Survey on Prediction of Kidney Disease by using Data Mining Techniques

S. Gopika[1], Dr. M. Vanitha[2]

Research Scholar, PG and Research Dept of Computer Science, J.J College of Arts and Science (Autonomous)

Pudukottai, Tamil Nadu[1]

Assistant Professor, Dept of Computer Applications, Alagappa University, Karaikudi, Tamil Nadu [2]

Abstract: Data mining is the process of extracting interesting patterns of huge database hiding. The medical domain contains heterogeneous data as text, figures and images that can be suitably removed to provide various useful medical information. The obtained medical data may be useful to the doctor to detect the pattern of the disease, the predicted survival of the patient after the illness, the severity of the disease, and the like. The main objective of this paper is to analyze the application of data mining in the medical field and some of the techniques used in predicting kidney disease. Today, kidney disease is a major medical problem mining. Key medical industry mining data are used to predict data sets in kidney disease. Thus, different techniques are used to identify the DM-related kidney disease related conditions.

Keywords: Classification, Data Mining, Kidney Disease, Prediction.

I. INTRODUCTION

Data mining is the knowledge discovery in the database. Data mining techniques help to process the data and transform it into useful information. Predictive data mining results are useful in a wide variety of data processing areas, such as business intelligence, bioinformatics, health management, finance and other medical fields, and there are also many difficult tasks. This field requires accurate and timely diagnosis, which can save many patients' lives. Data mining plays a vital role in health analysis. Early detection and accurate results are achievable by the data mining algorithm of the physician. Different algorithms can be used for the diagnosis of various diseases. Accuracy and performance are also different based on the data used.

The medical data mining potential in the medical field rises to extract the hidden pattern in the data set. These patterns are used for medical diagnosis and prognosis. Medical data is globally dispersed, heterogeneous, and exaggerated. In order to engage in a user-oriented approach to novel and hidden data patterns, the data must be arranged. One of the major problems in probing health science or bioinformatics is the correct diagnosis of some important information in management. In general, numerous trials involve categorizing or grouping large-scale data in order to estimate censorship.

Kidney disease is an increasingly serious problem. For most people, slow years of kidney damage occur, often due to diabetes or high blood pressure. This is the so-called chronic kidney disease. When someone is sick, injured kidney function, sudden changes in the cause, or taking certain medications, this is called acute kidney injury. This can occur in normal kidneys or people who already have kidney problems. The main risk factors for developing kidney disease are:
- Diabetes,
- High blood pressure,
- Cardiovascular (heart and blood vessel) disease,
- A family history of kidney failure.

In chronic kidney disease, the patient's kidney is impaired and impairs its function. Chronic kidney disease (CKD) is a condition in which the kidneys are damaged, blood cannot be filtered and possible. In the advanced stage of CKD called end-stage renal disease (CKD), renal function is severely degraded.

II. LITERATURE REVIEW

Manish Kumar et al. Experiments were carried out using six algorithms for machine learning, i.e., the task of learning from a pool of UCI machines for chronic kidney disease: sorter random forest (RF) optimized low order (SMO), NaiveBayes radial basis function (RBF) Classifiers Multilayer Perception (MLPC) and Simple Logistics (SLG).

K.R. Lakshmi et al. This work is compared to the performance of Bayesian classifier, support vector machine (SVM) classifier and K-nearest neighbour based on accuracy, accuracy and execution time CKD prediction (KNN).

Lambodar Jena et al. The main objective of this paper is to use data from this data set to predict the classification of chronic kidney disease in each case accurately. Many

authors have used the WEKA tool in their work to compare the performance of different classifiers in multiple datasets. However, there was no work on the prediction accuracy of chronic kidney disease data. Here, we consider the classifier's six problems, and study its performance on the basis of the data sets to which they apply the various parameters.

S.Vijayarani et al. The study was to predict kidney disease using vector-based vector machines (SVM) and artificial neural networks (ANN). The aim of this study is to compare the performance of the two algorithms based on accuracy and run-time. From the experimental results it shows that the yield of RNA is better than other algorithms.

S. Ramya et al. The purpose of this work is to reduce the diagnostic time and improve the accuracy of the diagnosis by the classification algorithm. The proposed work deals with the classification of the different stages of chronic kidney disease severity. The experiment is carried out in different algorithms as back propagation neural networks, radial basis functions and random forests. Experimental results show that the radial basis function of the algorithm gives the results of other sorting algorithms and produces an accuracy of 85.3%.

Basma Boukenze et al. The development of large data sets in the health system is outlined and used in a collection of medical data using three learning algorithms. The goal of this study was to predict kidney disease by using multiple machine learning algorithms, SVM, C4.5 and Bayesian Networks (BNs), and selected efficient.

Milandeep Arora et al. This paper is a knowledge flow window tool Weka's data mining, not only provides a statistical evaluation of the results, but also provides a data flow diagram. At the end, after performing all the experiments, it has been concluded that J48 is the best algorithm to achieve accurate and early detection of chronic kidney disease in the three algorithms used in the early stages.

III.DATA MINING TECHNIQUES

There are many different methods for data mining tasks. Data mining techniques require not only certain types of data structures, but also methods of certain types of algorithms. Some common techniques of data mining are classification, clustering, prediction, sequence patterns, etc. There are two types of techniques for those parameters and nonparametric techniques. Parameter Model Description By using algebraic equations, there is no relationship between the input and output of some of the parameters specified. Nonparametric methods are more suitable for data mining applications. The non-parametric model is a data-driven one. No explicit formula is used to determine the model. Nonparametric techniques include neural networks, decision trees, and genetic algorithms.

A. Kidney Disease Prediction

Kidney disease refers to kidney damage, cannot be blood, they should be screened. Such damage can be wasteful in the body. For most people, slow years of kidney damage occur, often due to diabetes or high blood pressure. This is the so-called chronic kidney disease. When someone is sick, injured kidney function, sudden changes in the cause, or taking certain medications, this is called acute kidney injury. This can occur in normal kidneys or people who already have kidney problems. Kidney disease is an increasingly serious problem. 10% of the world's population is affected by chronic kidney disease (CKD), as well as millions of deaths each year because they are unable to obtain affordable treatment. Chronic kidney disease is the world's leading cause of death at number 27 in 1990, but increased to 18 in 2010. Risk factors for nephropathy are smoking, obesity, hypercholesterolemia, diabetes (type 1 and 2) Obstructive renal disease, including obstruction of the bladder by BPH, atherosclerosis, cirrhosis and hepatic failure, narrowing of the arterial supply of the kidneys, kidney cancer, bladder cancer, kidney stones, kidney infections. (SVM) and artificial neural network (ANN) were used to predict the prediction of renal disease. The accuracy of the classification, and the time to perform these algorithms. The results show that the neural network achieves accurate classification of the yield and the SVM has taken the minimum runtimes.

A set of chronic kidney disease library UCI machine learning data. They use algorithms such as Naive Bayes, Multilayer Perceptron, Support Vector Machine, J48 and Conjugate Rule Tree to compare the accuracy of the classification. His proposed multi-layer perceptual algorithm gives better classification accuracy and performance prediction to predict chronic kidney disease.

In its application, the paper machine support vector (SVM), decision tree (C4.5) algorithm and Bayesian network machine learning. Data sets of the Chronic Kidney Disease UCI machine learning knowledge base were used to predict chronic renal failure in patients with and without chronic kidney disease. C4.5 classification with the shortest running time and higher accuracy results.

The proposed decision tree, Bayesian classification, rule - based classification algorithm back propagation, support vector machine (SVM) classification K - nearest neighbor basic knowledge. Classifiers more accurately multi-layer perceptrons, random forests, naive Bayesian, support vector machines, K-nearest neighbors and radial basis functions.

B. Support Vector Machine (SVM)

Support Vector Machine, a new approach to the classification of linear and non-linear data. In a shell, a Support Vector Machine (SVM) is an algorithm that operates as follows. The use of non-linear data maps to renew unique training at a higher level. Surrounded by this new concept, the super optimal linear separation plane is examined, i.e., the "decision boundary" classification from another class of tuples. With the appropriate non-linear

mapping to a higher dimension certain, the data from the two classes can always be separated by a hiperplan. The support vector machine is a support vector and profit margin hyperplane. Although time-consuming SVM training may be too slow, they are very accurate in their ability to model complex models of nonlinear constraints that are excellent. They are less likely to be tuned than other methods. The support vector initiates also provide a compact description of the learning model. Support vector machines can be used for prediction and classification. They have been used in a variety of fields, including time-series prediction references that identify handwritten numbers, object recognition and speaker recognition and verification.

C. Artificial Neural Network (ANN)

Further propagating the tuple training data set absorbed by the iterative processing for comparing the prediction network with each known tuple of actual target values. The target value may be a training tuple (for classification problems) or a continuous value (prediction) for which the tag is known. For each tuple training, the weights are modified to minimize the mean square error between the prediction of the network and the target value. These changes are made in the direction "backward", that is, from the output layer, through each hidden layer to the first hidden layer, so named reverse propagation. Although there is no guarantee that the weight of the general convergence, the final learning process will stop. The input, output, and error terms in the involved steps are represented. However, once familiar with the process, each step is inherently simple.

D. Naive Bayes

Naive Bayesian classifier is a strong independence hypothesis (naive) based on Bayesian theorem (Bayesian statistics) application of simple probabilistic classification. The more descriptive term for the potential probability model is the "model independent entity." This assumption of personality is often confined to the truth in practical applications, and is therefore characterized as naive, but the algorithm tilts well in several supervised classification problems Will soon be able to learn. One advantage of the Naive Bayes classifier is that it only needs to train a small amount of data to estimate the parameters (variance of the device and variables) required in the classification. Since the independent variables assume that the variance of the variables used for each category should be determined, not the entire covariance matrix.

E. Neuro-fuzzy system

Chronic kidney disease is a large and growing problem of population aging. Detection and prognosis of kidney disease is important to provide adequate and appropriate patient treatment. Is used to detect the use of data sets in patients with kidney disease and the results of conventional systems used in the if-then rule, and with or bodies generated. The drawback of using this standard is that the probability of getting the wrong result is higher than the accuracy. A predictive error in the disease may lead to a loss of life in the patient. A new technique for predicting and detecting kidney disease using a patient data set is presented. This new technique uses two neural networks and fuzzy systems called neuro fuzzy systems that will produce the results based on the resulting input dataset. This new system is derived from a combination of fuzzy systems, which produces results from mathematical calculations rather than neural networks based on probabilistic theory. The results produced using mathematical calculations tend to be more accurate, thus improving system efficiency. Accurate prediction of kidney disease system development.

IV. CONCLUSION

The purpose of this study was to analyze the application of data mining algorithms and algorithms in the medical field and to predict disease. Data mining provides good results when applied with appropriate tools and techniques in the diagnosis of the disease. Therefore, data mining is a promising area for healthy prediction. Kidney disease can be predicted by using multiple classifications in data mining. You can also predict the level of chronic kidney disease using the algorithmic classifier.

According to experts although the number of parameters used in this study KD is limited. Thus, they were attracted to some new results unless each finding of a relationship between the other studies in the relevant factor prior knowledge KD was confirmed. Provided the field is most effectively available, and may be used in this study can be more significant results. In view of the scope of the study, the following research environment has been provided to readers and research.

REFERENCES

[1]. Manish Kumar "Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm", International Journal of Computer Science and Mobile Computing, Vol.5 Issue.2, February 2016, pg. 24-33.

[2]. K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna,"Performance comparison of three data mining techniques for predicting kidney disease survivability", International Journal of Advances in Engineering & Technology, Mar. 2014.

[3]. Lambodar Jena, Narendra Ku.Kamila "Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 Vol.4, Issue11, November 2015.

[4]. S.Vijayarani, S.Dhayanand "Kidney disease Prediction Using SVM and ANN Algorithms", International Journal of Computing and Business Research (IJCBR) ISSN (Online): 2229-6166 Vol. 6, Issue 2, March 2015.

[5]. S. Ramya, N. Radha "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms", International Journal of Innovative Research in Computer and Communication Engineering. Vol. 4, Issue 1, January 2016.

[6]. Basma Boukenze, Hajar Mousannif and Abdelkrim Haqiq "Performance of Data Mining Techniques to Predict in Healthcare Case Study: Chronic Kidney Failure Disease", International Journal of Database Management Systems (IJDMS) Vol.8, No.3, June 2016`

[7]. Milandeep Arora and Er. Ajay Sharma "Chronic Kidney Disease Detection by Analyzing Medical Datasets in Weka", International Journal of Computer Application (2250-1797) Volume 6– No.4, July- August 2016.

[8]. K.Manimekalai, ─Prediction of Heart Diseases using Data Mining Techniques‖, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 2, ISSN (Online):2320-9801, ISSN (Print):2320-9798, February 2016.

[9]. Jyoti Rohilla, Preeti Gulia, ─Analysis of Data Mining Techniques for Diagnosing Heart Disease", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 7, ISSN: 2277 128X, July 2015.

[10]. Beant Kaur "Review on Heart Disease Prediction System using Data Mining Techniques", IJRITCC, Volume: 2 Issue: 10, pp 3003-3008, October, 2014.

[11]. Abdel-Motaleb "Artificial intelligence algorithm for heart disease diagnosis using Phonocardiogram signals", IEEE, ISSN: 2154-0357, pp 1-6, 2012.

[12]. Dr. S. Vijayarani, Mr.S.Dhayanand, "Kidney Disease Prediction using SVM and ANN algorithms", International Journal of Computing and Business Research (IJCBR), Volume 6, Issue 2, ISSN (Online):2229-6166, March 2015.

[13]. Lambodar Jena, Narendra Ku. Kamila, "Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease", International Journal of Emerging Research in Management &Technology, Volume-4, Issue-11, and ISSN: 2278-9359, November 2015.

[14]. Basma Boukenze, Hajar Mousannif and Abdelkrim Haqiq, "Performance of Data Mining Techniques to Predict in Healthcare Case Study: Chronic Kidney Failure Disease", International Journal of Database Management Systems (IJDMS), Vol.8, No.3, June 2016.

[15]. Pushpa M. Patil, "Review on Prediction of Chronic Kidney Disease using Data Mining Techniques", International Journal of Computer Science and Mobile Computing, Vol. 5, ISSN 2320–088X, Issue. 5, May 2016.

[16]. K.Rajeswari,, "A Novel Risk Level Classification of Ischemic Heart Disease using Artificial Neural Network Technique - An Indian Case Study "International Journal of Machine Learning and Computing, Vol. 1, No. 3,Pp 231-235, 2011.

[17]. Miss. Manjusha B. Wadhonkar, "Classification of Heart Disease Dataset using Multilayer Feed forward backpropogation Algorithm " (IJAIEM) Volume 2, Issue 4,Pp 214-220, 2013.

[18]. S. Vaishnavi (April, 2014) "Artificial Intelligence Approach for Disease Diagnosis and Treatment", IJIRCCE, Vol. 2, Issue 4, pp 4000-4007.

[19]. D.Sindhuja, R. Jemina Priyadarsini, "A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder", International Journal of Computer Science and Mobile Computing, Vol.5, Issue.5, ISSN 2320–088X, May 2016.

[20]. V.Shankar sowmien, V. Sugumaran, C.P.Karthikeyan, T.R. Vijayaram, "Diagnosis of Hepatitis using Decision tree algorithm", International Journal of Engineering and Technology (IJET), Vol 8 No 3, e-ISSN : 0975-4024, p-ISSN : 2319-8613, Jun-Jul 2016.

[21]. Dr. S. Vijayarani, Mr. S. Dhayanand, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms‖, International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, ISSN: 2278 – 7798, April 2015.

[22]. Jothi Prabha A, A. Govardhan, "Application of Classification Techniques on Various Attributes of Breast Cancer", Vol. 4, Issue 6, and ISSN(Online): 2320-9801, ISSN(Print): 2320-9798, June 2016.

[23]. Durairaj M, Deepika R, "Prediction of Acute Myeloid Leukemia Cancer Using Data Mining- A Survey‖, Volume I, Issue 2, ISSN: 2394 – 6598, February 2015.

[24]. Ms. Nilam chandgude, Prof. Suvarna pawar, "A survey on diagnosis of diabetes using various classification algorithm", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 12, ISSN: 2321-8169, 6706 – 6710, December 2015.

[25]. Thirumal P. C, Nagarajan N, ─Utilization of Data Mining Techniques for Diagnosis of Diabetes Mellitus- A Case Study", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 1, ISSN 1819-6608, January 2015.

[26]. K. Rajalakshmi, Dr. S. S. Dhenakaran, "Analysis of Datamining Prediction Techniques in Healthcare Management System", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, ISSN: 2277 128X, April 2015.

[27]. S.Vijayarani, S.Dhayanand "Kidney disease Prediction Using SVM and ANN Algorithms", International Journal of Computing and Business Research (IJCBR) ISSN (Online): 2229-6166 Vol. 6, Issue 2, March 2015.

[28]. Vikas Chaurasia, Saurabh Pal, "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability", Vol. 3, Issue. 1, ISSN 2320–088X, January 2014.

[29]. V.Krishnaiah, Dr. G. Narsimha, Dr. N. Subhash Chandra, "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques", International Journal of Computer Science and Information Technologies, Vol.4 (1), ISSN: 0975-9646, 2013.

[30]. Durairaj M, Deepika R, "Prediction of Acute Myeloid Leukemia Cancer Using Data Mining- A Survey‖, Volume I, Issue 2, ISSN: 2394 – 6598, February 2015.