

**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY****DATA MINING TECHNIQUES FOR THE ANALYSIS OF KIDNEY DISEASE-A
SURVEY****M. Mayilvaganan^{*1}, S. Malathi² & R. Deepa³**^{*1}Associate Professor, Department of Computer Science, PSG College of Arts & Science,
Coimbatore, India²M. Phil Research scholar, PSG College of Arts & Science, Coimbatore, India³Assistant Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore,
India

DOI: 10.5281/zenodo.829799

ABSTRACT

Data mining is a vital role in several applications such as business organizations, educational institutions, and government sectors, health care industry, scientific and engineering. In the health care industry, the data mining is predominantly used for disease prediction. In India chronic kidney disease is one of the measure causes of death today. Data mining classifiers are used for prediction which can also be used in health area where a large voluminous data is generated. The healthcare industry are used the data mining techniques for predicting the kidney disease from the data set. So the different DM techniques used to identify the accuracy for the kidney related diseases.

KEYWORDS: DM techniques, Kidney disease, Classification algorithms.**I. INTRODUCTION**

The purpose of data mining is to extract useful information from large databases or data warehouse. Data mining algorithms applied in health care industry play a significant role in prediction and diagnosis of the disease. Techniques of data mining help to process the data and turn them into useful information. Early detection and accurate results are achievable by physicians using data mining algorithms. Different algorithms will be used for varied disease diagnosis. Based on the data used the accuracy and performance also vary. Data mining techniques such as classification and prediction, clustering, association rule mining and various mining methods can be useful to apply on medical data.

II. MEDICAL DATA MINING

Data mining has been used to uncover patterns from the large amount of stored information and then used to build predictive models. Medical field contains large amount of data that are needed to be processed. huge and complex volumes of data are generated by healthcare activities; un-automated analysis has become impractical. DM can generate information that can be useful to all stakeholders in health care, including patients by identifying effective treatments and best practices; Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs [13].

III. KIDNEY DISEASE

The kidneys' functions are to filter the blood. All the blood in our bodies passes through the kidneys several times a day. The kidneys remove wastes, control the body's fluid balance, and regulate the balance of electrolytes. As the kidneys filter blood, they create urine, which collects in the kidneys' pelvis -- funnel-shaped structures that drain down tubes called ureters to the bladder. Each kidney contains around a million units called nephrons, each of which is a microscopic filter for blood.

[Mayilvaganan * *et al.*, 6(7): July, 2017]ICTTM Value: 3.00

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. For most people, kidney damage occurs slowly over many years, often due to diabetes or high blood pressure. This is called chronic kidney disease. 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. This can occur in a person with normal kidneys or in someone who already has kidney problems. Kidney disease is a growing problem. 10% of the population worldwide is affected by chronic kidney disease (CKD), and millions die each year because they do not have access to affordable treatment.

There are number of factors which increase the risk of Kidney disease:Diabetes, Hypertension, Smoking, Obesity, Heart disease, Family history of Kidney disease, Alcohol intake, Drug abuse/drug overdose, Age, Race/Ethnicity, Male sex

Symptoms of kidney disease

Changes in your urinary function, Difficulty or pain during voiding, Blood in the urine, Swelling & Pain in the back or sides Extreme fatigue and generalized weakness, Dizziness & Inability to concentrate, Feeling cold all the time, Skin rashes and itching, Ammonia breath and metallic taste, Nausea and vomiting, Shortness of breath.

Diagnosis methods

Diabetic nephropathy is diagnosed using a number of tests including:

- **Urine tests** - to check protein levels. An abnormally high level of protein in the urine is one of the first signs of diabetic nephropathy.
- **Blood pressure** - regular checks for raised blood pressure are necessary. Elevated blood pressure is caused by diabetic nephropathy and also contributes to its progression.
- **Blood tests** - to check the degree of kidney function.
- **Biopsy** - a small tag of tissue is removed from the kidney, via a slender needle, and examined in a laboratory. This is usually only performed when there is doubt about whether kidney damage is due to diabetes or to another cause.
- **Kidney ultrasound** - enables the size of the kidneys to be imaged and allows the arteries to the kidneys to be checked for narrowing that can cause decreased kidney function.

Treatment options

There is no cure for diabetic nephropathy.

Treatment must become ever more aggressive as the kidneys deteriorate towards failure. Medical options include:

- **Prevention**- this is the best form of treatment and includes good control of blood glucose levels and blood pressure.
- **Medications** - including medications to reduce high blood pressure, particularly angiotensin converting enzyme (ACE) inhibitors and angiotensin receptor blockers to curb kidney damage.
- **Dialysis** -End stage kidney failure is the failure of the kidney to function at all. Dialysis involves either shunting the patient's blood through a special machine (haemodialysis) that helps remove the wastes while preserving water and salts, or removing wastes through fluid introduced into the abdomen (peritoneal dialysis). Dialysis is required several times every week for the rest of the person's life.
- **Kidney transplant** - a healthy donor kidney, obtained either from someone who has died or from a relative or friend, replaces the function of the diseased kidneys.

IV. DATA MINING ALGORITHMS AND TECHNIQUES

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

Classification

Classification is the processing of finding a set of models (or functions) which describe and distinguish data classes or concepts, for the purposes of being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is known). The derived model may be represented in various forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae, or neural networks [14].

Types of classification models

- Classification by decision tree induction
- Bayesian Classification
- Back propagation
- Support Vector Machines (SVM)
- Classification Based on Associations
- K-Nearest Neighbor Classifies
- Case Based Reasoning
- Genetic Algorithm
- Rough Set Approach
- Fuzzy Set Approach

Decision Tree based algorithm

A decision tree consists of internal nodes that represent the decisions corresponding to the hyper planes or split points (i.e., which half-space a given point lies in), and leaf nodes that represent regions or partitions of the data space, which are labeled with the majority class. A region is characterized by the subset of data points that lie in that region [15].

The decision tree is a structure that includes root node, branch and leaf node. Each internal node denotes a test on attribute, each branch denotes the outcome of test and each leaf node holds the class label. The topmost node in the tree is the root node. The decision tree approach is more powerful for classification problems.

Bayes Classification

Bayesian classifiers are known as Naive Bayesian Classifier, has comparable performance with decision tree and selected neural network classifiers. Each training example can incrementally increase /decrease the probability that a hypothesis is correct prior knowledge can be combined with observed data. Even when Bayesian methods are computationally intractable, they can provide a standard of optimal decision making against which other methods can be measured [16].

Bayes Theorem

Let H be some hypothesis that the data tuple X belongs to a specified class C, X be a data tuple.

P (H/X) - is the posterior probability of H conditioned on X.

P (H) - is the prior probability of H.

P(X/H) - is the posterior probability of X conditioned on H.

P(X) - is prior probability of X.

$$P (H/X) = \frac{P(X/H) P (H)}{P(X)}$$

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. The k -nearest neighbour algorithm, which is most often used for classification, although it can also be used for estimation and prediction K-nearest neighbour is an example of instance based learning, in which the training data set is stored, so that a classification for a new unclassified record may be found simply by comparing it to the most similar records in the training set. They can be used to model complex relationships between inputs and outputs or to find patterns in data [17].



Clustering

Clustering is the method by which like records are grouped together. g. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group and so may be considered as a form of data compression. Although classification is an effective means for distinguishing groups or classes of objects, it requires the often costly collection and labeling of a large set of training tuples or patterns, which the classifier uses to model each group. [14]

Types of clustering methods

- Hierarchical (divisive) Methods
- Partitioning Methods
- Density Based Methods
- Grid-Based Methods
- Model Based Algorithms

Categorization of Clustering Algorithms:

Algorithms are key step for solving the techniques. In these clustering techniques, various algorithms are currently in the life, still lot more are evolving. But in general, the algorithm for clustering is neither straight nor canonical:

Hierarchical methods:

- Agglomerative Algorithms
- Divisive Algorithms

Partitioning methods:

- Relocation Algorithms
- Probabilistic Clustering
- K-Medoids Methods
- K-Means Methods

Density-based algorithms:

- Density-based connectivity clustering
- Density functions clustering

Grid-based methods:

- Methods based on co-occurrence of categorical data
- Constraint-based clustering
- Clustering algorithms used in machine learning
- Gradient descent and artificial neural networks
- Evolutionary methods

Model Based Algorithms:

- Algorithms for high dimensional data
- Subspace Clustering
- Projection Techniques
- Co-Clustering Techniques

Association Rules:

Association and correlation is usually to find frequent item set findings among large data sets

Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction

An association rule is basically an expression of the form $X \rightarrow Y$. X and Y are item-sets. Control the process of association rule mining is Support and Confidence. Support is the statistical signification of a rule while Confidence is the degree of certainty of the detected associations. Support is the probability $P(X \cup Y)$,

Confidence is the conditional probability of Y being a true subject to X or $P(Y|X)$ [9]

Neural Networks

Neural networks are computing models for information processing and are particularly useful for identifying the fundamental relationship among a set of variables and patterns in the data. They can be used to model complex relationships between inputs and outputs or to find patterns in **data**. Neural network is a set of connected input/output units and each connection has a weight present with it [18].

V. COMPARISON OF DM TECHNIQUES

Author	Kidney Disease	Classifier Technique	Accuracy
K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna[1]	Kidney failure	ANN	93.85 %
		Decision Tree	78.44 %
		Logistic Regression	74.74 %
S.Ramya, Dr.N.Radha [2]	Chronic kidney diseases	Random Forest	78.6 %
		Back Propagation	80.4 %
		Radial Basis Function	85.3 %
ParulSinha&PoonamSinha[3]	Chronic kidney diseases	k Nearest Neighbour	0.7875
		SVM	0.7375
Jicksy Susan Jose [4]	Kidney Image	Association Rule	92%
		Naive Bayes	
Dr. S. Vijayarani&Mr. S. Dhayanand [5]	Kidney failure	Naive Base	70.96
		SVM	76.32
P.Swathi Baby& T. Panduranga Vital [6]	Kidney disease	AD Trees	93.9
		J48	98.11

		KStar	100
		Naive Bays	
		Random Forest	
Dr.S.Vijayarani, Mr.S.Dhayanand[7]	Acute Nephritic Syndrome, Chronic Kidney disease, Acute Renal Failure and Chronic Glomerulonephritis	SVM	0.763
		ANN	0.877
Lambodar Jena, Narendra Ku. Kamila [8]	Chronic kidney diseases	Naïve Bayes	95%
		Multilayer Perception	99.75%
		SVM	62%
		J48	99%
		Conjunctive Rule	94.75%
		Decision Table	99%

VI. CONCLUSION

Data mining has been used in various fields and the importance of it is increasing day by day. In this paper survey of various data mining algorithms are made for the kidney disease. Any one of these techniques will be implemented in future.

VII. REFERENCES

- [1] 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, and Mar. 2014 pp. 242-254.
- [2] 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.
- [3] ParulSinha, PoonamSinha "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
- [4] Jicksy Susan Jose,etal"An Efficient Diagnosis of Kidney Images Using Association Rules "International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 2.
- [5] Dr.S.Vijayarani , Mr.S.Dhayanand ,"data mining classification algorithms for kidney disease prediction" International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 4, August 2015

- [6] P. Swathi baby, T. Panduranga Vital, "Statistical Analysis and Predicting Kidney Disease Using Machine Learning Algorithms", International Journal of Engineering Research and Technology (IJERT) ISSN: 2278-018, Vol 4, Issue 07, July -2015, Pg 206-210.
- [7] 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, March 2015.
- [8] Lambodar Jena, NarendraKu.Kamila "Distributed Data Mining Classification Algorithms for Prediction of ChronicKidney-Disease", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 Vol.4, Issue11, November 2015.
- [9] Dr.R.Shanmugavadivu, N. Pavithra." Survey on Data mining Techniques used in Kidney related Diseases", International Journal of Modern Computer Science (IJMCS) ISSN: 2320-7868 (Online) Volume 4, Issue 4, August, 2016
- [10] 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, Bellingham, WA, January 2003, pp. 1-8.
- [11] V.Kirubha, S.ManjuPriya "Survey on Data Mining Algorithms in Disease Prediction" International Journal of Computer Trends and Technology (IJCTT) – Volume 38 Number 3 - August 2016.
- [12] "Data Mining Methods& Techniques" by ABM Shawhat Ali, Saleh A. Wasimi.
- [13] "Data Mining of Medical Data :Opportunities andChallenges" by Dan A. Simovici.
- [14] "Data Mining: Concepts and Techniques" by Jiawei Han and MichelineKamber.
- [15] "Data Mining and Analysis" by Mohammed j. Zaki .
- [16] "Data Mining Concepts and Techniques"Third Edition by Jiawei Han.
- [17] "Discovering Knowledge in Data" an introduction to data mining by daniel t. larose.
- [18] "Data Mining and Knowledge Discovery Handbook" by OdedMaimon.

CITE AN ARTICLE

Mayilvaganan, M., S. Malathi, and R. Deepa. "DATA MINING TECHNIQUES FOR THE ANALYSIS OF KIDNEY DISEASE-A SURVEY." *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY* 6.7 (2017): 616-22. Web. 15 July 2017.