Prediction of Chronic Kidney Disease Using Weighted Associative Classifier (WAC)

D. Rajesh., MCA., M.Tech
Head of the Department of BCA
S.S.R College of Science and Management Studies
Kothanur

Abstract

Data Mining in Healthcare has become a present trend for obtaining accurate results of medical diagnosis. Chronic Kidney Disease (CKD) has become an international fitness problem and is a place of concern. It is a situation where kidneys turn out to be damaged and cannot filter toxic wastes within the frame. By using Data Mining Techniques, researchers have the scope to predict the Chronic Kidney Disease. Enormous data mining techniques are existing for predicting diseases namely classification, clustering, association rules, summarizations, regression and etc. The main objective of this research work is to predict kidney diseases using classification algorithms such as Associative classification. This helps doctors to diagnose and suggest the treatment at an early stage. It also helps the patients to know about their health condition at an earlier stage and follow necessary diet and prescriptions.

Keyword: *Data mining, Associative Classification, WAC*

1. Introduction

Data mining aids in healthcare to support for effective treatment, healthcare management, customer relation management, fraud and abuse detection and decision making. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. In chronic kidney disease, the patient's kidneys are damaged and decrease their functions. If Kidney decrease gets worse, waste can build to high levels in your blood and many complications may develop like high blood pressure, anemia, weak bones, poor nutritional health and nerve damage.

2. Related Work

There is a continuous study and research going on the field of medical diagnosis. A lot of work has been done on diseases like Kidney failure, Diabetes, Heart attack using several data mining techniques.

Parul Sinha and Poonam Sinha [1] developed a decision support system to predict chronic kidney disease. They have compared results of two techniques Support Vector machine and KNN (K Nearest Neighbor). Their experimental result shows that KNN has higher accuracy than SVM. J Chitra Devi et. al [3], the creator depicts that C4.5 grouping calculation manages numerical properties and additionally all out characteristics. A heart disease prediction is done using three data mining techniques namely Neural Network, Decision Tree and Naive Bayes. Their results disclose that neural networks with 15 features has surpassed two other techniques and accordingly is selected as the predictive model [4]. Prediction of four types of Kidney diseases namely Nephritic Syndrome, Chronic Kidney disease, Acute Renal Failure and Chronic Glomerulonephritis. Supervised classification algorithm Support Vector Machine (SVM) and Artificial Neural Network (ANN) is used to predict the kidney disease. Experimental results show that ANN is a best classifier Classification accuracy for ANN is higher compared to SVM and the execution time for SVM is lower compared to ANN. ANN has better classification accuracy [5]. Different machine learning classification algorithm for diagnosis of chronic kidney disease is discussed. Various classification techniques that are used are: Decision Tree, Linear Discriminant classifier, Quadratic Discriminant classifier, Linear SVM, Quadratic SVM, Fine KNN, Medium KNN, Cosine KNN, Cubic KNN, Weighted KNN, Feed Forward Back Propagation Neural Network using Gradient Descent and Feed Forward Back Propagation Neural [6].

3. The approach

I. Associative Classification

Associative classification is a new classification approach integrating association mining and classification. It becomes a significant tool for knowledge discovery and data mining. Association mining, or pattern discovery, aims to discover descriptive knowledge from database. A typical associative classification system is constructed in two stages: 1) discovering all the event associations (in which the frequency of occurrences is significant according to some tests) 2) generating classification rules from the association patterns to build a classifier. In the first stage, the learning target is to discover the association patterns inherent in a database also referred to as knowledge discovery. In the second stage, the task is to select a small set of relevant association patterns discovered to construct a classifier given the predicting attribute.

This approach used total twenty-four features, most of which are clinical in nature and the rest are physiological. Table 1 summarizes various parameters. As a part of data pre-processing, missing values and outliers are imputed with mean value of that feature for continuous data and attribute model value for categorical data. Nominal data are converted to numerical values. For example, Nominal values 'Normal' are labelled "1" and 'Abnormal' are labelled "0".

Table 1: Attributes of Chronic Kidney Disease

1. Age	13. Sodium
2. Blood pressure	14. Potassium
3. Specific gravity	15. Hemoglobin
4. Albumin	16. Packed cell volume
5. Sugar	17. White blood cell count
6. Red blood cells	18. Red blood cell count
7. Pus cell	19. Hypertension
8. Pus cell clumps	20. Diabetes mellitus
9. Bacteria	21. Coronary artery disease
10. Blood glucose random	22. Appetite
11. Blood urea	23. Pedal edema
12. Serum creatinine	24. Anemia

II. Weighted Associative Classifier (WAC)

Weighted Associative Classifier (WAC) is a new concept that uses Weighted Association Rule for classification. Weighted ARM (Associative Rule Mining) uses Weighted Support and Confidence Framework to extract Association rule from data repository. The WAC has been proposed as a new Technique to get the significant rule instead of flooded with insignificant relation. Figure 1 depicts Associative Classifier for Data Mining.

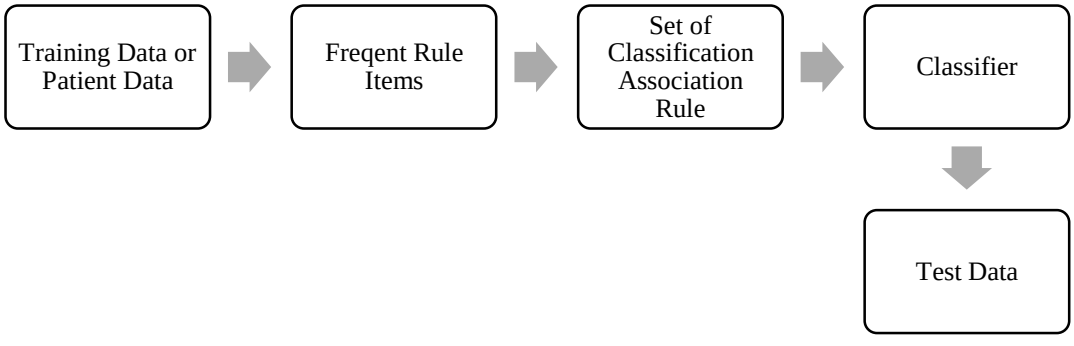

Figure 1: Associative Classifier for Data Mining

The major steps are as follows:

1. Initially, the Kidney disease data warehouse is pre -processed in order to make it suitable for the mining process.

2. Each attribute is assigned a weight ranging from 0 to 1 to reflect their importance in prediction model. Attributes that have more impact will be assigned a high weight (nearly 0.9) and attributes having less impact are assigned low weight (nearly 0.1).

3. Once the pre-processing gets over, Weighted Association Rule Mining (WARM) algorithm is applied to generate interesting pattern. This algorithm uses the concept of Weighted Support and Confidence framework instead of tradition support and confidence. Rules generated in this step are known as CAR (Classification Association Rule) and is represented as X ◊ Class label where X is set of symptoms for the disease. Example of such rules are (Hypertension, "yes") Kidney_Disease="yes" and {(creatinine,">1.3"), (Blood urea, ">45"), (Hypertension, "yes")} Kidney_Disease= "yes".

4. These rules will be stored in Rule Base.

5. Whenever a new patient's record is provide, the Classification Association Rule (CAR) rule from the rule base is used to predict the class label.

Conclusion

In this paper, I have presented an intelligent and effective kidney disease prediction methods using data mining. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. Data mining have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions. The proposed work can be further enhanced and expanded for the automation of Kidney disease prediction. Real data from Health care organizations and agencies needs to be collected and all the available techniques will be compared for the optimum accuracy.

References

1. Paul Sinha, Poonam Sinha, "*Comparative Study of Chronic Kidney Disease Prediction Using KNN and SVM*", International Journal of Engineering Research and Technology (IJERT) ISSN : 2278-0181-IJERV4 IS1 20622.

2. Vijayarani, S., Dhayanand, M. S., & Phil, M. (2015). Kidney disease prediction using svm and ann algorithms‖. International Journal of Computing and Business Research (IJCBR) ISSN (Online), 2229-6166.

3. J Chitra Devi, "Binary Decision Tree Classification based on C4.5 and KNN Algorithm for Banking Application", International Journal of Computational Intelligence and Informatics, September 2014.

4. Koh Bhatla, N., & Jyoti, K. (2012). An analysis of heart disease prediction using different data mining techniques. International Journal of Engineering, 1(8), 1-4.

5. Boukenze, B., Mousannif, H., & Haqiq, a. predictive analytics in healthcare system using data mining techniques. Computer Science & Information Technology, 1.

6. Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.