International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 06 | June -2020 ISSN: 2582-3930

Forecasting Phases of Kidney Diseases along Evaluating Data Mining Algorithm

Ms. Pooja Sonar[1], Prof. Mr. Kishor N. Shedge[2]

[1]PG Student, Dept. of Computer Engineering, SVIT Nashik, Maharashtra, India
[2]HOD, Dept. of Computer Engineering, SVIT Nashik, Maharashtra, India

---

Abstract -Information mining is an important procedure for ordering novels, legitimates, possibly helpful and reasonable examples in information. It precisely states as the extraction of data from a gigantic database. Information mining is an indispensable job in a few applications, for example, business associations, government organization, and human services industry. In the human services industry, the information digging is dominatingly utilized for infection expectation. Various information mining methods are existing for anticipating infections in particular arrangement, synopsis, affiliation rules, grouping, relapse and hence forth. The most important target of this research work is to predict kidney ailments utilizing characterization calculations, by help of Support Vector Machine. This research work is focusing on identifying the best characterization calculation depending upon the arrangements, exactness and the execution time with execution factors. From the test results it seems that the SVM is better classifier calculation for prediction of kidney diseases

Key Words:Prediction of kidney disease stages Data mining techniques Probabilistic Multilayer perception machine Radial basis function, SVM.

1. INTRODUCTION

Aglobalhealthissuewhichis widelygrowingis aChronickidney disease (CKD). It's a chronic condition associated with increased morbidity and mortality, a higher risk of several diseases includes cardiovascular disease, and higher health care costs. Moreover two millions of people across the world undergoes dialysis or kidney transplant treatment for stayingalive,yetthisnumbermay probably representonly10%ofaffected peoplewhoneeds treatment for staying alive. The majority out of the 2 million people who receives treatment for their kidney failure are in only five relatively wealthy countries, which only represent 12% of the total population across the world. By comparison, only 20% of the world's population is getting treated from about 100 developing countries, and they represent almost half of the global population. Annually,morethanonemillionpeoplein112lower-incomecountriesdie due to untreated kidney failure because of financial burden of dialysis or kidney transplantation treatment.

Thus, there is significant importance in the earlydetection, controlling,andmanagingofthedisease.Itisnecessaryforpredictingtheprogression of CKD with reasonable accuracy because of itsdynamicandcovert natureintheearlystages,andpatientheterogeneity.CKDisoftendescribedby aseveritystages which need to be identified at earliest.Clinicaldecisionsareinfluencedbythestagewhether apatientisprogressingandthe exactrateofprogression.Defining the disease stage is quite crucial task as it givesseveral indicationsthat supports the determination of various interventionand required treatments.Therefore, data mining can play a vital role in extractinghiddendata from the wide patient medical and clinical datasetthatphysiciansfrequently collect from patients to obtain insights aboutthediagnosticinformation, and to preplan implementation of precise treatment. Datamining can be further defined as the process of extracting hidden data from a large dataset. Several Data mining techniques are applied and used widely in various contexts and fields. With the help of data mining techniques we could predict, filter, classify, and cluster or group data. The prediction attribute refers to the algorithm processing of a training set containing a set of attributes and outcomes.

Machine learning algorithms have been used to predict and classify in the healthcare field. We have used the Support Vector Machine Algorithm to classify and predict diabetes and pre-diabetes patients, and as a result it shows that SVM is useful for classifying patients with common diseases. Similarly, It can be classified Alzheimer's disease by using a Support Vector Machine (SVM) to analyze whole-brain anatomical magnetic resonance imaging (MRI) for a particular setofpatients,andit's resultshowsthatSVMisapromisingapproach for Alzheimer's disease early detection. JoshiJ,[3] have done heart disease prediction using the Probabilistic Neural Network Algorithm,DecisiontreeAlgorithm,andNaïveBayesAlgorithm,andPRNN provides the best results as compared with the other algorithms for heart diseaseprediction.It can havepredictionofHBV-induced liver cirrhosis using the Multilayered Perceptron (MLP) Algorithm and as a result it shows that the MLP classifier gives satisfactory prediction outputsforliverdisease,mostlyinHBV-relatedlivercirrhosispatients.

2.RESEARCH METHODOLOGY

Data Mining is been utilized in our study because it is a process of identifying novel and is potentially useful, valid and ultimately understandable patterns in data [4]. Supervised and unsupervised learningtechniques are used for classification of data mining. A "supervised" learning technique requires the

building of a model based on previous performance analysis and is useful in both medical and clinical researches for classification, statistical regression and association rules [5]. On the other side, the"unsupervised" learning technique is not guided by prior analysis and it does not create a pre-analysis hypothesis. A model can beconstructed based upon the results and it can be useful for clustering [6].

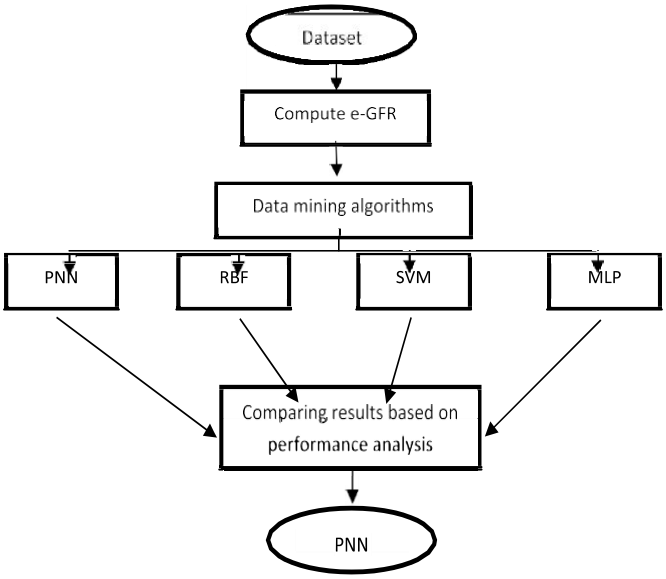


Fig. 1 Methodology workflow.

Three different types of commonly used artificial neural network algorithms and support vector machine algorithms have been used for this research, to determine which algorithm will generate the best classification results, so as to identify the different stages of chronic kidney disease, based on patient's clinical and laboratory data. (Shown in Fig. 1)

Machine learning techniques employ two phases to build the predictive/classification model as follows:

- A training phase that learns algorithmically how to build up a model by using training datasets with expected outputs.
- A validation phase that estimates how rigorously the model has been trained by using no. of validation datasets without the expected outputs

2.1 PROBABILISTICNEURALNetworks

Probabilistic Neural Networks (PNN) are a kind of Radial Basis Function neural network with one pass learning algorithm and highly parallelstructure.PNNwasintroducedbyDonaldF.Spechti n1990asa memory basednetworkthat helps in providingestimatesofcategoricalvariables. This algorithm provides a smooth approximation of a target function, evenwith a sparsedatainthemultidimensionalspace[8].Theadvantages ofPNNarefastlearningandeasytuning.ThePNNiscomposedoffour layers: input, pattern (RBF kernel function), summation, and a output, as shown in Fig. 2. Each neuron of the pattern layer uses a radial basis function as an activation function. This function is commonly taken to be as a Gaussian.

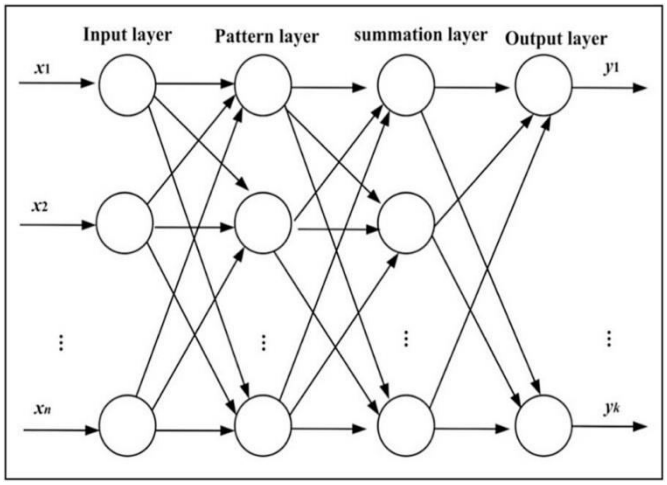


Fig. 2. Probabilistic Neural Network (PNN) Layers.

2.2 MULTILAYER PERCEPTRON ALGORITHM

The Multilayer Perceptron (MLP) is one of the most important class of neural networks, consisting of an input layer, one or more hidden layers,andtheoutputlayer,asshowninFig.3.MLPshavebeenappliedtosolvedifficultanddiverseproblems,bytraining themin a supervised manner using a well-known algorithm i.e., the error back-propagationalgorithm[3].Thisalgorithmisbased upontheerror correction learning rule. As such, it can be viewed as a generalization of an adaptive filtering algorithm.

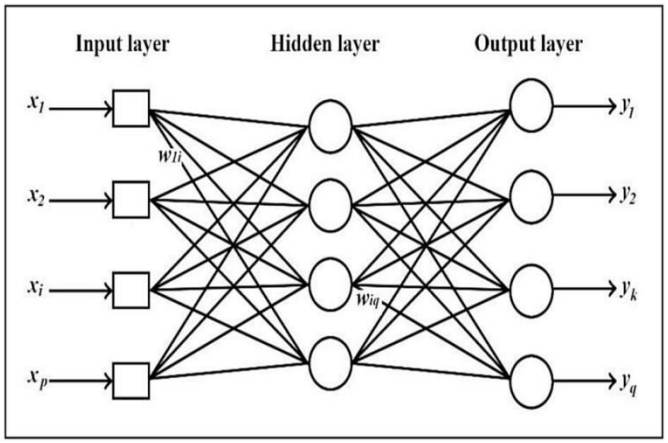
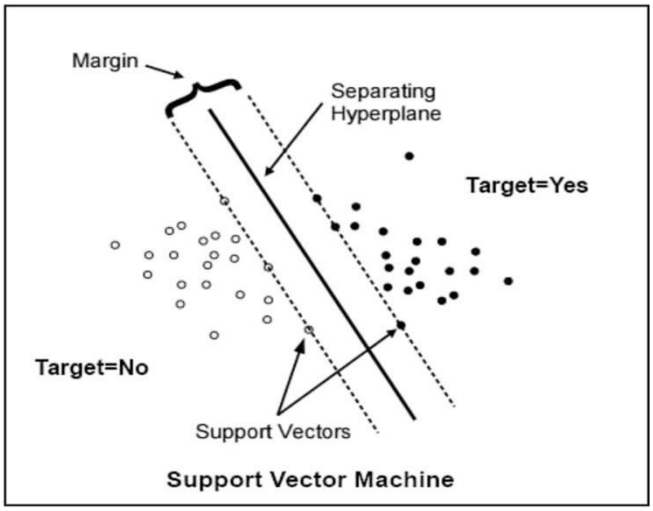


Fig. 3. Multilayer perceptron (MLP) layers.

2.1 SUPPORT VECTOR MACHINE ALGORITHM

The SVM is a method for the classification of both linear and non-lineardata.TheSVMalgorithmworksasfollows.Itusesan onlinear mapping technique to renovate the unique training data into a higher dimension. Surrounded by this new dimension, it examines the linear optimal separating hyperplane as shown in Fig. 4, i.e., a "decision boundary" sortingoutthetuplesofoneclassfromanother.Withasuitab



le

Fig. 4. Support Vector Machine (SVM) optimal hyperplane.

Nonlinear mapping to a necessarily high dimension, data from two classes can always be separated by a hyperplane. The SVM finds hyperplane using support vectors and margins. Although at the trainingtime,eventhefastestSVMscanbeexceedinglyslow, theyare accurate, and exemplary in their ability to model a complex nonlinear decision boundaries. They are much less prone to over fitting as compared with the other methods. SVM initiates also provide a compact description of the learned model. SVMs can be used for predicting along with classification.They have beenapplied to several areas, including handwritten digit recognition, object recognition, and speaker identification, as well as a benchmark time series predictiontests.

3. CHRONIC KIDNEY DISEASE (CKD)

CKD progression can be considered as a function of various parameters including underlying renal diseases, blood pressure, hypertension, proteinuria, and age. Early diagnosis of the CKD requires great attention among physicians, especially in determining the appropriate time for applying medical treatments and to control identified risk factors that reflects on the disease progression to End Stage Renal Disease (ESRD), such hypertension, proteinuria, and hyperphosphatemia.

3.1 STAGES of Chronic Kidney Disease

The stages of Chronic Kidney Disease (CKD) are mainly based on measured or estimated Glomerular Filtration Rate (eGFR). There are five stages in it, but kidney function is normal in the Stage 1, and minimally reduced in Stage 2.

Definitionofchronic: Labellingsomeoneashaving a CKDrequires mainlytwo samples at least 90 days apart. Historical values can be used for this. The estimated Glomerular Filtration Rate (eGFR) depends on creatinine measurement, sex, race and age. One of the most accurate methods to calculatetheeGFRistheModificationofDietinRenalDisease(MDRD).

$$eGFR=186x(Creatinine/88.4)^{-1.154}x(Age)^{-0.203}x(0.742iffemale)\ x\ (1.210\ ifblack)$$

3. CONCLUSIONS

TheProbabilisticNeuralNetworks algorithmgivesthehighestoverallclassificationaccuracypercentageof 96.7%, compared to other algorithms in classifying the stages of CKD patients. On the other hand, the Multilayer Perceptron requires a minimum execution time of 3s whereas the Probabilistic Neural Network requires 12s to finalize theanalysis.These algorithms have been compared with classification accuracy based on correctly classified stages of CKD patients, time taken to construct the model, and time taken to test the model. The Probabilistic Neural

Networks algorithm yields a better classification accuracy and prediction performance to predict the stages of chronic kidney disease patients butthe current study applied four data mining algorithms on a clinical/laboratory dataset consisting of 361 chronickidneydiseasepatients.Theresultsoftheaddressed algorithms have been compared to define the most accurate algorithm results in classifying the severity stage of CKD. This study recommends that the Support Vector Machine algorithm is the best algorithm that can be used by physicians in order to eliminate diagnostic and treatment errors at the early stage in beginning of kidney disease and to prevent in further harm and leading to get Kidney Transplant.

ACKNOWLEDGEMENT

It is my privilege to acknowledge with deep sense of gratitude to my Project Guide Prof.Mr.Kishor N. Shedge for his valuable suggestions and guidance throughout my course of study and timely help given to me in completion of Dissertation. I gladly take this opportunity to thank Prof. Mr. Kishor N. ShedgeHead of Department, for valuable guidance of Dissertation. I am thankful to all those who helped me directly or indirectly for Dissertation and research work.

REFERENCES

[1] Ramchoun H, Amine M, Idrissi J, Ghanou Y, Ettaouil M. Multilayer Perceptron: Architecture optimization and training. IJIMAI2016;4(1):26–30.

[2] Dhamodharan S. Liver disease prediction using conference on advance computing. Application Technologies; 2014, May.

[3] JoshiJ,DoshiR,PatelJ.Diagnosisandprognosisbreast cancerusingclassification rules. Int J Eng Res Gen Sci2014;2(6):315–23.

[4] Solanki AV. Data mining techniques using WEKA classification for sickle cell dis-ease.IntJComputSciInfTechnol2014;5(4):5857–60.

[5] Aljahdali S, Hussain SN. Comparative prediction performance with support vector machine and random forest classification techniques. Int J ComputAppl 2013;69(11).

[6] CouserWG,RemuzziG,MendisS,TonelliM.Thecontribu tionofchronickidneydisease to the global burden of major noncommunicable diseases. Kidney Int2011;80(12):1258–70.

[7] SchabackR.Apracticalguidetoradialbasisfunctions.Ele ctronicResource2007;11.

[8] LeveyAS,CoreshJ,BalkE,KauszAT,LevinA,SteffesMW,e tal.NationalKidney Foundation practice guidelines for chronic kidney disease: evaluation, classifica-tion,andstratification.AnnInternMed2003;139(2):[137]–47].

[9] Cristianini N, Shawe-Taylor J. An introduction to support vector machinesandother kernel-based learning methods. Cambridge university press; 2000.

[10] FukumizuK.Activelearninginmultilayerperceptro ns.Advancesinneuralin- formation processing systems. 1996. p.295–301.

[11] SpechtDF.Probabilisticneuralnetworks.NeuralNetwork1990; 3(1):109–18.

[12] YuW,LiuT,ValdezR,GwinnM,KhouryMJ.Applicatio nofsupportvectormachine modeling for prediction of common diseases: the case of diabetes and pre- diabetes. BMC Med InfDecisMak 2010;10(1):16.

[13] DessaiISF.Intelligentheartdiseasepredictionsystemusin gprobabilisticneural network. Int J Adv Comp Theory Eng (IJACTE)2013;2(3):2319–526.

[14] MagninB,MesrobL,KinkingnéhunS,Pélégrini-IssacM,ColliotO,SarazinM,etal. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology2009;51(2):73–83.