



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 3)

Available online at: www.ijariit.com

Predicting kidney related diseases

K Sriram

k.sriram4493@gmail.com

School of Engineering and Technology Jain University
(SET JU), Bengaluru, Karnataka

Ravilla Raviteja

ravillaraviteja@gmail.com

School of Engineering and Technology Jain University
(SET JU), Bengaluru, Karnataka

Manjunath C R

manjucl23@gmail.com

School of Engineering and Technology Jain University
(SET JU), Bengaluru, Karnataka

Sahana Shetty

sahanashe@gmail.com

School of Engineering and Technology Jain University
(SET JU), Bengaluru, Karnataka

ABSTRACT

Health care industry is observing a tremendous advancement along with upcoming innovations in the Information Technology and the Computer Science, and this is something which thrived this industry to come up with more added medical related information, which led to growth in several research sectors. Various steps are taken to handle the outburst of information related to medical sciences and acquisition of valuable knowledge. This eventually led the researchers and scientists towards applying their technical revolutions as well as inventions such as "predictive analytics", "machine learning", "big data analytics" and "learning algorithms" for gathering worthwhile understanding and support in better decision making. Big data analytics can be seen as one of the major emerging sides in the field of medical sciences. Big data is also being used for providing predictive intuitions in the healthcare field and it is also playing an important role in the analysis of chronic diseases and medical data with the help of predictive analytics. In health care industries prediction can turn out to be most useful as well as successful when the knowledge can be conveyed as action. In this case, we propose a method that gives real-time analyzed report about predicting kidney related diseases with the help of historical data and real-time data.

Keywords: *Kidney related diseases, Big data analytics, Predictive analytics, Machine learning techniques.*

1. INTRODUCTION

"Big Data sets a situation in which data sets have increased at such huge sizes that conventional technologies of information, can no longer manage them effectively, either the size or the extent and the growth of the data set". Tremendous amounts of data have led the world to be flooded with it. Thousands and Thousands of data are generated every second. Various areas, as well as their events, are tangled because of digitization, "Big Data" and "analysis of big data" always remain together. This reveals the mutual view that "Big data" is not used for stating the congestion of data, rather used for mentioning analytical implementations which help in managing the information movement and renovate into a foundation of expedient data. Scientists and researchers according to the prediction have stated that this flooding of medical related information has the capability to improvise resolutions from the clinical and care point of view. This also will enable docs to analyze each patient and gain appropriate understanding, eventually helping in healthier resolutions and outcomes.

Data mining algorithms have a useful role in health care industries in the prediction cum diagnosis of diseases. These data mining applications relate to medical device industries and pharmaceutical industries including hospital management, with the main aim of finding useful and hidden information from the database. The process of knowledge discovery includes developing, understanding, selection, the creation of data and pre-processing of data.

2. FACTORS THAT INFLUENCE IN CAUSING KIDNEY RELATED ISSUES

Some of the common factors that influence in causing kidney issues are as follows:

- High Blood Sugar: Diabetes results in damaged blood vessels with in the kidney.

- Hypertension: It increases the blood pressure and also results in hardening or weakening of kidney
- Smoking: It can have an effect on the blood pressure and there by resulting in the kidney to weaken
- Obesity: Obesity can lead to high blood sugar or high blood pressure can eventually result in weakening or damaging blood vessels within the kidney
- Heart disease: Heart is responsible for pumping blood and kidney purifies this blood which is circulating, heart-related issues can also lead to kidney related issues
- Heredity from family: some kidney diseases are as a result of family heredity.

Symptoms of kidney disease:

- Urinal functionality changes.
- Discomfort while voiding of urine
- Blood flow while voiding urine
- Inflammation as well as Soreness in back
- Exhaustion as well as faintness
- Issues related to skin
- Sickness and queasiness
- Restlessness in Breath taking.

3. VARIOUS KIDNEY RELATED DISEASES

- a) “Glomerulonephritis”: Swelling may occur in the kidney when a highly active immune system damages the kidney, this is a result of blood which comprises of urine having proteins. It can also result in failure of the kidney.
- b) “Pyelonephritis (Kidney pelvis infection)”: The main symptom for pyelonephritis which can be observed is discomfort in the back and illness, this is as a result of bacterial infection and it may spread faster in case of an infected as well as non-treated kidney bladder.
- c) “Nephrotic syndrome”: if kidney gets damaged then it pours more protein into the urine, it may also lead to edema disease (it’s a leg swelling symptom).
- d) “Nephrolithiasis (Kidney stones)”: Blockage in the flow of urine occurs as the minerals present within urine transforms into stones which can keep growing. The Results can be painful and the situation can worsen. Smaller kidney does not need any medication or treatment whereas the once which are large need treatment as well as medication

4. PROBLEM DEFINITION

Filtering the blood is one of the major functionality of kidneys. An average human being has around 1-1.5 gallons of blood circulating throughout the body and the kidneys help in filtering this blood as much as 400 times every day. It helps in cleansing the dirt and controlling the fluid stability within the body, it also regulates and keeps electrolytes stable. As the blood gets filtered the generation of urine occurs, this is segmented in the structure which is shaped like funnel also known as pelvis which exhibits the tube named ureters in the bladder. Nephrons which are said to be the filters (microscopic) of blood are present in millions of units inside a kidney. Without experiencing any of the signs or indications there is a possibility wherein the kidney and liver functionality might drop to an extent up to ninety percent.

5. PROPOSED SYSTEM

In this System, our main focus is on the machine learning techniques that help in predicting the chronic kidney diseases. The predictive models that we will be using for the above scenarios are for example “K-nearest neighbors”, “decision tree”, “support vector machine models” as well as “logistic regression”.

Machine Learning Techniques:

- 1) K -Nearest Neighbours: Searching for the data that is closest to the space arrangement helps in classifying the unknown patterns. Euclidean Distance is used in order to predict the class in case of K –Nearest Neighbours

$$d(x,y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (1)$$

Finding the k-closest patterns or examples within the space arrangement is done by Euclidean Distance. The majority voting from neighbors identifies the categories of anonymous pattern

2) Support Vector Machine: In case of classification problem one of the most commonly used “data mining technique” is Support Vector Machine which helps in predicting the class of the data.

Using the trained data finding an optimal hyper-plane amongst two classes, hyper-plane is discovered by Support Vector Machine with the help of optimal problem solution:

$$\max Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j$$

where $0 \leq \alpha_i \leq C$ for $i = 1, 2, \dots, n$ (2)

The digital function $f(x)$ which is defined as a “kernel function” is used by Support Vector Machine in order to calculate the output.

$$f(x) = \text{sign} \left[\sum_{i=1}^l \alpha_i d_i K(x, x_i) + b \right]$$

where $K(x, x_i)$ is the kernel function. (3)

3) Decision Tree: It is a classification method which is commonly used data mining tasks. The structure of Decision tree includes a root node, leaf nodes, and branch. The data is divided into classes according to the attribute present in the sample that is to be trained.

Algorithm for “Decision tree” is as follows:

Algorithm 1: Decision Tree

- Step 1: Create Node N;**
- Step 2: If tuples in D are all of the same class, C then
return N as leaf node labeled with class C;**
- Step 3: If attribute_list is empty then
return N as leaf node labelled with the
majority class in D;**
- Step 4: Apply Attribute_selection_method(D,attribute_list)
to find the “best” splitting_criterion;**
- Step 5: Label node N with splitting_criterion;**
- Step 6: If splitting-attribute is discrete-valued and multi
Waysplits allowed then
attribute_list ←
attribute_list=splitting_attribute;**
- Step 7: For each outcome j in splitting_criterion
let Dj be the set of data tuples in D
satisfying outcome j;
if Dj is empty then
Attach a leaf labelled with majority
Class in D to node N;**
- Step 8: else attach the node returned by Generate_de
cision_tree (Dj,attribute_list) to Node N;**
- Step 9: return N;**

4) Logistic-Regression: This is basically a linear regression model, Logistic Regression is used to compute the distribution between an example X and a bool class tag Y via P(X|Y).

It classifies the Boolean class label Y as follows:

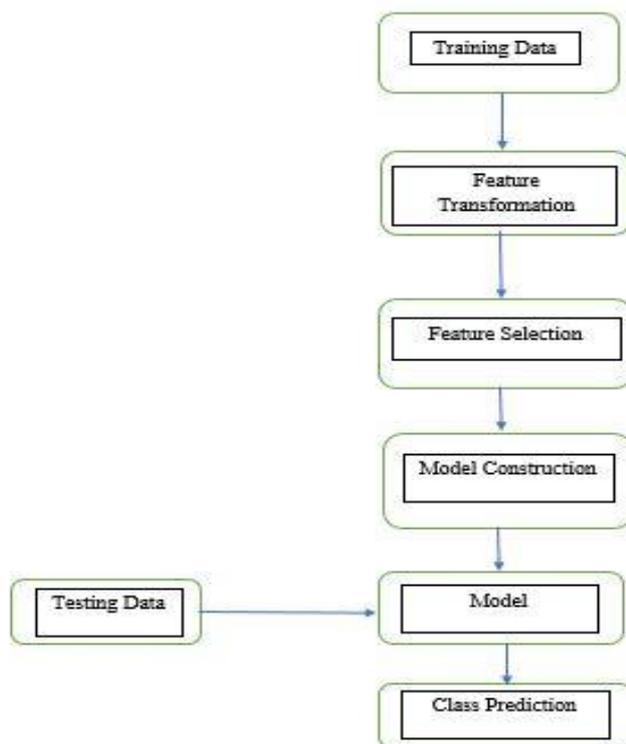
$$P(Y = 1|X) = \frac{1}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)}$$

$$P(Y = 0|X) = \frac{\exp(w_0 + \sum_{i=1}^n w_i X_i)}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)}$$

(4)

In case of our proposed system, we stress on some of the machine learning techniques in order for the prediction of kidney diseases. The main agenda of the method that is proposed is to compare the classification performance of K-nearest neighbors, decision tree, logistic regression, and support vector machine.

The proposed process for constructing the predictive models are given as follows:



- 1) Training of Data: In this step the transformation of nominal attributes to binary attributes takes place.
- 2) Feature Transformation: In this, we make use of the Best-First feature selection method in order to choose the feature subsets so as to cut down the training time as well as the number of attributes.
- 3) The Best-First is a feature selection method where it looks out for the feature subsets’ space with the help of greedy hill climbing amplified along with backtracking capability.
- 4) This step focuses on training the classifier model in order to generate a predictive model to predict the undetected data.
- 5) The final step focuses on the prediction of the class which includes kidney diseases as a result of testing the data.

6. CONCLUSION

This paper focuses on the predictive models such as K-nearest neighbors (KNN), logistic regression (LR), and support vector machine models (SVM) and decision tree classifiers. By using machine learning methods to predict kidney diseases.

7. ACKNOWLEDGEMENTS

The Authors would like to thank the Management of Jain Group of Institutions. The work has been carried out at project laboratory Department of Computer Science and Engineering, School of Engineering and Technology, Jain University.

8. REFERENCES

[1] Mohammed Shamim Rahman, Rajan Sharma, and Stephen J.D. Brecker, “Transcatheter aortic valve implantation in patients with pre-existing chronic kidney disease”, ELSEVIER International Journal of Cardiology Heart & Vasculature, Vol.5, 2015, pp. 9–18.

- [2] Ricardo T. Ribeiro, Rui Tato Marinho, and J. Miguel Sanches, "Classification and Staging of Chronic Liver Disease from Multimodal Data", IEEE transactions on biomedical engineering, Vol. 60, Issue- 5, 2013, pp.1336-134.
- [3] Anne Rogers, Anne Kennedy, Thomas Blakeman and Christian Blickem, "Non-disclosure of chronic kidney disease in primary care and the limits of instrumental rationality in chronic illness self-management", ELSEVIER Social Science & Medicine 131, 2015.
- [4] Srinivasa R. Raghavan, Vladimir Ladik, and Klemens B. Meyer, "Developing Decision Support for Dialysis Treatment of Chronic Kidney Failure", IEEE transactions on information technology in biomedicine, Vol. 9, Issue-2, 2005.
- [5] K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna, "Performance comparison of three data mining techniques for predicting kidney disease survivability", International Journal of Advances in Engineering & Technology, Mar. 2014.