International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Special Issue - NGEPT - 2019

Comparative study of attributes for a performance analysis of renal disease over big data using machine learning

Ayesha, EPCET, Bangalore, India ,ayeshakgn6@gmail.com

Ms. Netra S N, Assistant Professor, EPCET, Bangalore, India ,netra.sngpr@gmail.com

Deepthi S Babu, EPCET, Bangalore, India ,deepthisathish22@gmail.com

Archana N,EPCET, Bangalore, India ,archuputti143@gmail.com

Harshitha N S, EPCET, Bangalore, India , harshithans97@gmail.com@gmail.com

Abstract: Renal disease refers to the condition of kidneys caused by conditions, diabetes, glomerulonephritis or high blood pressure. These problems may happen gently for a long period of time, often without any symptoms. It may eventually lead to kidney failure requiring dialysis or a kidney transplant to preserve survival time. So the primary detection and treatment can prevent or delay of these complications. The aim of this work is to reduce the diagnosis time and to improve the diagnosis accuracy through classification algorithms based on different attributes. Here the study of attributes for the disease prediction may vary and our aim is to improve the diagnosis time for lesser attributes compared with more attributes. We approached machine learning classifier based on polynomial linear regression. The comparison will be done for different attributes. The computation time has to improve for lesser attributes.

Keywords: Chronic Kidney Disease (CKD), Machine Learning (ML), data mining, Stages Of CKD.

I. INTRODUCTION

People with chronic kidney diseases are at higher risk of cardiovascular disease and they should be recognized early so that appropriate preemptive measures can be taken. CKD is defined as the presence of kidney damage, revealed by the abnormal albumin excretion or decreased kidney function. The disease is quantified by measured or estimated by Glomerular Filtration Rate (GFR) that persists for more than 3 month of the CKD patients. The glomerular filtration rate (GFR) is the best indicator of how well the kidneys are working. The National Kidney Foundation published treatment guidelines for identified five stages of CKD based on diminishing GFR measurements. The guidelines mention different actions based on the stage of kidney disease.

Data mining is a used for the healthcare industry to enable health systems systematically. It uses data for analytics to identify incompetence and best practices that increase the care and reduce costs. Classification rules are typically useful for medical problems that have been applied mainly in the area of medical diagnosis. Moreover, various machine learning (ML) techniques have been applied to the field of medical treatments over the past few years.

Chronic kidney disease (CKD) is a worldwide common health problem, with predictable lifetime risk of >50%, higher than that for invasive cancer, diabetes and coronary heart diseases. CKD is a long term disorder caused by damage to both kidneys [1], [2]. There is no single cause and the damage is typically permanent and can lead to ill health. In some cases dialysis or transplantation may become essential. Diabetes mellitus is also becoming more common in one cause of CKD. Chronic kidney disease is become more frequently in older people and consequently is likely to increase in the population as a whole. People with chronic kidney diseases are at higher risk of cardiovascular disease and they should be recognized early so that appropriate preemptive measures can be taken [3, 4].CKD is defined as the presence of kidney damage, revealed by the abnormal albumin excretion or decreased kidney function.

The disease is quantified by measured or estimated by Glomerular Filtration Rate (GFR) that persists for more than 3 month of the CKD patients. The glomerular filtration rate (GFR) is the best indicator of how well the kidneys are working. The National Kidney Foundation published treatment guidelines for identified five stages of CKD based on diminishing GFR measurements. The guidelines mention different actions based on the stage of kidney disease [5]. A GFR of 90 or above is considered as normal. Even with a normal GFR, it may be at increased risk for developing CKD if the patients have diabetes, blood pressure in high,

 DOI : 10.18231/2454-9150.2019.0572

or a family history of kidney disease. The risk increases with age over 65 are more than twice as likely to develop CKD as people between the ages of 25 and 65. experimental and its results. Section V gives prediction of chronic kidney diseases with various performances and its future works.

II. LITERATURE SURVEY

Miguel A. et al. [6] proposed an approach for the management of alarms related to monitoring CKD patients within eNefro project. The result proof the paragmatism of Data Distribution services (DDS) for the activation of emergency protocols in terms of alarm ranking and personalization, as well as some observations about security and privacy**.**

Christopher et al. [7] discussed a contexualized method and possibly more interpretable means of communicating risk information on complex patient populations to time constrained clinicians. Data set was collected from American Diabetes Association(ADA) of 22 demographic and clinical variables related to heart attack risk for 588 people with type 2 diabetes. The method and tool could be encomparses to other risk- assessment scenarios in healthcare distribution, such as measuring risks to patients safety and clinical recommendation compliance.

Srinivas R. Raghavan et al. [8] explored reviews the literature on clinical decision support system, debates some of the difficulties faced by practitioners in managing Chronic Kidney Disease failure patients, and sets out of decision provision techniques used in developing a dialysis decision support system.

Ricardo T. Ribeiro et al. [9] proposed a method, called Clinical Based Classifier(CBC), discriminates healthy from pathologic conditions. A large multimodal feature database was specifically built for this study. It contains chronic hepatitis, 34 compensated cirrhosis, and 36 decomposed cirrhosis cases, all validated after histopathology examination by liver biopsy. The CBC classification outperformed the nonhierarchical, one counter to all scheme, achieving better accuracy.

Mitri F.G. et al. [10] presented ultrasound-based modality complex to stiffness and free from speckle noise and owns some advantages over conventional ultrasound imaging in terms of quality.

Chih- Yin Ho et al. [11] presented a computer-aided diagnosis tool based on analyzing ultra sonography images and system could detect and classify various stages of Chronic Kidney Disease(CKD). The calculation changed over locations are reference indicators could be responsible for physicians an auxiliary and objective computer-aid diagnosis tool for CKD identification and classification.

Al- Hyari et al. [12] proposed new clinical decision support system for identifying patients with CRF. Data classification algorithms including such as Artificial Neural Network, Decision Tree and Naive Bayes are developed and applied to diagnose patients with CRF and determine evolution stage of disease. The data set containing 102 instances is collected from patient's records and used for this study. The attained results showed that the developed decision tree algorithm is most accurate CRF classifier (92.2%) when compared to all the other algorithms used in the study.

Kuo- Su Chen et al. [13] established a detection system based on computer vision and machine learning techniques for simplifying diagnosis of CKD and different stages of CKD. The proposed system required average time of 0.016 seconds for feature extraction and classification of each testing case. The results presented that the system could produce reliable diagnosis based on noninvasive ultrasonography methods and which could be measured as most proper clinical diagnosis and medical treatment for CKD patients.

Mohammed Shamim Rahman et al. [15] described the effect of chronic kidney disease(CKD) on morbidity and mortality following Trans catheter aortic valve implantation(TAVI) including patients on hemodialysis, often excluded from randomized trials. There are 118 consecutive patients underwent TAVI 63 were considered as having (CKD) and 55 not having (No-CKD) significant pre-existing CKD. The result shows TAVI is a safe, suitable treatment for patients with pre-existing CKD, though carefulness must be trained, particularly in patients with pre-existing diabetes mellitus and elevated preoperative which is associated with increased short term post-operative mortality.

III. EXISTING SYSTEM

The existing work deals with classification of different stages in chronic kidney diseases using machine learning algorithms like Naive Bayes, Decision Tree, K-Nearest Neighbour and Support Vector Machine for fixed attributes.

Drawbacks:

1. The accuracy will be less.
2. Time computation takes will be more for all attributes: Training is to be done for all the attributes which is given in the dataset. It takes more time to extract each attribute features.

Stages of Chronic Kidney Disease		
Stage	Description	Glomerular Filtration Rate (GFR)*
1	Kidney damage (e.g., protein in the urine) with normal GFR	90 or above
2	Kidney damage with mild decrease in GFR	60 to 89
3	Moderate decrease in GFR	30 to 59
4	Severe reduction in GFR	15 to 29
5	Kidney failure	Lesser than 15

Table1: Different Stages of Renal Disease

 DOI : 10.18231/2454-9150.2019.0572

IV. PROPOSED SYSTEM

The proposed system deals with the Machine Learning Classification called Polynomial linear regression algorithm, which provides more accuracy compared to the existing algorithms. The implementation of the proposed system will be on hadoop platform, hence computation speed will be increased additionally the dataset for predicting disease has lesser attributes and showing the diagnosis time in lesser.

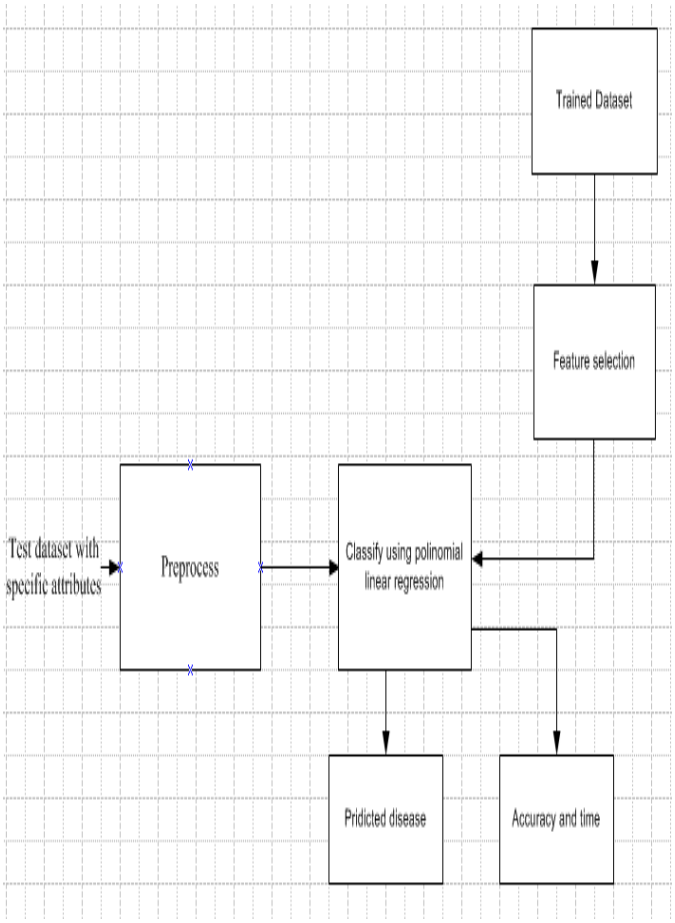

Fig 1. System architecture

A. SYSTEM REQUIREMENT SPECIFICATION

Requirement of the software Specification is a fundamental document, which forms the foundation of the software development process. Only listing the requirements of a system is not enough but also it has a description of its major feature. SRS is two-way end-to--end insurance policy that assures that both the client and the organization understand the other's requirements from that perspective at a given point in time. The organization also functions as a blueprint for completing a project with as little cost growth as possible.

B. FUNCTIONAL REQUIREMENT

Requirement defines a function of an software in a system and how the system must behave when it is presented with specific inputs or conditions. These may have the calculations, data manipulation and processing and other specific functionality. In this system following are the functional requirements:-

- Collect the data from the hospital.
- Train the Patient Data.
- Predict the Disease for patient.

C. NON-FUNCTIONAL REQUIREMENT

Non functional requirements are the requirements which are not directly concerned with the specific function delivered by the system. They specify the criteria that can be used to judge the operation of a system rather than specific behaviours. These is relate to emergent system properties such as reliability, response time and store occupancy. Requirements are collected through the user needs, because of budget constraints, organizational policies, the need for interoperability with other software and hardware systems or because of external factors such as:-

1. Product Requirements
2. Organizational Requirements
3. User Requirements
4. Basic Operational Requirements

D. PRODUCT REQUIREMENTS

Portability: Since the software is developed in java it can be executed on any platform for which the JVM is available with minor or no modifications.

Correctness: It followed a well-defined set of procedures and rules to compute and also rigorous testing is performed to confirm the correctness of the data.

Easy to Use: The front end is designed in such a way that it provides an interface which allows the user to interact in an easy manner.

Modularity: The complete product is broken up into many modules and well-defined interfaces are developed to explore the benefit of flexibility of the product.

Robustness: This software is being developed in such a way that the over all performance is optimized and the user can expect the results within a limited time with utmost relevancy and correctness. Java itself possesses the feature of robustness, which implies the failure of the system is negligible.

Non functional requirements are also called the qualities of a system. Qualities are of two types that is execution quality and evolution quality. Execution qualities are security & usability of the system which are observed during run time, whereas evolution quality involves testability, maintain ability, extensibility or scalability.

 DOI : 10.18231/2454-9150.2019.0572

Use case Diagram of the system

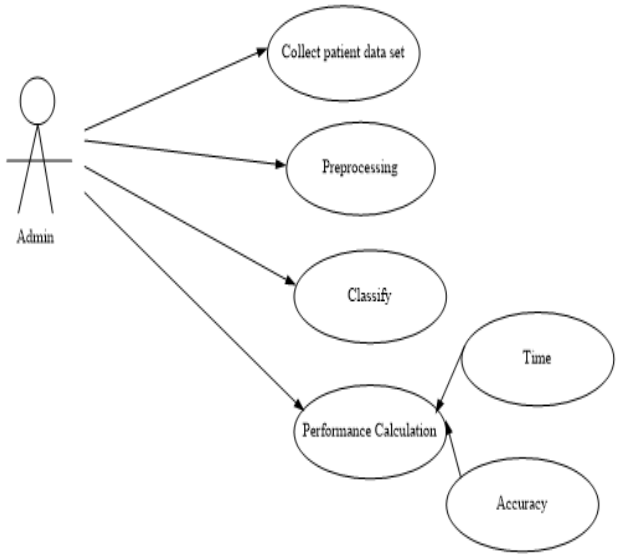

Fig 2. Use case diagram

Use case diagram is a behavioral diagram created from an analysis. It present a graphical overview of the functional requirements provided by a system in terms of actors, users, etc their goals (represented as use cases), and any dependencies between those use cases.

E. RESOURCE REQUIREMENT

Netbean IDE 7.2 : Netbean is a multi-language software development environment comprising an integrated development environment (IDE) and an extensible plug-in system. Netbean is written in Java first and can be used to develop applications in Java and, by means of the various plug-ins, in other languages as well, including C, C++, COBOL, Python, Perl, PHP, and others.

Swing: Swing is a Java Foundation Classes (JFC) consists of five major parts such as: AWT, Swing, and Accessibility, Java 2Dimensional, and Drag and Drop. Java 2Dimensional has become an integral part of AWT, Swing is built on top of AWT for making graphical front end, and Accessibility support is built into Swing.

Swing possesses several traits such as

1. Platform independence
2. Extensibility
3. Component-oriented
4. Customizable
5. Configurable
6. Look and feel.

V. ALGORITHM

Linear Regression Steps

1. Draw the scatter plot. Look for

1) Linear or non-linear pattern of the data and

2) Deviations from the pattern (outliers).

If the pattern is non-linear, consider a transformation. If there are) Deviations from the pattern, you may consider removing them only IF there is a non-statistical reason to do so. (Are those individuals "different" than the rest of the sampled individuals?)

2. The data set which has been ploted by a line to the data and check the assumptions of the model by looking at the Residual Plot (for constant standard assumption) and normal probability plot (for normality assumption).

3. Transform the data for preprocessing and to the regression line using the transformed data.

4. Once a "preprocessing" model is determined, then classify the regression line.

5. Classify the data using PLR

6. Determine if the explanatory variable is a significant predictor of the response variable by performing a E-test or P-test.

VI. RESULTS

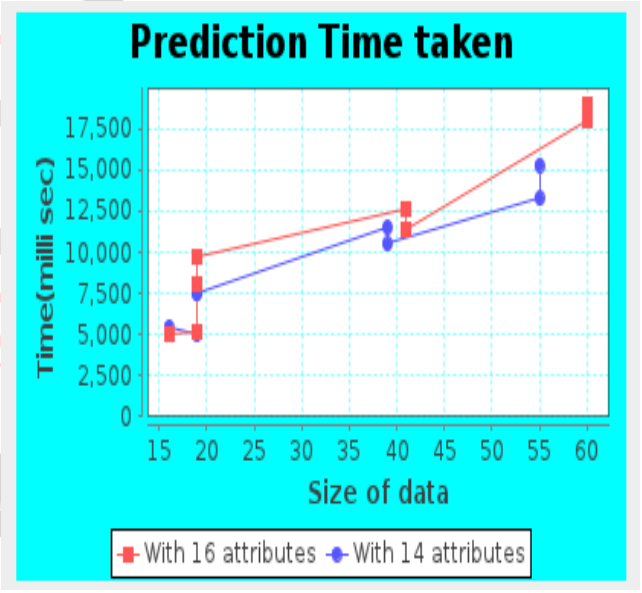

Fig 4. Comparison Graph

Information Extraction: In this patients kidney disease health care records are collected from various sources in hospitals and then organized as a electronic health care record (EHR).

Feature Selection: From among all the attributes collected for patient, the most important features for kidney prediction and modelling is extracted from the EHR.

Predictive Modelling: We construct a predictive model to say if the patient will get kidney disease or not. polynomial Linear Regression model are the predictive models. The predictive modelling part will run Big Data Hadoop Platform.

 DOI : 10.18231/2454-9150.2019.0572

VII. CONCLUSION

Accurate prediction of chronic kidney disease is one of the emerging topics in medical diagnosis. Even though some approaches using real-time features shows very good performance in terms of accuracy. This work proposes a classification model to predict the chronic kidney disease using various machine learning algorithms. All the four classification algorithms have been considered for diagnosis of chronic kidney disease. From the above results, the objective is to find the better model for chronic kidney disease. The polynomial linear regression is the better model for diagnosis of chronic kidney disease it attains the accuracy of 98%. It correctly classified the 980 instances from 1000 instances. Thus finally it is observed that PLR is better algorithm for chronic kidney diagnosis.

REFERENCES

[1] John R, Webb M, Young A and Stevens PE, "Unreferred chronic kidney disease: a longitudinal study", American Journal of Kidney Disease, Vol.5, Issue- 3, **2004**, pp.**825-35**.

[2] Coresh J, Astor BC, Greene T, Eknoyan G and Levey AS, "Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey", American Journal Kidney Disease, Vol.1, Issue- 4, **2003**, pp.**1-12**.

[3] Miguel A. Estudillo-Valderrama, Alejandro Talaminos Barroso and Laura M. Roa,"A Distributed Approach to Alarm Management in Chronic Kidney Disease", IEEE journal of biomedical and health informatics,Vol.18, Issue-6, **2014**, pp. **1796-1803**.

[4] Christopher A. Harle, Daniel B. Neill and Rema Padman, "Information Visualization for Chronic Disease Risk Assessment", IEEE Computer Society, **2012**, pp.**81-85**.

[5] 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.**

[6] Chih-Yin Ho, Tun-Wen Pai, Yuan-Chi Peng and Chien-Hung Lee, "Ultrasonography Image Analysis for Detection and Classification of Chronic Kidney Disease", IEEE conference published on Intelligent and Software Intensive Systems (CISIS),**2012**, pp.**624 – 629**.

[7] Kuo-Su Chen, Yung-Chih Chen and Yang-Ting Chen, "Stage diagnosis for Chronic Kidney Disease based on ultrasonography", IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), **2014**, pp. **525 – 530**.

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

 DOI : 10.18231/2454-9150.2019.0572