MANTECH Publications

Journal of Data Mining and Knowledge Engineering
Volume 7, Issue 2, July-December 2022

A Review on Data Mining Methods for the Study of Kidney Disease

Charu Thakur[1], Archana Pawar[2], Geetangali Sharma[3]

Assistant Professor[1], Students[2, 3]

Department of Computer Science and Engineering

Shivdhan Singh Institute of Technology and Management (SSITM)

***Corresponding Author's Email:** -geetangalisharma95@gmail.com[3]*

Abstract

Data mining plays an important role in a variety of applications, including corporate organisations, educational institutions, and government sectors, as well as the health care industry, scientific and engineering fields. Data mining is mostly employed in the healthcare business for illness prediction. Chronic renal disease is one of the leading causes of mortality in India today. Data mining classifiers are used for prediction and may also be employed in the health field, where enormous amounts of data are created. Data mining techniques are being utilised in the healthcare industry to predict renal disease from data sets. So the many DM methodologies utilised to determine the accuracy for kidney-related disorders.

***Keywords**: DM methodologies, Classification algorithms, Data mining, Kidney disease*

INTRODUCTION

Data mining is used to extract relevant information from big datasets or data warehouses. Data mining algorithms used in the health care business play an important role in illness prediction and diagnosis. Data mining techniques aid in the processing of data and the conversion of data into meaningful information. Using data mining techniques, clinicians can achieve early detection and accurate outcomes. Different algorithms will be used to diagnose various diseases. Accuracy and performance vary depending on the data used. Medical data can benefit from data mining techniques such as classification and prediction, clustering,

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

association rule mining, and other mining approaches.

MEDICAL DATA MINING

Data mining has been used to identify patterns in massive amounts of stored data, which has subsequently been used to develop prediction models. The medical industry generates a vast volume of data that must be handled. Healthcare activities generate enormous and complex volumes of data; un-automated analysis has become impractical. Data mining can generate information that can be useful to all stakeholders in health care, including patients, by identifying effective treatments and best practices. Data mining has great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practises that improve care and reduce redundancy.

KIDNEY DISEASE

The kidneys' primary role is to filter blood. The kidneys filter all of the blood in our bodies numerous times every day. The kidneys filter waste, maintain fluid balance, and regulate electrolyte equilibrium. As the kidneys filter blood, they produce urine, which gathers in the renal pelvis — funnel-shaped structures that drain to the bladder via tubes called ureters. Each kidney includes around a million nephrons, each of which is a minuscule blood filter.

Kidney disease occurs when the kidneys get damaged and are unable to filter blood as effectively as they should. Wastes might accumulate in the body as a result of this injury. Most people develop kidney disease gradually over time, generally as a result of diabetes or high blood pressure. This is referred to as chronic renal disease. Acute kidney damage occurs when a person's kidney function suddenly changes due to sickness, accident, or the use of certain drugs. This can happen to someone who has normal kidneys or someone who already has renal difficulties. Kidney disease is becoming more prevalent. Chronic kidney disease (CKD) affects 10% of the global population, and millions die each year due to a lack of affordable treatment.

Diabetes, hypertension, smoking, obesity, heart disease, family history of kidney disease, alcohol consumption, drug abuse/drug overdose, age, race/ethnicity, male sex are all risk factors for kidney disease.

Symptoms of kidney disease

Changes in your bladder function Difficulty or discomfort during voiding,

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

Blood in the urine Swelling and soreness in the back or sides Dizziness and inability to focus, Constant coldness, skin rashes and itching. ammonia breath and a 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

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

who has died or from a relative or friend, replaces the function of the diseased kidneys.

DATA MINING ALGORITHMS AND TECHNIQUES

For knowledge discovery from databases, several algorithms and methods such as Classification, Clustering, Regression, artificial intelligence, neural networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor approach, and so on are employed.

Classification

Classification is the process of locating a collection of models (or functions) that explain and separate data classes or ideas in order to use the model to forecast the class of unknown items. The model developed is based on an examination of a set of training data (i.e., data objects whose class label is known). The resultant model can take several forms, including classification (IF-THEN) rules, decision trees, mathematical equations, and 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

Internal nodes indicate the judgments relating to the hyper planes or split points (i.e., which half-space a particular point resides in), while leaf nodes represent sections or partitions of the data space that are labelled with the majority class. An area is defined by the subset of data points that fall within it [15].

The decision tree is a structure made up of root, branch, and leaf nodes. Each internal node represents a test on an attribute; each branch represents the result of the test; and each leaf node represents the class label. The root node is the tree's highest node. For classification problems, the decision tree approach is more powerful.

Bayes Classification

Choice Bayesian classifiers, also known as Naive Bayesian classifiers, outperform decision trees and chosen neural network classifiers. Prior knowledge can be paired with observed facts to progressively raise

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

or decrease the chance that a hypothesis is right. Even when Bayesian approaches are computationally intractable, they can provide an ideal decision-making benchmark against which other methods can be assessed [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 neighbor algorithm (K-NN) is a method for categorising objects in the feature space based on the closest training data. The k-nearest neighbour algorithm is most commonly 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 can be found simply by comparing it to the most similar records in the training set.

Clustering

Clustering is the process of grouping similar records together. g. A cluster is a group of data items that are comparable to one another within the same cluster but not to objects in other clusters. A cluster of data elements can be viewed as a single group and hence as a type of data compression. Although classification is a useful method for discriminating between groups or classes of items, it necessitates the gathering and labelling of a large number of training tuples or patterns, which the classifier utilises 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 a critical stage in solving approaches. Various algorithms are now in use in these clustering approaches, and many more are in the works. However, in

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

general, the clustering technique is neither linear 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:

The goal of association and correlation is to uncover common item sets found in huge data sets.

Given a set of transactions, find rules that can predict the appearance of an item based on the appearance 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 U Y), Confidence is the conditional probability of Y being a true subject to X or P (Y|X) [9]

Neural Networks

Neural networks are information processing computer models that are particularly effective for detecting the basic link between a collection of variables and patterns in data. They can be used to detect patterns in data or to represent complicated interactions between inputs and outcomes. A neural network is a collection of connected input/output units, each with its own weight[18].

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

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 %
Parul Sinha& Poonam Sinha[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	
		J48	93.9
		KStar	98.11
		Naive Bays	100
		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%
		Conjuctive Rule	94.75%
		Decision Table	99%

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

CONCLUSION

Data mining has been employed in a variety of sectors, and its value is growing by the day. In this study, a survey of several data mining techniques for renal illness was conducted. Any of these strategies will be used in the future.

REFERENCES

1. "Data Mining: Concepts and Techniques" by Jiawei Han and MichelineKamber.
2. "Data Mining Concepts and Techniques"Third Edition by Jiawei Han.
3. 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.
4. 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.
5. "Data Mining and Knowledge Discovery Handbook" by OdedMaimon.
6. 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
7. "Data Mining of Medical Data :Opportunities andChallenges" by Dan A. Simovici.
8. 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.
9. "Data Mining and Analysis" by Mohammed j. Zaki .
10. 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
11. "Data Mining Methods& Techniques" by ABM Shawhat Ali, Saleh A. Wasimi.
12. P. Swathi baby, T. Panduranga Vital, "Statistical Analysis and

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved

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.

13. 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.
14. "Discovering Knowledge in Data" an introduction to data mining by daniel t. larose.
15. 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.
16. 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
17. 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.
18. 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.

Page 74- 82 © MANTECH PUBLICATIONS 2022. All Rights Reserved