Identification and Prediction of Kidney syndrome Patients using Data Mining Practice

Mr. Smeet D. Thakur[1], Dr. Archana T. Bhise[2], Dr. Amit K. Gaikwad[3]

[1]*Research Scholar, Shri JJTU Rajasthan, India,333001*
[2] *Assistant Professor, H.V. Desai College, Pune and Research Scholar, JJTU, Rajasthan,333001*
[3]*Associate Professor, G.H. Raisoni College of Engineering, Amravati, India, 444701*

smeet87@gmail.com

Abstract – W*e are existing in a postmodern age and there are incredible change happening to our daily routines which make an impact on our physical condition optimistically and harmfully. As a result of these changes various kind of diseases are enormously increased. Especially, kidney disease has become more common these days. The being of inhabitants is at a danger. Variation in Weight, Appetite, Urine, UTI etc. can lead to kidney diseases that include poor kidney function. It may causes Kidney failure, Aneurysm, Peripheral artery disease, Kidney attack, Stroke and even sudden cardiac arrest. Many forms of kidney disease can be detected or diagnosed with different medical tests by considering ancestors medical history and other factors. But, the prediction of kidney diseases without doing any medical tests is quite difficult. The aim of this project is to diagnose different kidney diseases and to make all possible precautions to check at early stage itself with reasonable rate. We follow Data mining technique in which attributes are fed in to SVM, Randomforest, KNN, and ANN classification Algorithms for the prediction of kidney diseases. The preliminary readings and studies obtained from this method is used to know the possibility of detecting kidney diseases at initial stage and can be completely cure by proper identification.*

Keywords- FFBPNN, KNN, SVM, Clustering, data cleaning.

I- INTRODUCTION

There are so many diseases which affect us badly and one among them is kidney disease. It is a serious disease since we often hear that most of the inhabitants die out of kidney diseases and other kinds of similar diseases relates to kidney [1-3] It is observed by most of the medical scholars that at many times most of the kidney patients might not survive kidney attacks and they die with it. In this paper we would like to deal with the four classification techniques which is use to prediction of kidney disease [4-6]. Namely SVM, Random forest, KNN, ANN. The studies have been done by evaluating the medical profiles of people who undergoes treatment in PDMC (Medical College) Amravati, by categorizing their age, sex, Urine Sample, Dialysis sample etc. we chose those categories since it is observed that kidney diseases are mainly study similarly.

We expect there is always major in studying about kidney diseases syndrome. Our study we try to the chances of detecting the kidney diseases at premature stage [7-9].It can totally cure the disease by proper identification of Kidney Disease. Kidney is the prime part in a human body. It is an operating system of our body. Other functions human body will badly affect the irregular function of kidney [10-13]. Any disarray of the kidney is Kidney disease. Different from kidney disease is the problems with the urine and dialysis system as well as the kidney. According to the kidney disease, it is the leading cause of death in the India, European nations, US, and African Nations and will occur as a result of kidney disease. UTI, Kidney Stones, and Chronic Kidney Disease are some examples of kidney disease. Some important reasons of kidney disease are age, poor appetite, diabetics, fatness, genetic, Urine, Kidney Stones, cholesterol etc [14-17]. Usually, Kidney disease

International Journal of Innovations in Engineering and Science,   www.ijies.net

can be use with surgery or medication. But its effective prevention is not yet being done. The effective prevention kidney disease is also a target of the research [18-20]. The rest of this paper Section II describes the Data mining process. The survey is briefly discussed in Section III. Problems are identified in Section IV. Methodology discussed in Section V. Result are given in Section VI. finally gives Conclusion in Section VII.

I I - DATA MINING

It is an prosperity of data obtained in healthcare but there is no effectual study tool to find covert relationships in data. There are many applications and technologies to generate new information and among them Data mining is rapidly on the increase one which is useful for the large existing databases for having new information especially in medical field. Millions of people die of kidney diseases annually. Kidney disease diagnosis seems to be crucial application of data mining techniques. By using the information doctor can diagnosis of kidney disease. To evaluate different classification techniques in kidney disease diagnosis is the objective of this paper. The health care's are well known assertion is the data rich however knowledge poor. We hope Data mining methods can help as solution in this situation. Different data mining techniques can be using for the good result. Some of the new data mining methods are clustering, association, data cleaning, data visualization, classification, prediction, neural networks, data warehousing. The paper try to give details about various aspects of information by using data mining methods that to make enough precautions to fight against kidney disease. The patient movement is looking after continuously. There occurs any act. Then the risk antonyms of disease are enlightened to the doctor and patient. In machine learning algorithms with the help of computer technology is used to doctors are skill to predict kidney diseases at an first stage. This project provides an intuition about data mining technique used to forecast kidney diseases.

Fig. 1- System Architecture

The above block diagram Shows in Fig. 1 is systems market of data mining. We have selected less section of data and the selected data has undergone preprocessing and through post processing we get the knowledge.

III - LITERATURE SURVEY

Disease Forecasting System using data Mining methods Kidney disease is a term which is coined to describe the big quantity of healthcares that are related to the kidney. This defines the medicinal state of uncertain health conditions. It sways all the bits of the kidney. To diagnose the kidney disease by using Association rule mining, classification, clustering. Diagnosis of Kidney Disease Patients using Fuzzy Classification Technique. The kidney disease prediction is primary concern in medical science. It is easier to find out kidney diseases by using 'Data Mining' says similar earlier studies. By using various techniques in 'Data Mining' ,Fuzzy, KNN are classifier which belongs to the family of 'Data Mining' was used to classify the training and testing. Data belongs to different classes. Prediction and Diagnosis of Kidney Disease by Data Mining techniques in kidney disease diagnosis seems to be essential is the application of data mining techniques. Discovered Knowledge leads the physicians to its diagnosis of kidney disease and prevention. To evaluate different classification techniques in kidney disease diagnosis is the objective of this paper. Classifiers like J48 Decision Tree, K Nearest Neighbors (KNN), Naive Bayes (NB), and SMO are used to classify dataset. Prediction of Kidney Disease at early stage using Data Mining and Big Data Analytics: A Survey. This evaluation works out possibly the medical provisional's can help the patients by forecasting the kidney disease before it happens. There are many 'Data Mining' techniques available namely Classification techniques involving NB, Decision tree , Neural network , Genetic algorithm , AI. Clustering algorithms are SVM, KNN. Boosted Apriori: an Effective Data Mining Association Rules for Kidney Disease Prediction System. These day's numeral number of tests is given to the patient for find a disease. But this can be can reduce by using 'Data Mining' technique. At an untimely phase Kidney Disease is significant. A rapid and well organized detection technique to reduce numbers of deaths from kidney diseases. The study is to remove unseen patterns by using data mining techniques. The following Table I explains the data mining technique, accuracy of the above literature survey papers.
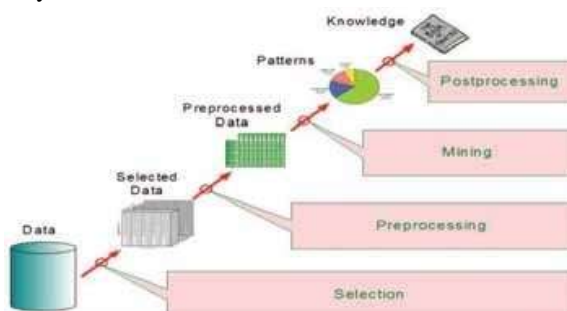
International Journal of Innovations in Engineering and Science,   www.ijies.net

Table 1-Comparison of Literature Survey

Authors	Data Mining Technique	Accuracy In (%)
aaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R., & Khovanova, N	KNN	63.6%
	MAFIA	57%
Perveen, S., Shahbaz, M., Keshavjee, K., & Guergachi, A.	Fuzzy	60.78%
	KNN	45.6%
in, A., & Stevens, P. E.	KNN	52%
	SMO	67%
	Naïve Bayes	81%
	Decision Tree	75%
)., & He, J.	Naïve Bayes	78%
	ANN	48%
	GA	66%
	SVM	73.2%
	KNN	81%
	AI	78%
Himani, and Mahesh H. Panchal	Decision Tree	68%

IV-PROBLEMS IDENTIFIED

The untimely detection of kidney diseases   can stop the death. But in every situation it is observe at the last stages of disease. Or after death .So the Health care's are point to detect the disease at early stages. In this case Data mining technique is the good technique can detect disease. Completely cure the disease by proper diagnosis. But the main problem of Data mining is using different algorithms for detection of kidney disease. Some algorithms are diagnose is less accurate and time consuming. Next problem is taking the survey less number of attributes is used in previous papers and also less number of patients is used.

V- METHODOLOGY

To diagnose the disease at early stage at affordable cost is the important aim of this paper. Using data mining Technique we can detect disease at first stage. We can completely cure the disease by proper diagnosis. Health care industry collect huge amount of information. Which are not mined to locate secret information? Solution of this problem is data mining technique. It is an approach method. This is used to examine large volumes of information. Extracts patterns that can be change to useful information. The data require to be gathering in a standardized form. Medical profiles 20 attributes are collected direct hospital, such as: These data are used to

forecast the patient getting cardio vascular disease. The real time data set collected from *"PDMC Medical College, Amravati"*. Collection of data was carried by interacting with patients one to one and jotting it down .The other mode of collecting data was from discharge summary of the respective patients .In such a way a total 20 attributes of nearly 2200 and above patients were collected.

Table 2- Different Attributes

Id	Attribute
1	Age
2	Sex
3	Blood pressure(DI)
4	Blood pressure (SI)
5	Sugar
6	Urea
7	Creatin
8	Bilirubin
9	SGOT (Serum glutamic oxaloacetic transaminase)
10	SGPT (Pyruvic)
11	PR (Per Rectum)
12	Sodium
13	Pottasium
14	HB(Hemoglobin)
15	Family history
16	Stress level
17	Sedentary lifestyle
18	Height
19	Weight

The Attributes are fed in to SVM shows in Table II Random forest, KNN, ANN classification Algorithms in kidney disease prediction. Applying the data mining technique to kidney disease treatment. It can provide as reliable performance as that achieved in diagnosing kidney disease. The main advantages of this project are early detection of kidney disease. Its diagnosis correctly on time. Providing treatment with affordable cost.

Here using 4 algorithms:

- Support Vector Machine(SVM)
- Random forest
- KNN
- ANN

These entire 4 algorithms are used by open CV in the platform of QT creator we find out the accuracy by testing and training of data.

Random Forest: Random forest is a decision tree is a decision support tool. That uses a tree like model of decisions It is predominantly an band of unprimed

International Journal of Innovations in Engineering and Science,   www.ijies.net

classification trees. It provides remarkable performance on a number of practical problems .Such as health care prediction problems .Iit is not sensitive to noise in the data set. It is not subjected to over fitting. It is built by combining the predictions of several trees. Each of them are trained separately. It works fast. Usually exhibits a significant performance improvement over many other tree-based algorithms .such as decision tree. Three main choices to be made a random tree

1. Splitting the leaves.

2 Predictor is to use in each leaf.

3. Inject randomness into the trees.

KNN: KNN is a non-parametric method in Fig. 5 Used for classification. And also used regression. It is a type of instance based learning. Where the function is only approximated locally. All computation is deferred vary till classification. This algorithm is among the simplest of all machine learning algorithms. Useful technique can be used to assign weight to the contributions of the neighbors. So that the nearer neighbors contribute more to the average than the more distant ones. A common weighting scheme consists in giving each neighbor a weight of 1/d. Where d is the distance to the neighbor. A peculiarity of the algorithm. It is sensitive to the local structure of the data.

Support Vector Machine: SVM is a supervised machine learning algorithm. It is mostly used in classification problems. This has a training phase in the initial .We need to give the data to the SVM algorithm shows in Fig. 2 The data which is classified. Then labeled already. Completing the prime stage of the training. Then data are given to the algorithm which will be classified with the help of the less stimulation of the human. It has a step called feature selection. Also in which they are identifying. When the unknown sample. Its prediction is node needed much then we will use SVM. Also the feature selection evenly. These both are also used for identifying the keysets .That is involved in different process that helps to differentiate the classes.
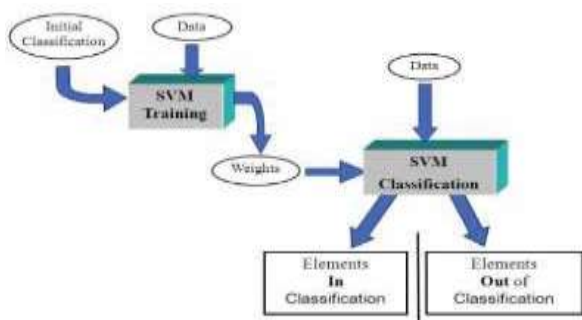
Fig 2- Process of SVM

ANN: ANN is a Computer system modeled on the human brain. And also the nervous system. The selected significant patterns for the neural network are trained .With effective prediction of kidney attack. Multi layer Perception Neural Network with Back propagation is being used as the training algorithm. In feed forward neural networks the neurons of the first layer forward

Their output to the neurons of the second layer. In a unidirectional fashion. This explains that the neurons are not received from the reverse direction. A kind of feed forward neural network mechanism .Is the Multilayer Perception Neural Networks. Multilayer feed forward neural network. ANN has three general steps and further divided into multi layers. These layers are input, output, hidden S Slayers.
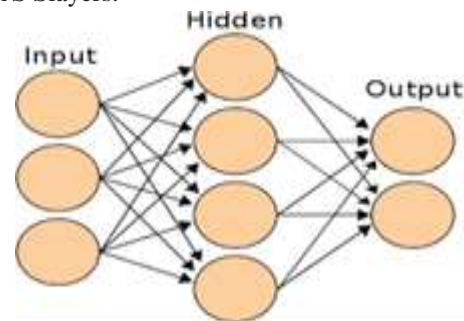
Fig. 3.- KNN-Layers

VI -RESULT
Table III -Comparison of Result

The Algorithms Used	Accuracy
Support Vector Machine(SVM)	85.88%
Random forest	85.88%
KNN	83.21%
ANN	92.21%

We have followed data mining technology in the research to get better results in Table III in order to diagnose and predict the kidney disease. We have used four algorithms namely ANN, KNN, SVM, RANDOM FOREST shows in Fig. 4, Fig. 5 these four algorithms have been used through Machine Learning and used through open CV we have practiced testing and training of each algorithms and could find out its accuracy and among these four algorithm ANN shows in Fig. more accuracy (92.21%) in comparing with others. Following pictures are the main outcome of data mining technique is that we can predict the kidney diseases and can receive precautions.

International Journal of Innovations in Engineering and Science,   www.ijies.net

Fig. 4-Using Svm Algorithm

Fig. 5-Using KNN Algorithm

VII - CONCLUSION

The major inspiration of this development is to provide an insight about detecting and curing kidney disease using data mining technique. For data mining, data were collected from PDMC Hospital Amravati. Collection of data was carried by interacting with patients one to one and scrawl it down .The other mode of collecting data was from discharge summary of the respective patients .In such a way, a total 22 attributes of nearly 2400 and above patients were collected. This collected data were then sorted and arranged systematically in Excel format. Using this data, it can be subjected to different data mining algorithms. From the medical profiles twenty attributes are extracted such as age, sex, urine sample etc. To predict the likelihood of patient getting kidney diseases these attributes are fed in to SVM, Random forest, KNN, and ANN classification Algorithms in which ANN gave the best result with the highest accuracy. Valid performance is achieved using ANN algorithm in diagnosing kidney diseases and can be further improved by increasing the number of attributes.

REFERENCES

[1] *Alizadehsani, R., Zangooei, M. H., Hosseini, M. J., Habibi, J., Khosravi, A., Roshanzamir, M., ... & Nahavandi, S. (2016). Coronary artery disease detection using computational intelligence methods. Knowledge-Based Systems, 109, 187-197.*

[2] *Janani, J., & Sathyaraj, R. (2021). Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(13), 6383-6390.*

[3] *Gupta, D., Khare, S., & Aggarwal, A. (2016, April). A method to predict diagnostic codes for chronic diseases using machine learning techniques. In 2016 International Conference on Computing, Communication and Automation (ICCCA) (pp. 281-287). IEEE.*

[4] *Neves, J., Martins, M. R., Vilhena, J., Neves, J., Gomes, S., Abelha, A., ... & Vicente, H. (2015). A soft computing approach to kidney diseases evaluation. Journal of medical systems, 39(10), 1-9.*

[5] *Purusothaman, G., and P. Krishnakumari. "A survey of data mining techniques on risk prediction: Kidney disease." Indian Journal of Science and Technology 8.12 (2015).*

[6] *Saly, D., Yang, A., Triebwasser, C., Oh, J., Sun, Q., Testani, J., ... & Wilson, F. P. (2017). Approaches to predicting outcomes in patients with acute kidney injury. PloS one, 12(1), e0169305.*

[7] *Wouters, O. J., O'donoghue, D. J., Ritchie, J., Kanavos, P. G., & Narva, A. S. (2015). Early chronic kidney disease: diagnosis, management and models of care. Nature Reviews Nephrology, 11(8), 491-502.*

[8] *Moreno-Sanchez, P. A. (2021). Development and evaluation of an Explainable Prediction Model for Chronic Kidney Disease Patients based on Ensemble Trees. arXiv preprint arXiv:2105.10368.*

[9] *Kazemi, Y., & Mirroshandel, S. A. (2018). A novel method for predicting kidney stone type using ensemble learning. Artificial intelligence in medicine, 84, 117-126.*

[10] *Elhoseny, M., Shankar, K., & Uthayakumar, J. (2019). Intelligent diagnostic prediction and classification system for chronic kidney disease. Scientific reports, 9(1), 1-14.*

[11] *Mohamadlou, H., Lynn-Palevsky, A., Barton, C., Chettipally, U., Shieh, L., Calvert, J., ... & Das, R. (2018). Prediction of acute kidney injury with a machine learning algorithm using electronic health record data. Canadian journal of kidney health and disease, 5, 2054358118776326.*

[12] *Neves, J., Martins, M. R., Vilhena, J., Neves, J., Gomes, S., Abelha, A., ... & Vicente, H. (2015). A soft computing approach to kidney diseases evaluation. Journal of medical systems, 39(10), 1-9.*

[13] *Saly, D., Yang, A., Triebwasser, C., Oh, J., Sun, Q., Testani, J., ... & Wilson, F. P. (2017). Approaches to predicting outcomes in patients with acute kidney injury. PloS one, 12(1), e0169305.*

International Journal of Innovations in Engineering and Science,   www.ijies.net

[14] *Tekale, Siddheshwar, et al. (2018). "Prediction of chronic kidney disease using machine learning algorithm." International Journal of Advanced Research in Computer and Communication Engineering 7.10,92-96.*

[15] *Janani, J., & Sathyaraj, R. (2021). Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(13), 6383-6390.*

[16] *Krittanawong, Chayakrit, et al. "Future direction for using artificial intelligence to predict and manage hypertension." Current hypertension reports 20.9 (2018): 1-16.*

[17] *Levin, A., & Stevens, P. E. (2011). Early detection of CKD: the benefits, limitations and effects on prognosis. Nature Reviews Nephrology, 7(8), 446-457.*

[18] *Saly, D., Yang, A., Triebwasser, C., Oh, J., Sun, Q., Testani, J., ... & Wilson, F. P. (2017). Approaches to predicting outcomes in patients with acute kidney injury. PloS one, 12(1), e0169305.*

[19] *Perveen, S., Shahbaz, M., Keshavjee, K., & Guergachi, A. (2018). A systematic machine learning based approach for the diagnosis of non-alcoholic fatty liver disease risk and progression. Scientific reports, 8(1), 1-12.*

[20] *Sathyadevan, Shiju, and Remya R. Nair. "Comparative analysis of decision tree algorithms: ID3, C4. 5 and random forest." Computational intelligence in data mining-volume 1. Springer, New Delhi, 2015. 549-562.*