International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A literature Review on Prediction of Chronic Kidney Disease Using Data Mining Techniques

Smitha Patil

Assistant Professor, Department of Information Science & Engg, Sir. MVIT Bangalore

Abstract: ***Nowadays Kidney Disease is a growing problem in the world wide. Due to the high possibility of death within a short period of time, a patient must be hospitalized and appropriately cured. Many Data Mining techniques are used in the health care industry for predicting the Kidney Disease. The Data Mining techniques, namely SVM, Naive Bayes, Decision Tree, Classification, Neural Network are used to analyze the accuracy for the kidney related disease.***

Keywords: Data Mining, Chronic Kidney Disease, Data Mining Techniques

1. Introduction

Data mining is 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 [11]. Data mining techniques can be classified into supervised learning technique and unsupervised learning technique. A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances**.** This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way. In Data mining, the problem of unsupervised learning is that of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution.

2. Kidney Disease

Kidneys filter extra water and wastes out of blood and make urine. Kidneys also help control blood pressure so that the body can stay healthy. Kidney disease means that the kidneys are damaged and can't filter blood like they should. This damage can cause wastes to build up in the body. It can also cause other problems that can harm your health [12].

For most people, kidney damage occurs slowly over many years, often due to diabetes or high blood pressure. This is called chronic kidney disease. CKD includes condition which affects the kidney and if the kidney gets worse, waste can build to high level in the blood, which damage both kidney and leads to death.

When someone has a sudden change in kidney function because of illness, or injury, or has taken certain medications—this is called acute kidney injury. Acute kidney failure happens when kidney suddenly lose the ability to eliminate excess salts, fluids and waste materials from the blood. It develops over a few hours or few days to week. This can occur in a person with normal kidneys or in someone who already has kidney problems.

Kidney disease is a growing problem. More than 20 million Americans may have kidney disease and many more are at risk. Anyone can develop kidney disease, regardless of age or race. The main risk factors for developing kidney disease are:

- Diabetes,
- High blood pressure,
- Cardiovascular (heart and blood vessel) disease, and
- A family history of kidney failure.

2.1 Five Types of Kidney Failure

Acute Perennial Kidney Failure
Insufficient blood flow to the kidneys can cause acute perennial kidney failure. The kidneys can't filter toxins from the blood without enough blood flow. This type of kidney failure can usually be cured once the cause of the decreased blood flow is determined.

Acute Intrinsic Kidney Failure
Acute intrinsic kidney failure can be caused by direct trauma to the kidneys, such as physical impact or an accident. Causes also include toxin overload and ischemia, which is a lack of oxygen to the kidneys. Ischemia may be caused by:

- Severe bleeding
- Shock
- Renal blood vessel obstruction
- Glomerulonephritis, which is an inflammation of the tiny filters in your kidneys

Chronic Perennial Kidney Failure
When there isn't enough blood flowing to the kidneys for an extended period of time, the kidneys begin to shrink and lose the ability to function.

Chronic Intrinsic Kidney Failure
This happens when there is long-term damage to the kidneys due to intrinsic kidney disease. Intrinsic kidney disease is caused by a direct trauma to the kidneys, such as severe bleeding or a lack of oxygen.

Chronic Post-Renal Kidney Failure
A long-term blockage of the urinary tract prevents urination, which causes pressure and eventual kidney damage.

2.2 Symptoms of Kidney Failure

Many different symptoms can be signs of kidney failure. No symptoms are present sometimes, but usually someone with kidney failure will see a few signs of the disease. Possible symptoms include:

- A reduced amount of urine
- Swelling of your legs, ankles, and feet from retention of fluids caused by the failure of your kidneys to eliminate water waste
- Unexplained shortness of breath
- Excessive drowsiness or fatigue
- Persistent nausea
- Confusion
- Pain or pressure in your chest
- Seizures
- A coma

3. Data Mining Techniques Used for Prediction

Classification is a very important data mining task, and the purpose of classification is to propose a classification function or classification model. The classification model can map the data in the database to a specific class. Classification construction methods include: Decision Tree, Naive Bayes, ANN, K-NN, Support Vector Machine, Rough Set, Logistic Regression, Genetic Algorithm, and Clustering [11].

Decision Tree: A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree & applying the tree to the dataset. There are many popular decision tree algorithms CART, ID3, C4.5, CHAID, and J48.

Artificial Neural Network (ANN): An artificial neural network (ANN), often just called a "neural network"(NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase

Naïve Bayes: Naive Bayes classifier is based on Bayes theorem. This classifier algorithm uses conditional independence, means it assumes that an attribute value on a given class is independent of the values of other attributes. The Bayes theorem is as follows: Let X={x1, x2... xn} be a set of n attributes. In Bayesian, X is considered as evidence and H is some hypothesis means, the data of X belongs to specific class C. We have to determine P (H|X), the probability that the hypothesis H holds given evidence i.e. data sample X. According to Bayes theorem the P (H|X) is expressed as P (H|X) = P (X| H) P (H) / P (X).

K-Nearest Neighbour: The k- nearest neighbour's algorithm (K -NN) is a method for classifying objects based on closest training data in the feature space. K-NN is a type of instance-based learning. The k -nearest neighbour algorithm is amongst the simplest of all machine learning algorithms. But the accuracy of the k -NN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance.

Logistic Regression: The term regression can be defined as the measuring and analyzing the relation between one or more independent variable and dependent variable. Regression can be defined by two categories; they are linear regression and logistic regression. Logistic regression is a generalized by linear regression. It is mainly used for estimating binary or multi-class dependent variables and the response variable is discrete, it cannot be modeled directly by linear regression i.e. discrete variable changed into continuous value. Logistic regression basically is used to classify the low dimensional data having non-linear boundaries. It also provides the difference in the percentage of dependent variable and provides the rank of individual variable according to its importance. So, the main motto of Logistic regression is to determine the result of each variable correctly.

Rough Sets: A Rough Set is determined by a lower and upper bound of a set. Every member of the lower bound is a certain member of the set. Every non-member of the upper bound is a certain non-member of the set. The upper bound of a rough set is the union between the lower bound and the so-called boundary region. A member of the boundary region is possibly (but not certainly) a member of the set. Therefore, rough sets may be viewed as with a three-valued membership function (yes, no, perhaps). Rough sets are a mathematical concept dealing with Uncertainty in data. They are usually combined with other methods such as rule induction or clustering methods.

Support Vector Machine: are a set of supervised learning methods used for classification, regression and outlier's detection. Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships.

Genetic algorithm: Genetic algorithm is stochastic search methods which have been inspired by the process of biological evolution. Because of gas robustness and their uniform approaches to large number of different classes of problems, they have been used in many applications. Genetic algorithms are a probabilistic search and evolutionary optimization approach. Genetic algorithms provide a comprehensive search methodology for machine learning and optimization.

Custering: In brief, cluster analysis groups data objects into clusters such that objects belonging to the same cluster are similar, while those belonging to different ones are dissimilar. Clustering plays an important role in a broad range of applications, from information retrieval to CRM. Such applications usually deal with large datasets and many

attributes. Exploration of such data is a subject of data mining.

4. Comparison of Data Mining Techniques

The Data mining techniques used in kidney related diseases, in some experiments the results may differ. The table has shown the results below.

Table 1: Accuracy Rate of Classification Techniques used for Chronic Kidney Disease

Author	*Kidney Disease*	*Method*	*Accuracy*
S.Ramya, Dr. N.Radha	Chronic kidney diseases	Random Forest	78.60%
		Back Propagation	80.40%
		Radial Basis Function	85.30%
`Lambodar Jena, Narendra Ku. Kamila	Chronic kidney diseases	Naïve Bayes	95%
		Multilayer perceptron	99.75%
		SVM	62%
		J48	99%
		Conjective Rule	94.75%
		Decision Table	99%
Dr. S. Vijayaran,Mr.S.Dhayanand	Acute NephriticSyndrome Chronic Kidney disease,	SVM	76.30%
	Acute Renal Failure and Chronic Glomerulonephritis	ANN	87.70%
Parul Sinha & Poonam Sinha	Chronic kidney diseases	K-Nearest Neighbour	78.75%
		SVM	73.75%
Manish Kumar	Chronic kidney diseases	Random Forest	100%
		Sequential Mininmal Optimization	95.60%
		Naïve Bayes	97.90%
		Radial Basis Function	98.80%
		Multilayer perception	98%
Abheer Y. Al-Hyari et al.	Chronic kidney diseases	Decision Tree	
K. R. Lakshmi, Y. Nagesh and M. VeeraKrishna	Kidney Failure	ANN	93.50%
		Decision Tree	78.44%
		Logistic Regression	74.74%

5. Conclusion

The main objective of this paper is to predict the Chronic Kidney Disease and analyzed the accuracy of chronic kidney disease using different DM techniques, also analyzed that there is no single classifier which produces best result for every dataset.

References

[1] 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.

[2] 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.

[3] 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, March2015.

[4] Parul Sinha, Poonam Sinha "Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM", International Journal of Engineering Research & Technology (IJERT)ISSN: 2278-0181 Vol. 4 Issue 12, December 2015.

[5] 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.

[6] Abeer, Ahmad and Majid," Diagnosis and Classification of Chronic Renal Failure Utilising Intelligent Data Mining Classifiers", International Journal of Information Technology and Web Engineering, 2014, Vol.9, No.4, pp. 1-12.

[7] K.R.Lakshmi, Y.Nagesh, M.VeeraKrishna "Performance Comparison Of Three Data Mining Techniques For Predicting Kidney Dialysis Survivability", International Journal of Advances in Engineering & Technology, ISSN: 22311963 Vol. 7, Issue 1, Mar. 2014 pp. 242-254.

[8] KoushalKumar, Abhishek "ArtificialNeural Networks for Diagnosis of Kidney Stones Disease", I.J. Information Technology and Computer Science,2012,7,20-25Published Online July 2012 in MECSDOI: 10.5815/ijitcs.2012.07.03.

[9] DSVGK Kaladhar, Krishna Apparao Rayavarapu, Varahalarao "Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis", Open Access Scientific Reportsvol.1, Issue12, 2012.

[10] Veerappan, Ilangovan and Abraham Georgi," Chronic Kidney Disease: Current Status, Challenges and Management in India", Indian Journal of Public Health Research and Development, 2014, Vol.6, No.1, pp. 1694-1702.

[11] J. Han and Kamber, "Data Mining: concepts and techniques", 2nd edition .The Morgan Kaufmann Series, 2006.
[12] http://www.webmd.com/urinary-incontinence-oab.

Author Profile

Mrs. Smitha Patil received the B.E degree in Information Science & Engineering from P.D.A College of Engineering Gulbarga, Karnataka and M.Tech in Computer Science & Engg. from VTU PG centre Gulbarga, Karnataka and currently working as Assistant Professor in the Dept. of Information Science & Engineering, SIR. MVIT Engg College Bangalore, Karnataka.