Predicting Chronic Kidney Disease using Data Mining Techniques

Seethal V1, Kuldeep Baban Vayadande2

1Master of Computer Application, 2Assistant Professor,
1,2Jain Deemed-to-be University, Bengaluru, Karnataka, India

ABSTRACT
Kidney is a significant aspect of a human body. Kidney infection or disappointments are expanded in every year. Presently a day’s chronic kidney infection is the most well-known disease for the individuals. Today numerous individuals pass on due to chronic kidney disease. The principle issue of CKD is, it will influence the kidney gradually. A few people don’t have side effects at all and are analysed by a lab test. It depicts the steady loss of kidney work. Early recognition and therapy are viewed as basic variables in the management and control of chronic kidney disease. Data mining techniques are utilized to extract data from clinical and laboratory, which can be useful to help doctors to recognize the seriousness stage of patients. Using Probabilistic Neural Networks (PNN) algorithm will get better prediction for determining the severity stage of chronic kidney disease.

KEYWORDS: Chronic kidney Disease (CKD), Data mining techniques, Probabilistic neural networks (PNN), Python, machine Learning

INTRODUCTION
Chronic kidney sicknesses have become a significant reason for worldwide dismalness and mortality even in developing nations. Earlier viewed as a medical issue just in created nations, 4 out of 5 ongoing infection passing currently happen in low-and centre pay nations. Constant kidney illness incorporates conditions that harm your kidneys and decrease their capacity to keep you healthy. You may create difficulties like hypertension, anaemia (low blood count), weak bones, poor nutritional and nerve harm. Likewise, kidney illness builds your danger of having heart and vein infection. These issues may happen gradually throughout an extensive stretch of time. Chronic kidney illness might be brought about by diabetes, hypertension and different problems.

Discovery controlling and overseeing of the sickness at the beginning phase is very significant. In view of its dynamic and covert nature in beginning phases and patient heterogeneity, it is important to foresee movement of ckd with greatest precision. As ckd is frequently depicted by seriousness of the stages, clinical choices are impacted by the stages. Characterizing the stage is significant as it gives data that strengthens the assurance of required mediation and medicines.

The purpose of data mining in medical field is to identify useful and understandable patterns by analysing large sets of data. This data pattern helps to predict seriousness stage of patients and doctors can take actions required for that. Through this doctors can improve patients quality of life and perhaps they can save the lives of more patients. One of the usages of data mining is assessing treatment suitability, it incorporates investigating symptoms, causes and courses of treatment to find the best technique for a particular condition. This application could help standardize a technique for treatment for express diseases, as needs be making the diagnosis and treatment measure quicker and easier.

Literature Survey
The purpose of data mining in medical field is to identify useful and understandable patterns by analysing large sets of data. This data pattern helps to predict seriousness stage of patients and doctors can take actions required for that. The results of applying Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Radial Basis Function (RBF) algorithms have been investigated, and our revelations show that the PNN calculation, gives better characterization and prediction for deciding seriousness stage in Chronic kidney disease.

[1] S.Dilli Arasu (2017) proposed a model using classification algorithm. In this project he discussed about the different data mining techniques and comes into conclusion that classification algorithm gives better results compare to other techniques. Classification is the easiest and one of the well known data mining methods. Where objects are separated
and assigned to different groups called classes. Each object must be scattered precisely to one class and not multiple and never to no classes. [2] Mrs Prasuna Kotturu, Mr VVS Sasank, G Supriya, Ch Sai Manoj, M V Maheshwaredy (2019) In this paper, they discussed various machine learning algorithms like decision tree, naïve Bayes classification, logistic regression (LR), support vector machine (SVM) and random forest. In this paper they detect the chronic kidney disease (CKD) using the best suited method and got 99.3% as the most exact using random forest method. [3] Siddheshwar Tekale, Pranjali Shingavi (2018) In this paper they have analysed 14 various attributes related to CKD patients and predicted the accuracy by using decision tree and support vector algorithms. And they observed that decision tree algorithm gives the accuracy of 91.75% and SVM gives 96.75% accuracy. [4] Shubha Vashisth Ishika Dhall; Shipra Saraswat (2019) in this paper they have analysed around 400 patients data and found that multi layer perceptron algorithm gives better results compare to other algorithms.

Proposed System
The purpose of data mining in medical field is to identify useful and understandable patterns by analysing large sets of data. This data pattern helps to predict seriousness stage of patients and doctors can take actions required for that.

Before starting the analysis, e-GFR were determined to recognize the seriousness phase of the kidney illness for every patient by applying the eGFR equation. Glomerular filtration rate (GFR) is a test used to check how well the kidneys are functioning. In particular, it appraises how much blood goes through the glomeruli every moment. Glomeruli are the little channels in the kidneys that filter waste from the blood. By using the result of GFR test we have to apply the data mining techniques into it and compare the results based on performance analysis. The results of applying Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Radial Basis Function (RBF) algorithms have been investigated, and our revelations show that the PNN calculation, gives better characterization and prediction for deciding seriousness stage in Chronic kidney disease.

Probabilistic Neural Networks (PNN) are a kind of Radial Basis Function neural network with a one pass learning algorithm and highly parallel structure. PNN was introduced by Donald F. Specht in 1990 as a memory-based network that provides estimates of categorical variables. The algorithm provides a smooth approximation of a target function, even with sparse data in a multidimensional space. The advantages of PNN are fast learning and easy tuning. The PNN algorithm contains four layers: input, pattern, summation, and output. Each neuron of the pattern layer uses a radial basis function as an activation function.

Conclusion
The dataset utilized in the analysis comprised of 401 CKD patients and contained 25 variables. Before starting the analysis, eGFR were determined to recognize the seriousness phase of the kidney illness for every patient by applying the eGFR equation. By using the result of GFR test we have to apply the data mining techniques into it and compare the results based on performance analysis. The probabilistic neural networks algorithm gives the better classification percentage of 96.7% compared to other algorithms.

Acknowledgement
I would like to express my sincere feeling and obligation to Dr MN Nachappa and Asst. Prof. Kuldip Baban Vayadande and project coordinators for their effective steerage and constant inspirations throughout my analysis work. Their timely direction, complete co-operation and minute observation have created my work fruitful.

Reference