J Nephrol Ther
Volume 11:1, 2021

Journal of Nephrology & Therapeutics

Research Article

Open Access

Computational Chronic Kidney Disease for Collaborative Healthcare Data Analytics using Random Forest Classification Algorithms

V. Shanmugarajeshwari[1*] and M. Ilayaraja[2]

[1]*Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India*

[2]*Department of Computer Science and Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India*

Abstract

Computational Collaborative Healthcare data analytics is a method of methodical data analysis that allows healthcare specialists to discovery opportunities used for development in health system management processing the various information are stored. This proposed approach entails three parts comparable to preprocessing, attribute selection, classification algorithms. The goal of this work is to plan a machine-based diagnostic approach using machine learning technique. This method is developed to mining the risk factors of chronic kidney diseases. In this work, Random forest, SVM C5.0, Decision Tree, C4.5 and ANN algorithms were used to identify an early diagnosis of CKD patients. This work comparing other algorithms the best for Random forest algorithm with good accuracy and less time complexity.

Keywords: Computational collaborative healthcare data analysis • Intelligent decision support system • Computational chronic kidney • Data mining • Random forest • Support Vector machine • Artiicial neural network C5.0 • Decision tree • C4.5 • Machine learning

Introduction

Data mining is an analyzing or discovering good knowledge to develop the meaningful collection of data from huge amount of data using the knowledge. The health specifying care is the solicitation of information using machine learning algorithms. To developing also exploring healthcare data records analytical surroundings are using various methods to superior raise the value of health related problem to prediction.

To developing also exploring healthcare data records, analytical surroundings are using various methods to superior raise the value of health-related problem to prediction Healthcare record data is mostly gorgeous derived from a worldwide diversity of foundations such as sensor devices, images and text in the form of electronic records. In this miscellaneousness in the collection of data and depiction method clues to several trials in together the handling process and analysis of the original data. Worldwide assortment in the methods is essential to evaluate dissimilar forms of records. The kidneys' operations are to pass through a filter of the blood. It eliminates unwanted blood to regulate the stability of electrolytes and fluid. It strains blood; they create urine, which two bean-shaped structure of the kidney. Every one kidney surrounds a million things of unit so-called nephrons.

Health-care record data is mostly gorgeous derived from a world wide diversity of foundations such as sensor devices, images, text in the form of electronic records. In this miscellaneousness in the collection of data and depiction method clues to several trials in together the handling process and analysis of the original data. World wide assortment in the methods is essential to evaluate dissimilar forms of records [1].

The kidneys' operations are to pass through a filter the blood. It eliminates unwanted blood regulate the stability of electrolytes and fluid. It strain blood, they create urine, which two bean shaped structure of the kidney. Every one kidney surrounds a million things of unit so-called nephrons.

Factors of CKD

The following are some of the factors which lead to CKD, the main cause is diabetes and others are hypertension, smoke, fatness, heart illness, family record, alcohol, and age problem.

Symptoms

Some of the warning sign are listed down, that could be variations to urinary function, plasma in the urine, bulge & pain, severe tiredness and weakness.

Types: Acute and Chronic

- Acute-Prerenal-Kidney-Failure APKF
- Acute-Intrinsic-Kidney-Failure AIKF
- Chronic-Prerenal-Kidney-Failure CPKF
- Chronic-Intrinsic-Kidney-Failure CIKF

CKD is a worldwide health crisis. In 2019, World Health Organization agreeing to fifty eight million deaths and 35 million recognized to chronic kidney disease. The world level 850 million people now predictable to have kidney diseases from many causes, CKD causes at least 2.4 million deaths world wide-reaching per year sixth fastest growing cause of disease and death. Dialysis is a fashion of life for many patients pain with kidney sicknesses in the India. The medical record of Government of Tamil Nadu, India, Every one year 2.2 Lakh fresh patients of affected by final point renal disease or end stage renal disease. According to GBD learning, kidney disease was hierarchical 27th 1990, but rose to 18th in 2010 and 9th in

***Corresponding Author:** V. Shanmugarajeshwari, Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India, E-mail: v.shanmugarajeshwari@gmail.com*

Received: February 02, 2021; **Accepted:** February 17, 2021 ; **Published:** February 24, 2021

2019. Focuses on the development and application of machine learning algorithms for classical methods using other machine learning approaches to achieve high accuracy [2].

Figure 1 represents the various factors are affecting the patient data are evaluated with healthcare data analytics.

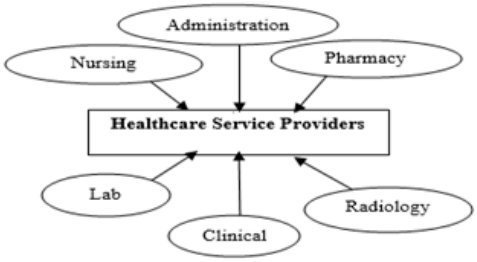

Figure 1. Affecting factors of the healthcare data analytics.

Literature review

In this collected works review mentions to a serious summary. The analyses have been done in various topic of an outline. The root of the fundamental information, everybody to building an advanced knowledge and thought for advance study perseverance.

Dowluru, et al. [3] used to affected by the patients and relate machine learning techniques to properly categorized records. Kidney nugget creation is one of the greatest collective due to altering human hierarchical generat ion methods. This paper showed with statistically and data mining method on kidney gravels and diseases. An efficient analysis and metadata using data about data scrutiny takes similarly too formed in the current task. To predicted good accuracy with C4.5, SVM. Logistic method has also good accuracy results. Bhaskar, and Suchetha, [4] the main aim of this work have been studying sensor devices to enhance the classification accuracy of the model. The future ideal effectively secret the models with an accu racy of deep learning algorithm CNN are high. Lakshmanaprabu, et al [5]. Present suggested context assembles the patient data using internet of Shankar et al. [6] have proposed to classify CKD in direction of indicate top structures aimed at cataloging procedure. This compared by surviving determining classify utmost correctness such as good accuracy of a Deep Neural Network. Zhang, Suggested model based on performance of all the model is equivalent to respectively other, both of them reach a great accurateness performance [7].

Findings

From this review, it is concrete that healthcare decision support clinical performance can be assessed by covering, machine learning techniques can be valued by various algorithms. In this survey, our research work ordered as three parts. It deal with huge datasets, using R programming language is used. This research work presents algorithm on the classification structure by various machine learning algorithm is resulted with good accuracy. In future the proposed research work has been successfully implemented in R with GUI environment.

Data set description

Data collection:

The chronic kidney data set files are composed from UCI Machine Learning Repository and it predicts CKD based on the given attributes. The dataset has thirty-two attributes that predict the CKD. To build on both numerical and nominal data types. Initially, data size is 4050 records 33 features are preprocessing; attribute variety techniques, cataloguing or classification algorithms toward spread over chronic kidney data using performance evaluation.

Methodology

The first objective is an early diagnosis of CKD patients with risk level by analyzing Chronic Kidney Disease dataset. This objective plays a valuable role in current research since many patients suffer from this disease around the world.

It represents the various parts are represented the patient data is evaluated with the collaborative healthcare data analytics block diagram of CKD.

Phase-Part 1: Preprocessing Phase, Dataset Description, Removing Missing Values, Cataloging and Categorization Methods.

Phase-Part 2: Classification Algorithms, C4.5, C5.0, ANN, Random Forest and SVM.

Phase-Part 3: CKD Prediction Performance Evaluation.

It is represented by loading CKD dateset into R envirnoment. The attributes are applied to preprocessed methods.

It mention about the preprocess and feature selection attributes are selected by R envirnoment.

Categorization

Ckd dataset using categorization process:

Input: Chronic kidney disease data

Output: Categorized value

Step 1: Choose the attributes for categorization.

Step 2: Compute the condition for the particular attributes.

Step 3: Conditional values results to be changed into the categorical values.

Step 4: Repeat steps 2-3 until all the conditions are changed for the particular attributes

Step 5: Repeat steps 1-4 until all relevant attributes have been used.

Step 6: Finally write the attributes in new csv file.

It shows the procedure for categorization process is used to convert numerical values into categorical value format.

The second objective is the power of the feature selection using machine-learning methods to detect the patients with the risk level of chronic kidney diseases while affected by particular symptoms of a particular disease.

Part 2: Feature extraction

Feature selection

It is also well-known as attribute selection. Now in choosing the applicable attributes and neglect the inappropriate attributes, many feature selection methods were applied to the preprocessed data set which has 4050 samples, concentrated on picking out all attributes.

It is attentions on seven attribute variety techniques are associated. The greatest attribute variety technique is to gain ratio feature selection that has been functional to the preprocessed records. It is a modification of the information gain that reduces its bias. It takes the number and size of branches into account when choosing an attribute [8].

The whole attribute set can be assessed by filter attribute selector to generate attribute subset.

This attribute subset can more estimate by proposition to get classification model.

• One-R

• Random Forest

• Symmetrical Uncertainty

Pseudo code for one R algorithm: for each attribute {for each value of that attribute{ compute the class distribution based on attribute value Class-label=select most frequent class create a rule : attribute = value => Class_label} Calculate the error rate of the rule on the whole dataset} Select rule with lowest error rate.

This work suggests One-R attribute selection method for attribute selection by calculating weights.

One-R method

One-R short for "One-Rule", is a simple, yet correct, classification method that makes one rule regulation for the prediction of information. To selects the rule with the smallest total error as its "one rule". This method, the attributes of the chronic kidney data to be evaluated the features are selected by individually. This method is based on error rate and missing values are performance as a benchmark for other learning schemes [9].

Symmetric uncertainty

This method is used to an information theoretical measuring called symmetric uncertainty in instruction to assess the worth of created solutions.

This process work focuses on three main feature selection methods compare than other methods best one is random forest feature selection. Random Forest attribute selection method is verified on preprocessing data. Third objective is the optimization of classical machine learning algorithms using other machine learning approaches to achieve high accuracy.

The third objective is the optimization of classical machine learning algorithms using other machine learning approaches to achieve high accuracy.

Classification: Random Forest, ANN, SVM, Decision Tree, C4.5 and C5.0 Algorithms.

Random forest algorithm

It is an important classification algorithm for random forest. These algorithms are handling continued values or categorical values. It is a supervised classification learning algorithm that works on randomly creating and merging numerous decision trees into one forestry.

Conclusion

In conclusion, chronic kidney disease study to approach emerging endorsements for machine learning techniques in healthcare has become a real world emerging for obtaining accurate results of medical diagnosis, using the machine learning techniques involved the collaborative healthcare is developing interested in a hopeful field for successful outcomes with reducing costs. Thus, system can improve the efficiency of mining risk factors of chronic kidney disease, the structure consuming machine learning methods over large volume of dataset for making better decision and prediction. The random forest feature selection is the good method for feature selection, which takes less time compares to the other feature selection methods. The reports are tested using machine learning algorithms, to predict the chronic kidney disease. Random forest decision tree classification algorithm is high accuracy resulted and less time complexity in 98.97% cataloguing accuracy.

References

1. Reddy, Chandan K., and Charu C. Aggarwal. "Healthcare Data Analytics." *CRC Press* 36 (2015).
2. Bala, Suman, and Krishan Kumar. "A Literature Review on Kidney Disease Prediction Using Data Mining Classification Technique." *Int J Comput Sci Mob Comput* 7 (2014): 960-967.
3. Krishna, K., A. Rayavarapu, and Varahalarao Vadlapudi. "Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-Analysis." *Open Access Sci Rep* 12 (2012).
4. Bhaskar, Navaneeth, and Suchetha Manikandan. "A Deep-Learning-Based System for Automated Sensing of Chronic Kidney Disease." *IEEE Sensors Letters* 10 (2019): 1-4.
5. Lakshmanaprabu, S.K., Mohanty S.N., Rani S.S, Krishnamoorthy S. and Uthayakumar J., et al. "Online Clinical Decision Support System Using Optimal Deep Neural Networks." *App Soft Comput J* 81 (2019):105487.
6. Shankar, K., P. Manickam, G. Devika, and M. Ilayaraja. "Optimal Feature Selection for Chronic Kidney Disease Classification Using Deep Learning Classifier." *In 2018 IEEE international conference on computational intelligence and computing research* (ICCIC) (2018):1-5.
7. Zhang, Hanyu, Che-Lun Hung, William Cheng-Chung Chu and Ping-Fang Chiu, et al. "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks." *In 2018 IEEE International Conference on Bioinformatics and Biomedicine* (BIBM) (2018): 1351-1356.
8. Quinlan, J. Ross. "Induction of Decision Trees." *Machine Learning* 1 (1986): 81-106.
9. Shanmugarajeshwari, V., and R. Lawrance. "A Scrutiny of Teachers' Pursuance Using Classification Techniques." *In 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing* (INCOS) (2017):1-7.
10. Novakovic, Jasmina. "The Impact of Feature Selection on The Accuracy of Naive Bayes Classifier." *In 18th Telecommunications forum* TELFOR 2 (2010):1113-1116.

How to cite this article: Shanmugarajeshwari, V and Ilayaraja M. "Computational Chronic Kidney Disease for Collaborative Healthcare Data Analytics using Random Forest Classification Algorithms." *J Nephrol Ther* 11 (2021): 344.