IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 6, June 2019

Performance Comparison of Some Classifier for Chronic Kidney Disease

Ayesha Nasuha[1]

M. Tech Student, Computer Science & Engineering, P.A College of Engineering, Mangalore, India[1]

Abstract: Nearly 10 percent of the world's population is affected by a major chronic kidney disease health issue. However, systematic and automatic methodologies are evidently used to predict chronic kidney disease. Machine training is one of the very kind methodologies. The classifier in the machine learning algorithms can provide known features and unknown class to the test samples with class labels. Existing works with machine learning algorithms do not provide the predictive accuracy to the extent required. To satisfy the gap, this initiative offers a new approach for the classification of acute renal illness with environmental variables from the medical dataset. The goal of this initiative is to forecast renal illness through the use of machine learning algorithm that supports vector machine (SVM). The primary goal of this study is to forecast renal disease using classification algorithms such as closest neighbour Optimal Fuzzy-K and Support-Vector-Machine. This study work concentrates primarily on discovering the finest classification algorithm based on the precise classification and output time variables. It is noted from the experimental outcomes that the SVM's output is a lot productive than the closest neighbouring machine Optimal Fuzzy-K. The precision is regarded to be the main measure for quality evaluation, and it is proven that the suggested technique offers greater precision in classification.

Keywords: Data mining, machine learning, chronic kidney disease

I. INTRODUCTION

Data mining is nothing but the main process of extracting information that's hidden in an informative database Predictive and descriptive are two preferred models of data mining which performs many tasks, estimation of values done by predictive model whereas descriptive model mainly identifies the relationship in data.[1,2]. Scientific analysis of data mining says that there is a different method from the nature of dataset from market-driven applications which leads to a detailed survey in data mining application in health care sector furnishing the types of data used and information details. 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 [3]. Data mining tool is found complex and time consuming instead, for advanced information findings use of a database is recommended information. In health care, data mining is a benefit in grouping patients with the similar disease or health issues so that the health care might give them effective treatments and to inform the hospitality information for patients. Recent technologies have to be established in health care sector for providing medical service in cost effective ways. These data mining techniques also analyse the factors that are responsible for the cause of disease this include type of food, living conditions, availability of pure water, etc. Health care organization develops an extensive data which is found difficult to analyse and to make a proper decision regarding patient health, treatment costs and other details of hospitals; in this case, data mining tool is helpful [4].

Kidney disease is increasing, and its prevailing seeks public attention, the high cost of treatment for this illness and the negligence may lead to cardiovascular [5]. Chronic Kidney Disease (CKD) is said to affect kidney and also causes damage to the kidney and fails to purify the blood. Kidneys functions decrease up to half its working efficiency it is said as the chronic renal failure. The advanced stage of this is End Stage Renal Disease (ESRD) results in very severe malfunction of the kidney, in this case, the function of the organ is reduced a bit, and the only possibility for survival is either transplantation or dialysis of the kidney [6].

II. EXPERIMENT

Data Mining is used in this research to perform a classification in the presence of missing data. The dataset that is used is a set of chronic kidney disease from the UCI database. The data set consist of medical examination collected from 400 patients, where 250 of them have chronic kidney disease and 150 don't have it. These two groups are classified to Chronic Kidney Disease Class (ckd) and Don’t Have Chronic Kidney Disease Class (notckd)[11]. certain feature values

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 8, Issue 6, June 2019

are missing in this data set. This research will review methods for dealing with these missing data, and to propose and implement a classifier to solve the challenge of detecting chronic kidney disease based on medical tests and to produces good classification accuracy based on cross validation [12]

Medical information mining has the ability to elevate the concealed trends in the dataset in the medical sphere. For medical treatment and prognosis, these models are used. Medical information are spread around the world, heterogeneous, in nature exaggerated [9]. To induce a user-oriented attitude to the data's novel and concealed models, the data should be coordinated together [10]. The management of the right identification of certain significant data is a significant issue in health research or bioinformatics exploration. Generally, multiple experiments require the classification or clustering of large-scale information for valued analysis purposes.

III. METHODOLOGY

It is presumed that the sample methods are vital to achieve the final diagnosis. Else more numerical trials could obscure the primary diagnostic method, which could lead to difficulties in obtaining the end outcomes, many experiments should be conducted overwhelmingly in the perceptive detection of disease [12].

Classification is one of information mining's most significant methods. To execute the classification method, it is necessary to classify the information by identifying and then placing it in a portion that is submissive by a human being [3]. This study paper defines algorithms for classification and also analyzes these algorithms' efficiency. Classification precision and implementation time are the efficiency variables used for assessment [13].

The test procedures are assumed to be essential in order to reach the ultimate diagnosis. Else, number tests could obfuscate the main diagnosis process which may result in trouble in gaining the end results, predominantly in the perceptively of finding disease many tests should be performed [12]. This sort of difficulty could be fixed with the support of machine learning which could be used directly to obtain the end result with the assistance of several artificial intelligent algorithms which perform the role as classifiers [14].

IV. SYSTEM ARCHITECTURE

Dataset: The synthetic Kidney Function Test (KFT) dataset have been created for analysis of kidney disease. This dataset contains five hundred and eighty four instances and six attributes are used in this comparative analysis. The attributes in this KFT dataset are Age, Gender, Urea, Creatinine and Glomerular Filtration Rate (GFR). This dataset consists of renal affected diseases.

Blood Urea Nitrogen: Urea is a surplus product that is eliminated by the kidneys. Nitrogen is a derivative product from urea, also eliminated by kidneys. When kidney function reduces, the BUN may be elevated.

Creatinine: this is an excess product of muscles and is normally eliminated by the kidneys. When kidney function reduces, the creatinine may be elevated.

Glomerular Filtration Rate (GFR): This is an essential measure and it is used to calculate the creatinine clearance. Normally this measure is calculated by using the following attributes; they are, age, body, sex of the patient and creatinine. This measure is considered as the best measure for finding the kidney function level and it is represented in percentage (i.e.30%).

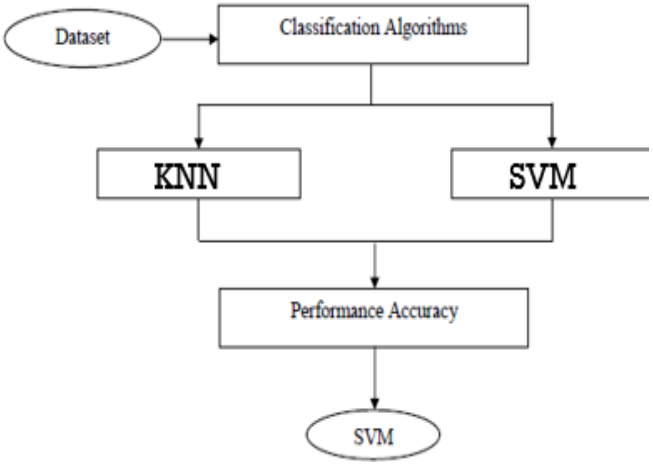

Figure 1: system architecture

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 8, Issue 6, June 2019

Classification: Classification – it maps data into predefined groups or classes as shown in Figure 1 . In classification the classes are indomitable before examining the data thus it is often mentioned as supervised learning. Classification is the process which classifies the collection of objects, datas or ideas into groups, the members of which have one or more characteristic in common[15]. In this research work KNN and SVM are used to classify different stages of Chronic Kidney Failure disease from the dataset [13].

KNN: A KNN classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model"[14]. This restricted individuality assumption infrequently clutches true in real world applications, hence the characterization as Naive yet the algorithm inclines to perform well and learn rapidly in various supervised classification problems[16]. An advantage of the KNN classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

Optimal Fuzzy KNN

Contribution: Chronic Kidney Disease is a kind of global health issue, can cause by several factors including environmental and genetic. In most of the past works, the CKD has analysed and forecasted based on the common factors without much concentration on environmental factors. Hence this work planned to include the environmental factors along with the patient's activities. In medical data prediction, the accuracy must be in acceptable range so that the system can replace a human source [15]. Hence the overall objective of this work is to develop a novel classification technique for the prediction of CKD with better accuracy [16].

Novel Method: The modelling of a novel CKD predictor is the motivated research topic in this work. An Optimal Fuzzy K-nearest neighbor (OF-KNN) technique is developed for the prediction of CKD. In the OF-KNN technique an optimization algorithm such as Bat is utilized to tune fuzzy then the OF is used to measure similarity in KNN. Thus the proposed technique can be capable of providing better accuracy and acceptable speed [17].

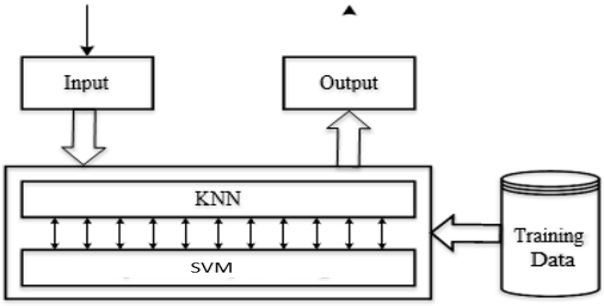

Figure 2 : Diagrammatic representation

Data set: In the implementation validation, the CKD dataset is used and serene from the UCI Machine Laboratory. It is composed of 400 instances of 24 +5 attributes. Among the 24 attributes, 11 possess numerical values, and 13 have nominal values. The remaining five attributes are the environmental factors smoking habit, drinking habit, body mass, hereditary kidney diseases and congenital kidney diseases. In this data set, 250 instances are identified as chronic kidney diseases, and the remaining are 150 are not identified. The diagrammatical representation of proposed method is portrayed in Figure 2.

Initially, the test data is feed to the input port of KNN. The KNN acts as the predictor or classifier in our work. The KNN predict the output class based on the distance between the input data and the training set. In the conventional system Euclidean distance is applied to find the nearest data in KNN [14]. To attain accurate prediction, the Euclidean distance is changed into the Optimal Fuzzy [13]. The Optimal Fuzzy system can helpful to find the optimal data which is similar to the test data. This OF-KNN system can provide improved accuracy than the conventional system, especially for the chronic kidney disease.

The steps involved in the proposed Optimal Fuzzy K-nearest neighbour (OF- KNN) technique are described below.

Steps in Optimal Fuzzy K-Nearest Neighbour (OF-KNN) the process flow of the planned methodology is shown in Figure.3.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 8, Issue 6, June 2019

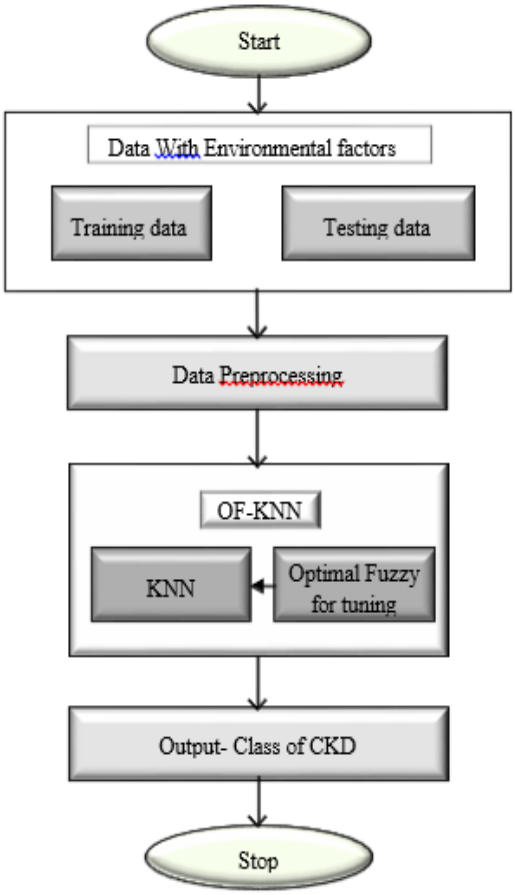

Figure 3: Flowchart of working of KNN

Data Pre-Processing: In this step, the suitable and useable format of the data is prepared, and then the knowledge extraction is applied [5]. The data pre-processing is utilized to solve the real world problems associated with the original dataset.

Examples of these problems are:

(i) Missing, and corrupted data elements
(ii) Data with noise
(iii) Different granularity measure of data
(iv) Dependent and large kind of data with irrelevant information
(v) Data form multiple sources.

V. RESULT AND CONCLUSION

This paper proposes a different approach for prediction of chronic kidney diseases with a modified dataset with five environmental factors. The methodology introduced in the paper for the prediction is OF-KNN (Optimal Fuzzy-K nearest Neighbour) and SVM accuracy. The planned work has proved as the better classifier with two different classes of diseases with perfect classification rate. The comparative analysis has done with, support vector machines and K nearest neighbours. The performance metrics utilised for the analysis are accuracy, precision, recall, specificity and F-Measure. From the analysis, it has concluded that the proposed method is well organised to obtain the perfect classification. Thus for diagnosis of CKD, the proposed OF-KNN machine learning tool is resulting in high classification accuracy rate. Figure 4 shows the comparison of accuracy of KNN and SVM.

```
ckd
ACCURACY OF KNN = 33.33333333333333
ACCURACY OF SVM = 81.81818181818183
execution time  :--- 96.24631977081299 seconds ---
```

Figure 4: accuracy comparison of KNN and SVM

Here the accuracy produced by KNN is 33.33 and accuracy produced by SVM is 6.666.as SVM produced higher accuracy than KNN we can say that SVM is more efficient than KNN. Figure 5 shows the graphical representation of KNN and SVM accuracy comparison.

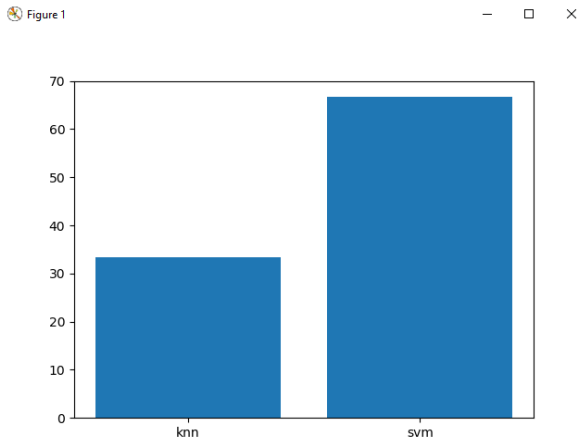

Figure 5 : Graphical Representation

ACKNOWLEDGMENT

The satisfaction of completing any task would be incomplete without mentioning the people without whom this endeavour would have been a difficult one to accomplish. I would like to extend heartfelt gratitude to **Dr. M. Sharmila kumari** , HOD, Department of Computer Science and Engineering for her continuous guidance and support in every stage of this main project. I would like to thank our Principal **Dr. Abdul Sharif f**or his support and encouragement during the process of this main project. I am indebted to the entire faculties Department of CSE for their help and co-operation. It is a great pleasure for me to acknowledge the valuable suggestions, help, and ideas of all my friends. Most importantly I thank the Almighty and my family for their blessings and continuous support.

REFERENCES

[1]. Awasthi Nikhita, Bansal Abhay, "Application Of Data Mining Classification Techniques On Soil Data Using R" International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 , 2017.
[2]. Anand V. Saurkar,"A Review Paper on Various Data Mining Tech- niques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4,Pp 98- 1o1 ,2014.
[3]. Bhargavi, P. and Dr.Jyothi, S. "Soil Classification Using Data Min- ing Techniques: A Comparative Study", International Journal of Engineering Trends and Technology, July to Aug Issue 2011
[4]. Baskar, S, S. Arockiam,L. and Charles, S. "Applying Data Mining Techniques on Soil Fertility Prediction", International Journal of Computer Ap-plications Technology and Research, Volume 2– Issue 6, 660 - 662, 2013, ISSN: 2319–8656,2013
[5]. Bhuvaneswari, S., Pramananda Perumal, T., Jagadhesan, B. "An analysis and impact factors on Agriculture field using Data Mining Techniques" International Journal of Business Intelligents Volume: o5 Issue Page No.41-44 ISSN: 2278-2400,2016.
[6]. Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E. D., & Goldschmitt, M. "Digital soil mapping using artificial neural networks. " Journal of plant nutrition and soil science, 168(1),21-33, 2005.
[7]. Chandan, Ritula Thakur , "Recent Trends Of Machine Learning In Soil Classification: A Review" International Journal of Computational Engineering Research (IJCER) ISSN (e): 2250– 3005,Volume, 08 ,2005.
[8]. Freund.Y and Schapire R, "A Short Introduction to Boosting", Journal of Japanese Society for Artificial Intelligence, 14(5):771-780,1999.2011.
[9]. Gholap,Jay "Performance Tuning Of J48 Algorithm For Prediction Of Soil Fertility" Asian Journal of Computer Science and Information Technology ISSN 2249-5126. 2011.
[10]. Jignasha,M, Jethva,Nikhil Gondaliya, Vinita Shah,"A Review on Data Mining Techniques for Fertilizer Recommendation" International Jour of Scientific Research in Comp Science, Engineering and Information Technology © 2o18 IJSRCSEIT | Vol 3 | Issue 1 | ISSN : 2456-3307, 2018
[11]. Khan, Huma, Navaz, Shahista Dr Ghosh, S, M. "A Survey on Various Data Mining Techniques in Field of Agri for Prediction of Crop Yield" International Journal of Science and Research (IJSR) ISSN (Online): 2319- 7o64, 2017.
[12]. M. Kovacevic, B. Bajat, B. Gajic, "Soil Type Classification and Estimation of Soil Properties using Support Vector Machines", Geoderma 154(3– 4), 34o– 347, 2010.
[13]. Madhuri Kommineni, Someswari Perla, Divya Bharathi Yedla, "A Survey of using Data Mining Techniques for Soil Fertility", International Journal of Engineering & Technology, 7 (2.7) (2o18) 917-918, 2018.
[14]. Murugesh Kumar, B, Dr.Ananda Kumar, K and Dr.Bharathi,(2016) A. "A survey on soil classification methods using data mining techniques", International Journal of Current Trends in Engineering & Research (IJCTER) e-ISSN 2455–1392 Volume 2 2016.
[15]. Neha Sharma, Damayanthi Sharma "Classlification & Prediction based Data Mining Techniques" Vol 4, Iss 11, Nov 2o16 International Journal of Advance Research in Computer Science and Management Studies ISSN: 2321-7782 (Online), 2016.