RESEARCH ARTICLE OPEN ACCESS

A Review Paper on Chronic Kidney Disease Detection

Milandeep Arora [1], Er. Ajay Sharma [2]
M. Tech Research Scholar [1], Associate Professor [2]
Department of Computer Science and Engineering
PTU Regional Centre, ACET, Amritsar
Punjab - India

ABSTRACT

As the world is advancing day by day, health problems are also getting multiplied with it. Similarly, the problem of chronic kidney disease is getting worsened day by day. It is also known as chronic renal disease and is a life threatening disease, it has various symptoms such as high blood pressure, anemia, rashes, muscle pain, conjuctivitus, etc. So, in order to tackle this problem it has to be detected at earliest stages possible and given suitable treatment before it get worsened. In this paper we have given a review about the symptoms of chronic kidney disease in detail and how to deduce a suitable method to effectively detect chronic kidney disease at early stages possible with the help of data mining. Data mining is an effective and reliable tool in performing several classification operations on the data collected for chronic kidney disease. I will also discuss the suitable algorithms proposed in this paper and also the methods by which the classification operations are performed. The platform used to further carry out our research work in WEKA data mining tool which supports many machine learning algorithms which we will be using in our research to get suitable results. The main aim of this paper is present a clear view of the chronic kidney disease, its symptoms and the criteria to detect it at earliest stages possible which will help the mankind to get safe from this life threatening disease.

Keywords:-Chronic kidney disease, classification, data mining, WEKA, etc.

I. INTRODUCTION

Chronic kidney disease also known as chronic renal disease is a very dangerous and life threatning disease which is becoming very common now a days. In order to deal with it, either it has to be detected at earliest stage possible or has to be given suitable treatment if detected. It has various physical symptoms such as tiredness, poor appetite, cramping, swollen feets and ankles, etc. but we cannot detect the disease effectively based on only these symptoms. To effectively detect and classify this disease we have to take certain minute and particular symptoms which could help us in determining the disease with high accuracy and these are as follows such as age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, anemia and last one is class. These twenty four are the particular symptoms that I will be taking in order to detect the chronic kidney disease accurately and the last value that is a class is not a symptom, it is just used a variable which help us to verify either a patient is suffering from chronic kidney disease or not. These symptoms are can also be viewed in tabular format with their short forms which can be used easily at the time of making data sets of the readings of various patients, this can be as follows

S.NO.	SYMPTOM	SHORT FORM
1.	Age	age
2.	Blood pressure	bp
3.	Specific gravity	sg
4.	Albumin	al
5.	Sugar	su
6.	Red blood cells	rbc
7.	Pus cell	pc
8.	Pus cell clumps	pcc
9.	Bacteria	ba
10.	Blood glucose random	bgr
11.	Blood urea	bu
12.	Serum creatinine	sc
13.	Sodium	sod
14.	Potassium	pot
15.	Hemoglobin	hemo
16.	Packed cell volume	pcv

17.	White blood cell count	wc
18.	Red blood cell count	rc
19.	Hypertension	htn
20.	Diabetes mellitus	dm
21.	Coronary artery disease	cad
22.	Appetite	appet
23.	Pedal edema	pe
24.	Anemia	ane
25.	Class	class

Table 1 list of the symptoms of chronic kidney disease to be used in WEKA

These are the actual and precise symptoms of chronic kidney disease which we will be using in detecting chronic kidney disease in WEKA data mining tool with the help classification algorithms of data mining. [1]

II. DATA MINING

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

EVOLUTION	ENABLING TECHNOLOGIES	CHARACTERISTICS
Data collection	Computers, tapes and disks	Retrospective, static data delivery
Data access	RDBMS and SQL	Retrospective, dynamic data delivery at record level
Data warehousing and decision support	OLAP, multidimensional databases and data warehouses	Retrospective, dynamic data delivery at multiple levels
Data mining	Algorithms, multiprocessor computers and massive databases	Prospective, proactive information delivery

Table 2 showing evolution of data mining

The most commonly used techniques in data mining are:

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) .

- Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

- Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset

(where k ³ 1). Sometimes called the k-nearest neighbor technique.

- Rule induction: The extraction of useful if-then rules from data based on statistical significance.

Many of these technologies have been in use for more than a decade in specialized analysis tools that work with relatively small volumes of data. These capabilities are now evolving to integrate directly with industry-standard data warehouse and OLAP platforms. The appendix to this white paper provides a glossary of data mining terms. [2]

Architecture for Data Mining

To best apply these advanced techniques, they must be fully integrated with a data warehouse as well as flexible interactive business analysis tools. Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. Furthermore, when new insights require operational implementation, integration with the warehouse simplifies the application of results from data mining. The resulting analytic data warehouse can be applied to improve business processes throughout the organization, in areas such as promotional campaign management, fraud detection, new product rollout, and so on. Figure 2 illustrates an architecture for advanced analysis in a large data warehouse. [2]

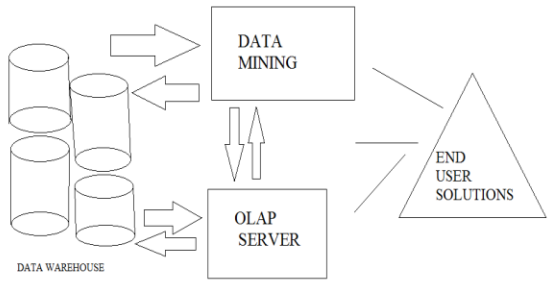

Fig 1 architecture of data mining

Process of Data Mining

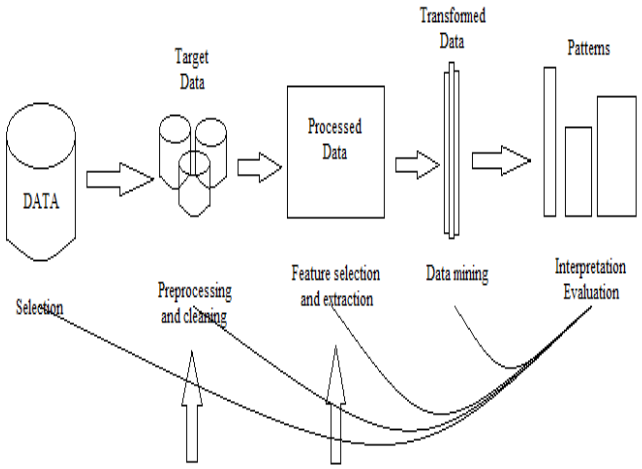

Fig 2 process of data mining [3]

Data mining is a promising and relatively new technology. Data mining is defined as a process of discovering hidden valuable knowledge by analyzing large amounts of data, which is stored in databases or data warehouse, using various data mining techniques such as machine learning, artificial intelligence(AI) and statistical. [4]

Data Mining Algorithms

An *algorithm* in data mining (or machine learning) is a set of heuristics and calculations that creates a model from data. To create a model, the algorithm first analyzes the data you provide, looking for specific types of patterns or trends. The algorithm uses the results of this analysis over many iterations to find the optimal parameters for creating the mining model. These parameters are then applied across the entire data set to extract actionable patterns and detailed statistics.

The mining model that an algorithm creates from your data can take various forms, including:

- A set of clusters that describe how the cases in a dataset are related.
- A decision tree that predicts an outcome, and describes how different criteria affect that outcome.
- A mathematical model that forecasts sales.

- A set of rules that describe how products are grouped together in a transaction, and the probabilities that products are purchased together.

The algorithms provided in SQL Server Data Mining are the most popular, well-researched methods of deriving patterns from data. To take one example, K-means clustering is one of the oldest clustering algorithms and is available widely in many different tools and with many different implementations and options.

Choosing the best algorithm to use for a specific analytical task can be a challenge. While you can use different algorithms to perform the same business task, each algorithm produces a different result, and some algorithms can produce more than one type of result. For example, you can use the Microsoft Decision Trees algorithm not only for prediction, but also as a way to reduce the number of columns in a dataset, because the decision tree can identify columns that do not affect the final mining model.

Choosing an Algorithm by Type

Data Mining includes the following algorithm types:

- Classification algorithms predict one or more discrete variables, based on the other attributes in the dataset.
- Regression algorithms predict one or more continuous numeric variables, such as profit or loss, based on other attributes in the dataset.
- Segmentation algorithms divide data into groups, or clusters, of items that have similar properties.
- Association algorithms find correlations between different attributes in a dataset. The most common application of this kind of algorithm is for creating association rules, which can be used in a market basket analysis.
- Sequence analysis algorithms summarize frequent sequences or episodes in data, such as a series of clicks in a web site, or a series of log events preceding machine maintenance.

However, there is no reason that you should be limited to one algorithm in your solutions. Experienced analysts will sometimes use one algorithm to determine the most effective inputs (that is, variables), and then apply a different algorithm to predict a specific outcome based on that data. SQL Server Data Mining lets you build multiple models on a single mining structure, so within a single data mining solution you could use a clustering algorithm, a decision trees model, and a Naïve Bayes model to get different views on your data. You might also use multiple algorithms within a single solution to perform separate tasks: for example, you could use regression to obtain financial forecasts, and use a neural network algorithm to perform an analysis of factors that influence forecasts. [5]

Algorithms To Be Used In This Paper

Naive Bayes

This is a classification technique based algorithm which is based on Bayes' Theorem and its main feature is that it doesn't have inter dependence among the various predictors to be used in the experiments.

J48

This algorithm helps in information gain, a concept that measures the amount of information contained in a set of data. It gives the idea of importance of an attribute in a dataset.

SMO

This algorithm devides the large quadratic problems into smallest possible sets and solve data analytically which helps in saving the excess time, reduced memory space is required for data and allow to use huge datasets very efficiently.

Rep Tree

It is a fast decision tree learner, it builds a decision regression tree using information gain, variance reduction and prunes it using reduced error pruning. It only sorts the numeric attributes once. The missing values are dealt with by splitting corresponding instances into pieces.

Random Tree

This tree is constructed randomly from a set of all possible tree having K random features at each node. All the trees have equal chance of being sampled or means uniform distribution.

III. LITERATURE SURVEY

Lambodar Jena et al [6] has suggested the use various algorithms on chronic kidney disease datasets and compare the results based on different parameters but only on single interface of WEKA and it has been calculated that multilayer perceptron algorithm performs the best among used algorithms.

Morteza Khavanin Zadeh et al [7] suggested the prediction of early chronic kidney disease and data of 193 patients who underwent homodialysis in Hasheminejad Kidney Center were explored. Eight common attributes of the patients including age, sex, hypertension level, Diabetes Mellitus state, hemoglobin level, smoking behavior, location of Arteriovenous fistula, and thrombosis state were used in the machine learning process and only two algorithms are used for prediction process.

L.Jerlin Rubini et al [8] used only three different algorithms such as radial basis function network, multilayer perceptron, and logistic regression. Also the interface used in this study only one.

Shital Shah et al [9] suggested the data mining approach which helps in detecting the chronic kidney effectively and realating it to the survival of the patient but again only with the help of limited algorithms and single interface.

Dr. S. Vijayarani et al [10] simply concludes in this paper that comparison two different algorithms such as SVM and Naïve Bayes. It is concluded that SVM performs better than Naïve Bayes algorithm in case of chronic kidney disease with the help of single platform.

K.R.Lakshmi et al [11] has compared the results of three data mining techniques that is Artificial Neural Networks, Decision tree and Logical Regression. It is used to elicit knowledge about the interaction between these variables and patient survival. Finally, ANN is suggested for Kidney dialysis to get better results with accuracy and performance.

Dr. S. Vijayarani et al [12] The objective of this research work is to predict kidney diseases by using Support Vector Machine (SVM) and Artificial Neural Network (ANN). The aim of this work is to compare the performance of these two algorithms on the basis of its accuracy and execution time. From the experimental results it is observed that the performance of the ANN is better than the other algorithm.

IV. PROPOSED METHODOLOGY

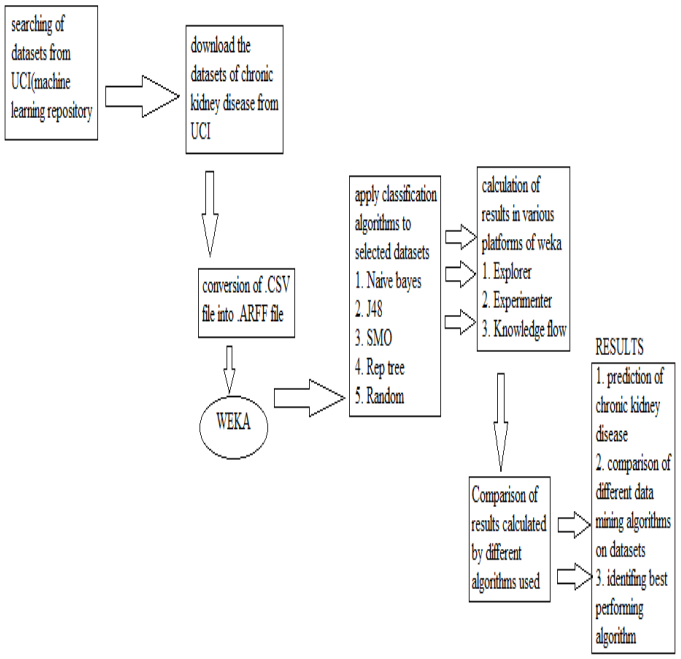

Fig 3 flow diagram of proposed work

V. TOOL USED

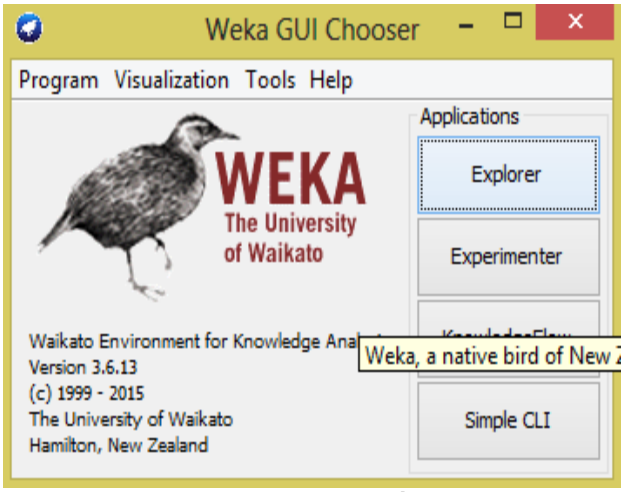

Fig 4 WEKA opening window

The Waikato Environment for Knowledge Analysis (WEKA) is a machine learning language tool written in java. Weka was developed at the University of Waikato in New Zealand. Weka is open source data mining tool developed in java. It is use for research, education, and application. It can be run on Windows, LINUX and Mac.

Weka is landmark system in the history of the data mining and machine learning research, communities because it is the only toolkit that has gained such widespread adoption and survived for and expended period of time. Other data mining and machine learning systems that have achieved this are individual system, such as C4.5, not toolkit. Since weka is freely available for download and ofer many powerful feature, it has become one of the most widely used data mining systems. Weka also became one of the favourite vehicles for data mining research and helped to advanced it by making many powerful features available to all. To sum up the weka team has made an outstanding contribution to the data mining field. There are mainly 2 ways to use weka to conduct your data mining tasks

Use Weka graphical user interfaces(GUI)

GUI is straightforward and easy to use. But it is not flexible. It cannot be called from you own application.

Import Weka java library to your own java application

Developer can leverage on Weka java library to develop software or modify the source code to meet special requirement. It is more flexible and advanced. But it is not as easy to use an GUI. [13]

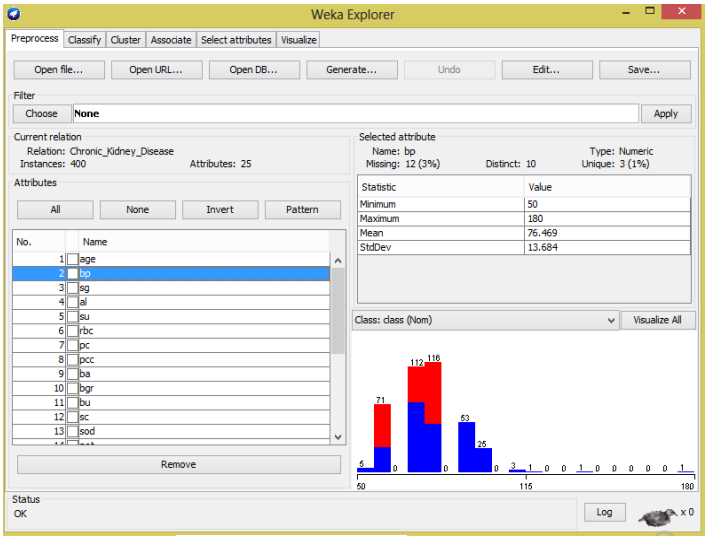

Fig 5 Explorer interface

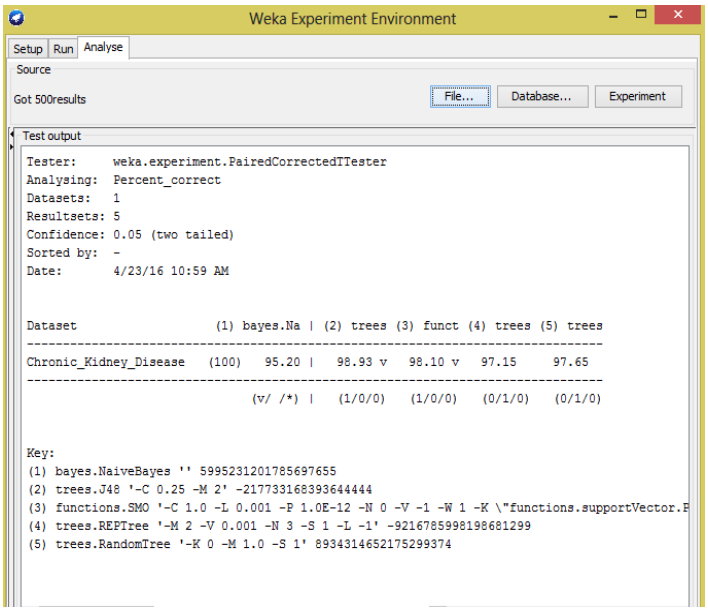

Fig 6 Experimenter interface

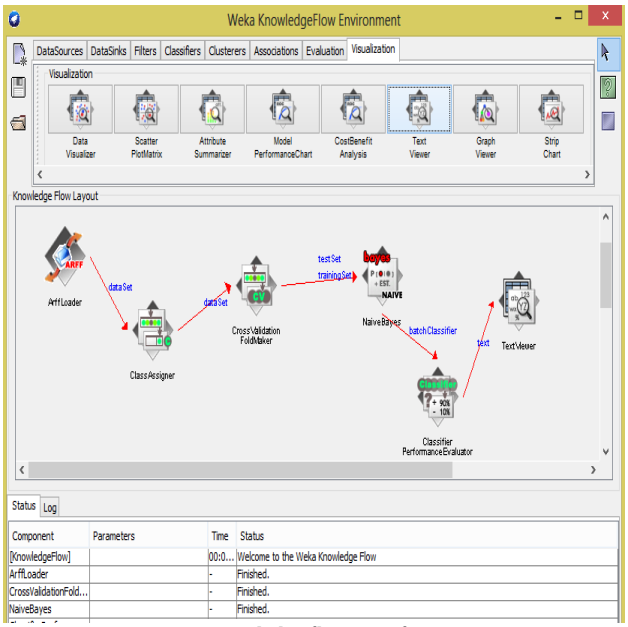

Fig 7 Knowledge flow interface

VI. CONCLUSION

In this paper, we have proposed to use five different type of classification algorithms with the help of WEKA data mining tool. The major difference between this paper and other research papers published in this field is that, all others have used only one or mostly two algorithms to accomplish the results and only one interface either explorer, interface or knowledge flow. Although one has used five algorithms but only with one interface, but we have used all five algorithms with all three interfaces and then carry out the results, compare them and computed the results. This paper will certainly help the patients of this disease to have a better detection. Also for future purpose we can compare results of same disease on two different platforms such WEKA and MATLAB tool and comparing both their results for further more effective results. This can be used as a potential Phd topic. Also this can be used very effectively by various health organizations to detect not only this disease but by taking an example from this can also take help to detect other diseases as well.

REFERENCES

[1] archive.ics.uci.edu/ml/datasets.html

[2] www.thearling.com/text/dmwhite/dmwhite.htm

[3] recommender-systems.readthedocs.io/en/latest/datamining.html

[4] www.zentut.com/data-mining/data-mining-processes/

[5] msdn.microsoft.com/en-us/library/ms175595.aspx

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

[7] International Journal of Hospital Research 2012, 2(1):49-54 Data Mining Performance in Identifying the Risk Factors of Early Arteriovenous Fistula Failure in Hemodialysis Patients Morteza Khavanin Zadeh , Mohammad Rezapour , Mohammad Mehdi Sepehri

[8] International Journal Of Modern Engineering Research (IJMER) | IJMER | ISSN: 2249–6645 | www.ijmer.com | Vol. 5 | Iss. 7 | July 2015 | 49 | Generating comparative analysis of early stage prediction of Chronic Kidney Disease L.Jerlin Rubini, Dr.P.Eswaran

[9] S. Shah, A. Kusiak, and B. Dixon, Data Mining in Predicting Survival of Kidney Dialysis Patients, in Proceedings of Photonics West - Bios 2003,Bass, L.S. et al. (Eds), Lasers in Surgery: Advanced Characterization, Therapeutics, and Systems XIII, Vol. 4949, SPIE, Belingham, WA, January 2003, pp. 1-8.

[10] International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 4, August

2015 DOI: 10.5121/ijci.2015.4402 13 DATA MINING CLASSIFICATION ALGORITHMS FOR KIDNEY DISEASE PREDICTION Dr. S. Vijayarani, Mr.S.Dhayanand

[11] International Journal of Advances in Engineering & Technology, Mar. 2014. ©IJAET ISSN: 22311963 242 Vol. 7, Issue 1, pp. 242-254 PERFORMANCE COMPARISON OF THREE DATA MINING TECHNIQUES FOR PREDICTING KIDNEY DIALYSIS SURVIVABILITY K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna

[12] International Journal of Computing and Business Research (IJCBR) ISSN (Online) :2229-6166 Volume 6 Issue 2 March 2015 KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS Dr. S. Vijayarani1, Mr.S.Dhayanand

[13] www.cs.waikato.ac.nz/ml/weka/