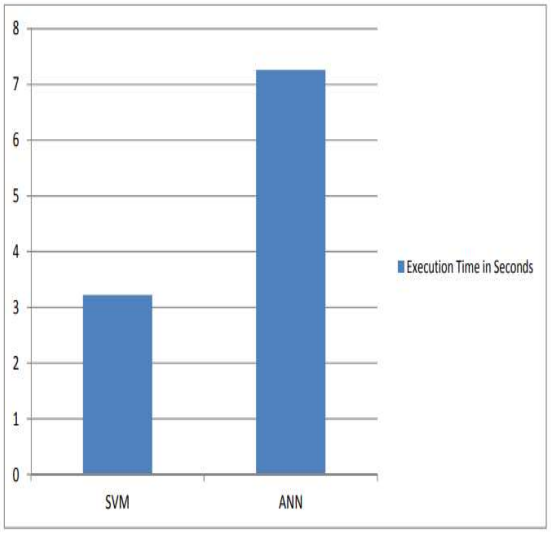

Table 5. Classification of Kidney Diseases

Data Mining Algorithms → / Kidney Disease	SVM	ANN
Normal	435	439
Acute Nephritic Syndrome	45	43
Chronic Kidney disease	42	42
Acute Renal Failure	19	16
Chronic Glomerulonephritis	42	43

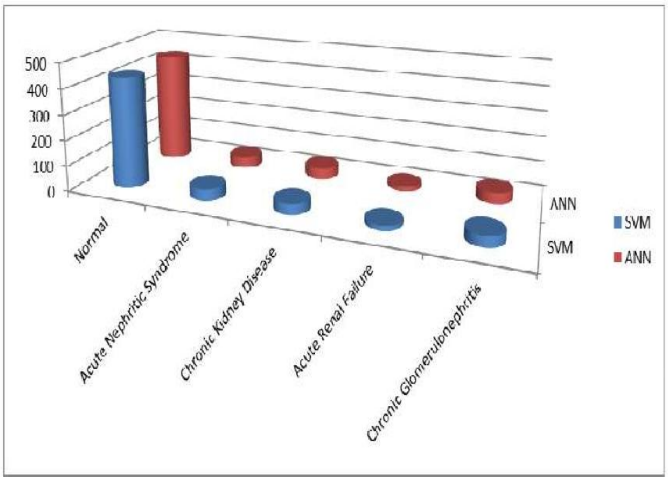

Fig 4. Classification of Kidney Diseases

Figure 4 represents the Kidney diseases classified by SVM and ANN algorithms. Based on chart analysis, ANN gives the overall best classification result than other algorithm.

5. RESULT AND DISCUSSION

The optimal algorithm has been determined to be the one with the highest accuracy and shortest execution time. Each classifier in this categorization has a different accuracy rate. Because ANN has the highest level of classification accuracy, it is regarded as the best classification algorithm. The SVM algorithm requires the least amount of execution time.

6. CONCLUSION

In this study, four different forms of kidney illnesses are categorised using a categorization approach.Based on the performance metrics classification accuracy and execution time, comparisons between the Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms are made. According t the findings, classification theANN outperforms the SVM classifier algorithm in terms of performance and produces more accurate results; for these reasons, it is regarded as the best classifier. SVM classifier may classify the data with the least amount of processing time. The execution time. ANN algorithm will be improved in the future to reduce

REFERENCE

1. Abhishek, GourSundarMitra Thakur, Dolly Gupta, (2012) "Proposing Efficient Neural Network Training Model for Kidney Stone Diagnosis", International Journal of Computer Science and Information Technologies, Vol. 3 (3) , page no 3900- 3904

2. AndrewKusiak, Bradley Dixonb, ShitalShaha, (2005) "Predicting survival time for kidney dialysis patients: a data mining approach", Elsevier Publication, Computers in Biology and Medicine 35, page no 311–327

3.Ashfaq Ahmed K, Sultan Aljahdali and Syed Naimatullah Hussain,(2013) "Comparative Prediction Performance with Support Vector Machine and Random Forest Classification Techniques", International Journal of Computer Applications Volume 69– No.11, May page no 12-16

4.Cristianini. N and Shawe-Taylor. J, (2000) "An Introduction to support vector machines and other kernel based learning methods". Cambridge University Press

5. Dipak V. Patil, Bichkar. R.S, (2011) "An Optimistic Data Mining Approach for Handling Large Data Set using Data Partitioning Techniques", International Journal of Computer Applications (0975 – 8887), Volume 24– No.3, June, page no 29- 33

6. IshaniMandli, Mahesh Panchal, (2014) "Selection of Most Relevant Features from High Dimensional Data using IG-GA Hybrid Approach", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.2, February, pg. 827-830

7.Jainesh Patel, Neha R Soni, (2014) "Survey of Supervised and Unsupervised Algorithms in Email Management", International Journal of Scientific & Engineering Research, Volume 5, Issue 3, March, page no 1241-1243

8. Jiawei han and Micheline Kamber, (2006) "Data mining Concepts and Techniques, Second edition", Morgan Kaufmann publishers, by Elsevier Inc. All rights reserved.

9. Khyati Chaudhary, Bhawna Mallick, "Exploration of Data mining techniques in Fraud Detection: Credit Card", International Journal of Electronics and Computer Science Engineering, Volume1,Number 3, page no 1765-1771

10.Koushal Kumar, Abhishek, (2012) "Artificial Neural Networks for Diagnosis of Kidney Stones Disease", International Journal of Information Technology and Computer Science, 7 July, page no 20-25

AUTHOR PROFILES

Dr.M.Rajaiah , Currently working as an Dean Academics & HOD in the department of CSE at ASCET (Autonomous), Gudur, Tirupathi(DT).He has published more than 35 papers in, Web of Science, Scopus Indexing, UGC Journals.

Ms.EswararajuHemasri, as B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur

Ms.DamaiVenkamma, as B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur

Mr. AnnapureddyTeja, as B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur

Mr.BulagakulaGowtham, as B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur