International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER SSRN

IJIEMR Transactions, online available on 22nd Jul 2019. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-07

Title: **PREDICTIVE ANALYSIS TECHNIQUES OF DATA MINING FOR CHRONIC KIDNEY DISEASE**

Volume 08, Issue 07, Pages: 293–300.

Paper Authors

SHAMEEMUNNISA, Y. ANJANI

V. K. R, V. N. B AND A. G. K COLLEGE OF ENGINEERING, Gudivada

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

PREDICTIVE ANALYSIS TECHNIQUES OF DATA MINING FOR CHRONIC KIDNEY DISEASE

SHAMEEMUNNISA[1], Y. ANJANI[2]
[1]M. Tech, CSE, V. K. R, V. N. B AND A. G. K COLLEGE OF ENGINEERING, Gudivada
[2]Associate Professor, CSE, V. K. R, V. N. B AND A. G. K COLLEGE OF ENGINEERING, Gudivada

Abstract: Chronic Kidney Disease (CKD) is a noteworthy medical problem and can be restored whenever treated it in the beginning times. More often than not, individuals don't know that medical tests we take for various purposes could contain significant data concerning kidney diseases. Thusly, properties of different medical tests are explored to recognize which characteristics may contain supportive data about the disease. The goal of this paper is to utilize such properties. The data says that it encourages us to quantify the seriousness of the problem, the anticipated survival of the patient after the ailment, the example of the disease and work for restoring the disease. Subsequently we considered an informational index with various qualities that can be found by and large medical tests, AI is connected by building up a choice tree utilizing the C4.5 calculation and anticipated whether the individual is ordinary or experiencing kidney problem. This proposed model will be created utilizing Java language and is executed in Net-Beans stage.

Keywords: Chronic Kidney Disease (CKD), Decision Tree, C4.5 Algorithm, Machine Learning.

I. INTRODUCTION

Data mining is a methodology of isolating profitable information from huge proportion of dataset. Data mining has been in various territories like picture mining, appraisal mining, web mining, content mining, outline mining, etc. The employments of data mining fuse eccentricity area, money related data examination, therapeutic data examination, relational association examination, publicize examination, etc. What's more, besides it has ended up being predominant in prosperity relationship as there is an essential of explanatory system for predicting and finding dark models. Data mining expect a basic part in discovering new examples in social protection industry. It is particularly important in prosperity field when no openness of assertion supporting a particular treatment decision is found. Extensive proportion of complex data is being created by social protection industry about patients, diseases, recuperating offices, therapeutic equipment's, cases, treatment cost, etc that requires taking care of and examination for information extraction. It contains a course of action of gadgets and techniques which when associated with arranged data, gives profitable information to therapeutic administrations specialists for choosing right

decisions and that improves the execution of patient organization errands. Patients with same medical issue can be solidified together and great treatment courses of action could be given in perspective on data assembled from past patient. The overall medical issue which has been gone up against now days is consistent kidney contamination (CKD) this is the area of concern. Relentless kidney sickness is where kidneys end up clearly hurt and can't divert harmful materials in our body. Our work dominatingly focuses on recognizing life incapacitating infirmities like Chronic Kidney Disease (CKD) using Classification estimations like Naive Bayes and Artificial Neural Network(ANN) like C4.5 to predicts periods of Chronic kidney disease(CKD).

2. RELATED WORKS

AnuChaudhary et al., [1] built up a forecast framework abuse A-priori and k-implies recipe for cardiovascular disease and renal disappointment expectation. In her review An earlier and k-mean recipe calculations are wont to anticipate renal disappointment understanding with forty two properties. They dissected the data misuse AI devices like dispersion and property insights, trailed by An earlier and k-implies calculations. They assessed the data abuse Receiver operational Characteristic (ROC) plot and institutionalization plots.

Ashfaq Ahmed et al., [2] have given a piece abuse AI procedures, especially Support Vector Machine [SVM] and Random Forest [RF]. These were wont to consider, arrange and think about malignancy, liver and cardiovascular disease learning sets with variable pieces and bit parameters. Consequences of Random Forest and Support Vector Machines were looked at for different information sets like carcinoma unwellness dataset, disease dataset and cardiovascular disease dataset. It's over that variable outcomes were resolved with SVM characterization procedure with totally extraordinary portion capacities.

Giovanni Caocci et al., [3] to foresee future urinary organ Transplantation Outcome, they taken separation between a man-made Neural Network and providing Regression. Examination has been done bolstered the Sensitivity and explicitness of providing Regression and a man-made Neural Network inside the forecast of urinary organ dismissal in 10 training and supportive datasets of urinary organ transplant beneficiaries. From the exploratory outcomes that each the equation methodologies were integral and their consolidated calculations won't to improve the clinical basic leadership technique and visualization of urinary organ transplantation.

Lakshmi.K.R et al., [4] dissected Artificial Neural Networks, call tree and Logical Regression managed AI calculations. These calculations are utilized for urinary organ compound examination. For characterization technique they utilized a data mining instrument named Tanagra. The ten times cross approval is utilized to measure the arranged learning continued by the examination of these information. From the test result they assimilated that ANN performed higher than the decision tree and Logical Regression calculations.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Neha Sharma et al., [5] identified and expected urinary organ diseases as a prelude to address treatment to patients. The framework was utilized for identification in patients with kidney disease and furthermore the aftereffects of their IF-THEN standards anticipated the nearness of a disease. Their method utilized 2 fluffy frameworks and a neural system alluded to as a neural haze framework, bolstered the consequences of the information record set got. Their framework was a blend of fluffy frameworks that made outcomes misuse right numerical estimations, instead of probabilistic based for the most part orders. Generally results bolstered number juggling will in general have higher exactnesses. Their work was prepared to gain supportive learning related to enhancements in results.

Swathi Baby P et al., [6] incontestable that information handling procedures might be adequately utilized in medical applications. Their investigation gathered learning from patients influenced with excretory organ diseases. The outcomes indicated learning mining's congruity during a kind of medical applications. K-implies (KM) principle will check scope of bunches in monstrous learning sets. Their investigation broke down tree AD, J48, star K, hypothesis shrewd, irregular backwoods and tree - put together ADT innocent hypothesis with respect to J48 renal issue information Se and noticed that the procedures offer connected arithmetic examination on the usage of calculations to anticipate excretory organ diseases in patients.

VeenitaKunwar et al., [7] in their examination had predicted chronic renal issue (CKD) abuse guileless hypothesis order and counterfeit neural system (ANN). Their outcomes demonstrated that guileless hypothesis made right outcomes than fake neural systems. it had been conjointly discovered that arrangement calculations were wide utilized for examination and distinguishing proof of CKDs.

III.PROPOSED METHODOLOGY

Chronic kidney contamination (CKD) has transformed into an overall medical problem and is a zone of concern. It is where kidneys end up obviously hurt and can't divert harmful wastes in the body. Our work dominatingly focuses on recognizing life crippling infections like Chronic Kidney Disease (CKD) using Classification figurings and desire for chronic kidney sickness stages. Proposed structure is computerization for perpetual kidney disease desire using portrayal technique "Guileless bayes" and counterfeit neural framework method for stage conjecture "C4.5". In the flow work of expectation of chronic kidney disease(CKD) that utilizations innocent bayes calculation and C4.5 calculation. In which the new patient disease forecast relies upon the past patient dataset. This dataset works powerfully.

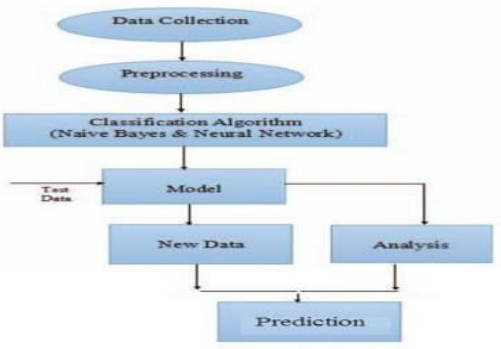

Fig 1 Flowchart of application

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Above fig demonstrates the exploratory arrangement. The procedure that includes the accumulation of patient information from different source at that point sent to pre-preparing to get the information that is picked then the information is changed into specific appropriate organization .Then the information mining strategy is connected on the information to remove valuable data and the assessment is done toward the end.

IV.PROPOSED ALGORITMS

A. Methodology for Disease prediction:

Proposed framework makes use of "Naïve Bayes Algorithm" This algorithm predicts whether the patient is suffering from ckd or notckd.

Step 1: Scan the dataset (storage servers)

Step 2: Calculate the probability of each attribute value. [n, n_c, m, p]

Step 3: Apply the formula

e P(attributevalue(ai)/subjectvaluevj)=(n_c + mp)/(n+m)

Where:

n = the number of training examples for which v = vj

•nc = number of examples for which v = vj and a = ai

- p = a priori estimate for P(aijvj)
- m = the equivalent sample size

Step 4: Multiply the probabilities by p

Step 5: Compare the values and classify the attribute values to one of the predefined set of class.

B. Methodology for Stage Prediction:

Proposed framework makes use of “C4.5” algorithm. This algorithm predicts the stages of ckd.

Step 1: Scan the dataset (storage servers)

Step 2: for each attribute a, calculate the gain [number of occurrences]

Step 3: Let a_best be the attribute of highest gain [highest count]

Step 4: Create a decision node based on a_best – retrieval of nodes[patient] where the attribute values matches with a_best.

Step 5: recur on the sub-lists [list of patient] and calculate the count of outcomes[Stages] – termed as subnodes. Based on the highest count we classify the new node.

C. C4.5 Algorithm

C4.5 algorithm is a type of classification algorithm. C4.5 is an enhancement of ID3 algorithm. It handles both discrete values and continuous values. And it is used to generate the decision tree. Improvement of C4.5 algorithm made over ID3 algorithm

Handling each continuous and distinct attributes threshold value has been created then it split the attributes according to the threshold value above and those value which are less than or the same to it.

Handling the dataset with missing attribute values also – C4.5 allows for missing value and the attribute value is marked as?, the missing attribute values part unit is simply not utilized in the calculations of gain and entropy .

Handling attributes at variance prices

Algorithm: C4.5

Input: The output of the EM algorithm as input to the C4.5 algorithm.

Output: C4.5 algorithm generates decision tree classifiers. Based on C4.5 algorithm data classification representation construct the decision. To calculate gain ratio the formula are given below.

E(S) =- pilogpi

Where, i - 1..... Count of class labels Pi – Probability of occurrence of class label in dataset.

$$I(S, A) = \sum_i^n \frac{|Si|}{|S|} E(S_i)$$

$$Splits(S, A) = -\sum_i^n \frac{|Si|}{|S|} \log\left(\frac{|Si|}{|S|}\right)$$

$$Gain(S, A) = E(S) - I(S, A)$$

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{Split(S, A)}$$

•Choosing dataset as an input to the rule for process

• Then calculate the Normalized information gain for each one attribute.

• And the attribute which has the highest information gain that attribute will be selected as best attribute.

• Then create a decision tree and attribute which has highest information gain make it as root node of that decision tree

• Repeat the process and calculate the information gain for each attribute and add that attribute as children node

V.CONCLUSION

This project is a medical sector application which helps the medical practitioners in predicting the disease types based on the symptoms. Patients can also predict diseases by entering symptoms in the form of sentences. It is automation for disease prediction and it identifies the disease, its types and complications from the clinical database in an efficient and an economically faster manner.It is successfully accomplished by applying the Naïve Bayes algorithm for classification and C4.5 for stages prediction. This classification technique comes under data mining technology. This algorithm takes symptoms as input and predicts the disease based on old patients data. Hence C4.5 is more accurate for prediction of Chronic Kidney Disease whether the patient is affected from disease or not. In this study it shows that classification strategies are used for prediction of the disease whether the patient is affected from CKD or not. The clustering techniques are used to cluster the similar kind of affected person under one cluster. This technique allows the doctors to suggest the encouraged medicine and minimize the fee. The essential aim is to decrease the cost and offer better treatment. In this project consists of only one clustering algorithm and two classification algorithm. In future work we can implement different clustering and classification algorithm for different healthcare dataset and calculating the accuracy between different algorithms and find out which algorithm is more efficient.

REFERENCES

1) AnuChaudhary, PuneetGarg, “Detecting and Diagnosing a Disease by Patient Monitoring System”, International Journal of Mechanical Engineering and Information Technology, Vol. 2 Issue 6 //June //Page No: 493-499, 2015.

2) Ashfaq Ahmed, K., Aljahdali, S., Hussain, S.N.: “Comparative prediction performance with support vector machine and random forest classification techniques”, International Journal Computer Applications. 69 (11), 12–16, 2016.

3) Giovanni Caocci, Roberto Baccoli, Roberto Littera, SandroOrrù, Carlo Carcassi and Giorgio La Nasa, “Comparison Between an Artificial Neural Network and Logistic Regression in Predicting Long Term Kidney

Transplantation Outcome", Chapter 5, an open access article distributed under the terms of the Creative Commons Attribution License, http://dx.doi.org/10.5772/53104,2017.

4) Lakshmi. K.R, Nagesh. Y and VeeraKrishna. M "Performance Comparison Of Three Data Mining Techniques For Predicting Kidney Dialysis Survivability", International Journal of Advances in Engineering & Technology, Mar., Vol. 7, Issue 1, pg no. 242-254,2016.

5) Neha Sharma, Er. Rohit Kumar Verma, "Prediction of Kidney Disease by using Data Mining Techniques", Prediction of Kidney Disease by using Data Mining Techniques, 2016.

6) Swathi Baby P and Panduranga Vital T, "Statistical Analysis and Predicting Kidney Diseases using Machine Learning Algorithms", International Journal of Engineering Research & Technology (IJERT), 2015.

7) VeenitaKunwar, KhushbooChandel, A. SaiSabitha, and AbhayBansal, "Chronic Kidney Disease Analysis Using Data Mining Classification Techniques", IEEE, 2016.

[8] RajniGarg and VikasMongia (2018), "A Comparative Study of Different Classification Algorithms on Kidney Disease Prediction", International Journal for Research in Applied Science & Engineering Technology, 6(2), pp 741-746.

[9] S.Gopika and Dr.M.Vanitha (2017), "Efficiency of Data Mining Techniques for Predicting Kidney Disease", International Journal of Engineering and Technology, 9(5), pp 3586-3591

[10] BasmaBoukenze, AbdelkrimHaqiq and HajarMousannif (2017), "Predicting Chronic Kidney Failure Disease using Data Mining Techniques", Advances in Ubiquitous Networking, Springer, pp 701-712.

[11] Duraiaj M, Ranjani V, Data Mining applications in healthcare sector a study. Int . J. Sci. Technol. Res. IJSTR, 2(10), 2013.

[12] Giovanni Caocci, Roberto Baccoli, Roberto Littera, SandroOrru, Carlo Carcassi and Giorgio La Nasa, Comparison Between an ANN and Logistic Regression in Predicting Long Term Kidney Transplantation Outcome, Chapter 5, an open access article distributed under the terms of the Creative Commons Attribution License, http://dx.doi.org/10.5772/53104, 2012.

[13] Su, J.L., Wu, G.Z., & Chao, I.P. The approach of data mining methods for medical database. In Engineering in Medicine and Biology Society, Proceedings of the 23rd Annual International Conference of the IEEE(Vol.4, pp.3824-3826), IEEE, 2001.

[14] Joshi J, Rinal D, Patel J, Diagnosis and Prognosis of Breast Cancer using Classification Rules, International Journal of Engineering Research and General Science, 2(6): 315-323, October 2014.

[15] Kumar M.N., Alternating Decision trees for early diagnosis of dengue fever .arXiv preprint arXiv: 1305.7331, 2013.

Author's Profile:

Shameemunnisa is a student of V.K.R, V,N.B & A.G.K College Of Engineering,Gudivada-521301, Andhra Pradesh. Presently she is pursuing her M.Tech [C.S.E] from this college.

Y. Anjani, M.TECH (PhD) well known Author and excellent teacher. She is currently working as Associate Professor, Department of CSE, of V.K.R, V.N.B & A.G.K College of Engineering, Gudivada-521301, Andhra Pradesh she has 11 years of teaching experience.