

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 5.258



IJCSMC, Vol. 5, Issue. 5, May 2016, pg.135 – 141

REVIEW ON PREDICTION OF CHRONIC KIDNEY DISEASE USING DATA MINING TECHNIQUES

Pushpa M. Patil

Department of Computer Science

Mahatma Gandhi Shikshan Mandal's Arts, Science and Commerce College, Chopda, Maharashtra, India

pd_salunkhe@yahoo.co.in

ABSTRACT

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. In this paper I have done a review of several research papers on prediction of chronic kidney disease using data mining classifiers. In health area chronic kidney disease can be very well predicted using many classifiers in data mining.

KEYWORDS: CKD, Data mining, Classification.

INTRODUCTION

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 [1].

In India, the projected number of deaths due to chronic kidney disease was around 5.21 million in 2008 and it is expected to rise TO 7.63 MILLION IN 2020 (66.7 % OF ALL DEATHS) [2].

REVIEW OF DIFFERENT CLASSIFICATION TECHNIQUES APPLIED FOR PREDICTION OF CHRONIC KIDNEY DISEASE

- In August 2015, Dr. S. Vijayarani & Mr. S Dhayanand have considered six different attributes of renal affected disease, among those GFR i.e. Glomerular filtration Rate is a measure attribute for prediction of kidney disease. They have implemented and compared two classification techniques naïve Bayes and SVM (Support Vector Machine). Their experimental results show that SVM are more accurate than Naïve Bayes.
- In January 2016, S. Ramya and Dr. N. Radha [4] have developed a system to predict the kidney function failure by applying four classification techniques on test data from patient medical report. They have 1000 records with 15 attributes. They also compared these four techniques like Back propagation Neural network, Radial Basis Function and Random Forest. Their results show that RBF (Radial Basis Function) has better accuracy for predicting the chronic kidney disease.
- In November 2015, Lambodar Jena and Narendra ku. Kamila [5] have analyzed chronic kidney disease dataset by various classification techniques like Naïve Bayes, Multilayer Perceptron, Support Vector Machine, J48, Conjunctive Rule, Decision table. They have used a weka software. They have used 25 different attributes for classification. Their research shows that for chronic kidney disease prediction comparatively Multilayer Perceptron give higher accuracy than other techniques i.e. 99.75 % of accuracy.
- In December 2015, Parul Sinha and Poonam Sinha [6] 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.
- In July 2015 P Swathi Baby and Panduranga Vital [7] have used machine learning algorithms like AD Trees, J48, KStar, Naïve Bayes Random Forest for prediction of kidney disease. Their research shows that Naïve Bayes has the highest 100 percent accuracy.
- In October 2014 Abeer & Ahmad [8] have implemented two data mining classifiers SVM and Logistic Regression (LR) Their results showed that SVM has more accuracy than other techniques with 93.14 percent.
- In July 2015 Jurlin Rubini and Dr. P. Eswaran [9] have proposed a new chronic kidney disease dataset and implemented three classifiers radial basis function network, multilayer perceptron and logistics regression. Finally they found that multilayer perceptron has the highest accuracy than other two classifiers.

- In 2015, Ruey Key [10], he implemented three different neural network models for chronic kidney disease detection which includes (BPN) backpropagation neural network, (GRNN) generalized feed forward neural network and (MNN) modular neural network. In his research he further implemented these models by embedding (GA) genetic algorithm in to their respective neural factor. All three models in experiment have better accuracy i.e. above 85 percent. Till among these models as per observation (BPN) back propagation neural network has the highest accuracy that remaining two models.
- In February 2016, Manish Kumar [12], predicted chronic kidney disease by performing six different data mining techniques like Random Forest Classifiers, (SMO) Sequential Minimal Optimization, Naïve Bays, (RBF) Radial Basis Function, Multilayer perceptron classifier (MLPC), (SLG) Simple Logistic. He has used total 400 records for the training to prediction algorithm. Among these techniques he has found Random Forest has highest accuracy.

Table 1: Summary of Research Results

Author	Publication Year	Tool	Classifier Technique	Accuracy
Dr. S. Vijayarani & Mr. S. Dhayanand	Aug 2015	MATLAB	Naïve Base	70.96
			SVM	76.32
S.Ramya & Dr. N. Radha	Jan 2016	R Tool Weka	BPN	80
			RBF	85.3
			RF	78.6
Lambodar Jena & Narendra ku. Kamila	Nov 2015	Weka	Naïve Base	95
			MLP	99.75
			SVM	62
			J48	99
			Conjunction Rule	94.75
			Decision Table	99
Parul Sinha & Poonam Sinha	Dec 2015	Weka & Orange	SVM	73
			KNN	78
P. Swathi Baby & T. Panduranga Vital	July 2015	Weka & Orange	AD Trees	93.9
			J48	98.11
			KStar	
			Naïve Bays	100
			Random Forest	

Abeer & Ahmad	Oct 2014		SVM Logistic Regression	93.14
Ruey Kei	2015	NBuilder	BPN GRNN MNN	
Manish Kumar	Feb 2016	Weka	RF SMO Naïve Bayes RBF MLPC SLG	100 97 95 98 98 98
Jurlin Rubini and Dr. P.Eswaran	July 2015	Weka	RBF Network MLP Logistic Regression	High

Data Mining Classifiers

Data mining is used for exploring and analyzing large quantities of data in order to discover knowledge i.e. hidden facts. Mining data for prediction of various diseases is now a day's very helpful in health area. The data generated at various specialist hospitals is voluminous. There are a vast number of attributes i.e. features of the particular disease and from it the specialist will diagnose particular disease & its severity. Many research scholars have implemented various data mining techniques for prediction of many diseases.

Supervised classification means we know the number of classes and their names. There is some training data available where classes are assigned to it. So we can build the model from this available training data and may then it will be used to assign new data to a predefined class. In prediction the records are classified according to future behavior [11].

Decision Trees.

It is a predictive model. A decision tree algorithm called ID3 was introduced by J. Ross Quinlan. Afterwards C4.5 is an improvement of ID3 algorithm. Leo Breiman, Jerome Friedman, Richard Olshen & Charles Stone developed CART (Classification and Regression Trees) algorithm. J48 is based on C4.5, it is an open source

Java implementation of C4.5 algorithm. Random Forest is the collection of trees. Random Forest will classify the instances from many decision trees.

From training data (i.e. available data in which class labels are assigned) a decision tree is build which we called as a model. Then we can use this model in predicting non classified data (i.e. test data). In a decision tree each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node holds a class label.

Bayes Classification

Bayesian classifiers are known as Naïve Basian Classifier. It uses membership probabilities, like probability that a given tuple belongs to a particular class.

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.

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

Rule Bases Classification

Here if-then ules are generated to cover all the cases from training dataset. e.g.

If SCL <= 1 Then Class= KFA

If SCL < 2 Then Class= KFB

These rules directly related to the corresponding decision tree that could be created. Classification rules can also be generated from a neural network [13].

Radial Basis Function (RBF)

An RBF is a three layer neural network in which data is input to the input layer, a Gaussian activation function is used at the hidden layer and a linear activation function is used at the output layer.

BackPropagation Algorithm

BP learns by iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with the actual known target value.

Multilayer Feed Forward Neural Network.

Support Vector Machines

SVM is used to classify linear and also non linear data. In this technique the tuples of one class are separated from another class by using a decision boundary [14]. In classification SVM finds the decision boundary with maximum margin as the best hyper plane.

K Nearest-Neighbour Classifier

Here in this classifier distance between test tuple and each training tuple in n dimensional space is found. The closeness is calculated with the Eclidean distance [14].

CONCLUSION

The chronic kidney disease can be very well predicted using many classifiers in Data Mining. One can also predict the level of chronic kidney disease using classifiers. As per the observation of different experiments there are some classifiers which gave highest accuracy are Multilayer Perceptron, Random Forest, Naïve Bayes, SVM, KNN and Radial Basis Function.

REFERENCES

- [1] National Kidney Foundation (NKF), “*The Facts About Chronic Kidney Disease (CKD)*”, National Kidney Foundation, 2012, <http://www.kidney.org/kidneydisease/aboutckd>
- [2] Global Status report on non communicable disease 2010 [online] Available from www.who.int/nmh/publications/ncd_report_full.pdf.
- [3] D. S. Vijayarani, Mr. S. Dhayanand, “*Data Mining Classification Algorithms for Kidney Disease Prediction*”, International journal of Cybernetics and informatics (IJCI) Vol. 4, No. 4 August 2015.
- [4] S. Ramya , Dr. 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] Lambodar Jeena, Narendra Ku. Kamila, “*Distributed Data Mining Classification Algorithms for Prediction of Chronic Kidney Disease*”, International Journal of Engineering Research in management and Technology ISSN : 2278-9359 (Vol-4, issue-11)
- [6] 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.
- [7] 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.
- [8] Abeer, Ahmad, “*Diagnosis and Classification of Chronic Renal failure Utilizing Intelligent Data Mining Classification*”.
- [9] Jurlin Rubini, “*Generating Comparative Analysis of Early Stage Prediction of Chronic Kidney Disease*”, International Journal of Modern Engineering Research.
- [10] Ruey Key, “*Constructing Models for Chronic Kidney Disease Detection and Risk Estimation*”, IEEE International Symposium on Intelligent Control.

[11] Book. *Data Mining Techniques*, Second Edition, Michael J. A. Berry, Gordon S. Linof, Wiley Publication, India.

[12] Manish Kumar, “*Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm*”, International Journal of Compute Science and Mobile Computing, Vol 5, Issue 2, Feb-2016 Pg. 24-33.

[13] Book. *Data Mining Introductory and Advanced Topics*, Margaret H. Dunham, Pearson Publication

[14] Book. *Data Mining Concepts and Techniques*, Jiawei Han, Micheline Kamber, Jian Fei. Third Edition, MK Publication.