WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Chronic Kidney Disease Cast Using SML Technique

***R. Rakesh*[1], Dr. G. S. Nishanthi*[2], *Dr. S. Kevin Andrews*[*3]**

[*1]PG Student, Department of Computer Application, Dr. M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India. rakeshraman2906

[*2]Professor, Department of Computer Application, Dr. M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India. nishanthi.mca@drmgrdu.ac.in

[*3]Professor, Department of Computer Application, Dr. M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India. kevin.mca@drmgrdu.ac.in

DOI: https://doi.org/10.55248/gengpi.5.0524.1137

ABSTRACT—

The term “chronic renal disease” means lasting damage to the kidneys that can get worse over time. If the damage is very bad, your kidneys may stop working This is known as kidney failure or end-stage renal disease (ESRD). Patients with kidney illness may progress to the chronic stage, and chronic kidney disease (CKD) is characterised by a progressive loss of kidney function. So, doctor can diagnosing of the kidney disease patients. So, our is predicting whether patients with renal disease have entered a phase of chronic kidney disease or not by showing best accuracy result of comparing supervised classification machine learning algorithms. The aim is to investigate machine learning based techniques for CKD forecasting by prediction results in best accuracy. The full dataset will be analysed using the supervised machine learning technique (SMLT) to gather various data, including variable identification, uni-, bi-, and multivariate analysis, treatments for missing values, validation of the data, data cleaning and preparation, and data visualisation.

Keywords— Chronic renal disease, kidney failure, diagnosis, prediction , accuracy, supervised machine learning technique.

1. Introduction

Worldwide, chronic kidney disease (CKD) is a public health concern.Between 8%-10% of the adult population have some form of kidney damage, and every year millions die prematurely of complications related to CKD. Medicare costs for people with CKD were more than $23,000 per year. Glomerular Filtration Rate (GFR), which measures the kidney’s ability to filter blood, is the primary clinical tool for evaluating kidney health [1]. Chronic kidney disease (CKD) is far more prevalent worldwide than was previously assumed In western countries, it affects 10–15% of adults, and many of them need expensive medications or renal replacement therapy. Nearly 26 million Americans fit into this category, and an additional 20 million are at an increased risk of developing chronic kidney disease (CKD), according to the National Kidney Foundation Kidney Disease Report and the Third National Health and Nutrition Examination Survey. Moreover, it has been recognized that CKD is a major risk factor for increased cardiovascular disease and death [12]. Globally, chronic kidney disease (CKD) is a major health and socioeconomic issue. According to US data, there are around 5000 patients with covert illness (stages I or II) and over 200 patients with overt CKD (stages III or IV) for every patient with end-stage renal disease (ESRD) [13]. Regarded as a serious health issue, chronic kidney disease (CKD) affects over 13% of Americans. The number of people with prevalent chronic kidney disease (CKD) will keep rising due to an ageing population and an increase in patients with diabetes and hypertension. Primary care physicians will have to treat the complicated medical issues specific to individuals with chronic renal impairment as the number of CKD patients rises [14].

2. Literature Survey

John R. Sedor, MD; Robert Thomas, MD; and Abbas Kanso, MD. "Impairments Associated with Chronic Kidney Disease" [1]. Regarded as a serious health issue, chronic kidney disease (CKD) affects over 13% of Americans. The number of people with prevalent chronic kidney disease (CKD) will keep rising due to an ageing population and an increase in patients with diabetes and hypertension. Primary care physicians will have to handle the complex medical issues that are specific to patients with chronic renal impairment as the number of CKD patients rises. Nephrologists seldom take care of the medical needs of patients with chronic kidney disease (CKD) until they need renal replacement treatment, as the literature has clearly documented. In this piece, we define CKD staging and go over five CKD-related complications: osteodystrophy, anaemia, hyperlipidemia, nutrition, and cardiovascular risk.

Muhammad SalmanI, Amer Hayat KhanII, Azreen Syazril AdnanIII, Syed Azhar Syed SulaimanIV, Khalid HussainV , Naureen ShehzadiVI, Fauziah Jummaat[2]. In 2015, a tertiary-care hospital in the northeast of the Malaysian Peninsula conducted a five-year retrospective study titled "Attributable causes of chronic kidney disease in adults." Globally, the prevalence of chronic kidney disease (CKD) is rising as a medical and social issue. Insufficient

data exists in Malaysia regarding the aetiology of chronic kidney disease (CKD), a necessary condition for mitigating the illness's impact. Thus, the purpose of this study was to assess the adult population at a tertiary referral hospital's contributing factors of chronic kidney disease (CKD). There were 851 qualified cases in all. The average age of the patients was 13.37 ± 61.18 years. 333 cases (39.1%) had CKD stage V, while 240 cases (28.2%), 186 cases (21.9%), 74 cases (8.7%), and 18 cases (2.1%) had stages IV, IIIb, IIIa, and II, respectively. The proportion of CKD stage V patients receiving renal replacement therapy was 15.6%. The foremost attributable causes of CKD were diabetic nephropathy (DN) (44.9%), hypertension (HPT) (24.2%) and obstructive uropathy (9.2%). The difference in the prevalence of CKD due to DN, HPT and glomerulonephritis between patients ≤ 50 and > 50 years old was statistically significant. According to our findings, the two main factors that can be linked to CKD in patients at a tertiary care hospital in Malaysia are DN and HPT. Moreover, the findings raise the likelihood that a stronger focus on primary prevention of diabetes and hypertension will significantly lower the number of CKD hospital admissions in Malaysia.

Mirjana Sabljar Matovinovic. “PATHOPHYSIOLOGYAND CLASSIFICATION OF KIDNEY DISEASES” [3].Chronic kidney disease (CKD) is far more prevalent worldwide than was previously assumed. In western countries, it affects 10–15% of adults, and many of them need expensive medications or renal replacement therapy. Nearly 26 million Americans fit into this category, and an additional 20 million are at an increased risk of developing chronic kidney disease (CKD), according to the National Kidney Foundation Kidney Disease Report and the Third National Health and Nutrition Examination Survey. Furthermore, it is well known that CKD is a significant risk factor for an elevated risk of cardiovascular disease and mortality. The 2007 European recommendations for the Management of Arterial Hypertension and the most recent cardiologic recommendations both take this information into account. At the same time, conditions like obesity, diabetes, and hypertension that increase a person's risk of developing chronic kidney disease are becoming more common and Early CKD identification is a top goal for healthcare in both developed and developing nations.

Robert Centor, MD, Barbara J. Turner, MD, MSED, Raquel C. Greer, MD, MHS, Michael Choi, MD, Joseph A. Vassalotti, MD, and Thomas D. Sequist, MD. "The Primary Care Clinician's Practical Guide to the Identification and Treatment of Chronic Kidney Disease"[4]. This useful method was created by a panel of internists and nephrologists for the Kidney Disease Outcomes Quality Initiative to help primary care physicians with the diagnosis and treatment of chronic kidney disease (CKD). A glomerular filtration rate (GFR) of less than 60 mL/min/1.73 m2 and/or kidney damage indicators for a minimum of three months are indicative of chronic kidney disease. GFR computed from the serum creatinine concentration (eGFR) and albuminuria from the urinary albumin-to-creatinine ratio are the most often used diagnostics for CKD in clinical practice. People with diabetes and/or hypertension should have their eGFR and albuminuria assessed; however, this is not advised for the general public. The goal of CKD management is to lower the patient's risk of developing the illness's associated consequences, including anaemia, metabolic acidosis, acute renal damage, cardiovascular disease, and mineral and bone disorders. Blood pressure <140/90 mm Hg, the use of angiotensin-converting enzyme inhibitors or angiotensin receptor blockers for patients with albuminuria and hypertension, haemoglobin A1c <=7% for patients with diabetes, and the correction of CKD-associated metabolic acidosis are all necessary to prevent the progression of chronic kidney disease (CKD). When prescribing, the level of eGFR should be taken into account, and nephrotoxins, such as nonsteroidal anti-inflammatory drugs, should be avoided in order to limit the risks to the patient's safety.

Lai Seong Hooi, Rashidah Ambak, Fatimah Othman, Hamizatul Akmal Abd Hamid, Tahir Aris, Irene Wong, Halizah Mat Rifin, Tania Gayle Robert, Hasimah Ismail, Norazizah Ibrahim Wong, Ghazali Ahmad, Sunita Bavanandan, Wan Shakira Rodzlan Hasani, and Loke Meng Ong. The findings of a countrywide population-based cross-sectional investigation on the prevalence of chronic kidney disease and its related variables in Malaysia are presented in reference [5]. Based on our research, the prevalence of chronic kidney disease (CKD) in Malaysia increased from 9.07% in 2011 to 15.48% (95% CI: 12.30, 19.31) in 2018. 3.85% of the projected cases were of stage 1 CKD, 4.82% of stage 2 CKD, 6.48% of stage 3 CKD, and 0.33% of stage 4-5 CKD. Significant correlations were found between CKD and hypertension (aOR 3.72), diabetes mellitus (aOR 3.32), rising BMI (aOR 1.06), and ageing (aOR 1.06). According to our research, CKD has emerged as one of Malaysia's most important public health concerns. Therefore, at the national level, there is an urgent need to screen for CKD and prevent its progression, related morbidity, and mortality.

3. METHODOLOGY

3.1 Data Pre-processing:

The machine learning (ML) model's error rate, which is as near to the actual error rate of the dataset as possible, is obtained through validation approaches. You might not require the validation approaches if the volume of data is sufficiently large to be representative of the population. Working with data samples that might not be an accurate representation of the population in a particular dataset is common in real-world situations. Identifying duplicate values, missing values, and the data type—integer or float—must be done. the data sample that is utilised to tune model hyperparameters while offering an objective assessment of a model fit on the training dataset. As skill from the validation dataset is included into the model setup, the evaluation gets increasingly skewed. A specific model is assessed using the validation set, although this is done on a regular basis. This data is used by machine learning experts to adjust the hyperparameters of the model. A time-consuming to-do list may result from data collection, analysis, and the process of addressing the content, quality, and structure of the data. Understanding your data and its characteristics is helpful during the data identification phase since it will guide your decision regarding which algorithm to employ when creating your model. Several distinct Specifically, it focuses on missing values—possibly the largest data cleaning task—and can clean data more quickly by utilising Python's Pandas package. It would rather spend more time researching and modelling than cleaning data.

A few of these sources are merely careless errors. In some cases, missing data may have a deeper cause. It's critical to comprehend these various kinds of missing data from a statistical perspective. The nature of the missing data will determine how to handle detecting and filling in the gaps, as well as

how to handle basic imputation and more complex statistical methods. It's critical to comprehend the origins of any missing data before diving into the code. The following are some common causes of missing data:

- When manually moving data from a legacy database, some was lost.
- A programming error occurred.
- Due to their concerns regarding the use or interpretation of the results, users choose not to fill out a particular field.

Variable identification with Uni-variate, Bi-variate and Multi-variate analysis:

- Import libraries for access and functional purpose and read the given dataset
- General Characteristics of the Analysed Dataset: • Present the dataset as a data frameshow columns
- shape of the data frame
- To describe the data frame
- Verifying the type of data and dataset details
- Checking for duplicate data
- Checking Missing values of data frame
- Checking unique values of data frame
- Checking count values of data frame
- Rename and drop the given data frame
- To specify the type of values
- To create extra columns

3.2 DataValidation/ Cleaning/Preparing Process

Loading the specified dataset while importing the library packages. Analysing the variable identification by data type and shape, assessing duplicate and missing values, etc. A validation dataset is a portion of data that was withheld from your model's training process. It is intended to provide an estimate of model skill during model tuning. You can utilise processes to optimise the utilisation of test and validation datasets during model evaluation. Renaming the provided dataset, removing a column, and other actions are examples of data cleaning and preparation for uni-, bi-, and multi-variate analysis. Data cleaning procedures and methods differ depending on the type of dataset. Finding and eliminating mistakes and abnormalities is the main objective of data cleaning in order to raise the value of data in analytics and decision making.

3.3 Exploration data analysis of visualization

In applied statistics and machine learning, data visualisation is a critical competency. In fact, the main focus of statistics is on quantitative data descriptions and estimations. An essential set of tools for developing a qualitative understanding is provided by data visualisation. This can be useful for discovering trends, faulty data, outliers, and much more while examining and getting to know a dataset. Plots and charts that are more visceral and stakeholders than measures of association or significance can express and illustrate important relationships with the use of data visualisations, provided the user has some topic knowledge. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end.

Sometimes data does not make sense until it can look at in a visual form, such as with charts and plots. Being able to quickly visualize of data samples and others is an important skill both in applied statistics and in applied machine learning. It will reveal the various plot types that you should be aware of when using Python to visualise data and show you how to apply them to enhance your understanding of your own data.

IV. ALGORITHMS

Logistic Regression

A dichotomous variable—one in which there are only two possible outcomes—is used to quantify the result. Finding the best fitting model to explain the relationship between a set of independent (predictor or explanatory) factors and the dichotomous feature of interest (dependent variable = response or outcome variable) is the aim of logistic regression To determine the probability of a categorical dependent variable, logistic regression, a machine learning classification technique, is employed.The dependent variable in logistic regression is a binary variable that has information coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).

Support | Vector Machine

a classifier that divides the data set into groups according to the best hyperplane between each group of data. I choose this classifier since it can produce a high prediction rate and is quite flexible in terms of the range of kernelling functions that may be used. Perhaps one of the most well-known and discussed machine learning methods is support vector machines. They were quite well-liked when they were first invented in the 1990s, and people still use them when they want a high-performing algorithm that requires little adjustment.

- How to sort through the many names for support vector machines.
- The representation that SVM use for storing the model locally on a disc.
- How a learned SVM model representation can be used to make predictions for new data.
- How to learn an SVM model from training data.
- The optimal way to get your data ready for the SVM algorithm.
- Where you might look to get more information on SVM.

Gaussian Naïve Bayes

- The Naive Bayes algorithm is a user-friendly technique that predicts using the likelihoods of each attribute belonging to each class. It is the supervised learning strategy you would devise in order to simulate a probabilistic predictive modelling problem.
- By assuming that the probability of each attribute belonging to a given class value is independent of all other characteristics, Naive Bayes simplifies the calculation of probabilities. Although this is a significant assumption, the technique is quick and efficient.
- Conditional probability is the likelihood of a class value given an attribute value. We may calculate the chance of a data instance belonging to a class by multiplying the conditional probabilities for each attribute for a particular class value that class.
- We can determine the class value with the highest probability and use that result to construct a forecast by calculating the probabilities of each occurrence belonging to each class.
- The statistical classification method known as Naive Bayes is founded on the Bayes Theorem. Among the most basic supervised learning algorithms is this one.The algorithm that is quick, accurate, and dependable is the naive Bayes classifier. On big datasets, naive Bayes classifiers perform quickly and accurately.
- Idle The Bayes classifier makes the assumption that a feature's influence within a class is unaffected by other characteristics. For instance, a loan applicant's desirability varies based on age, region, income, and past loan and transaction history.
- These traits are nonetheless regarded as separate even though they may be linked.

This assumption is regarded as naive because it makes calculation easier. We refer to this presumption as class conditional independence.

Random Forest Classifier

During training, a large number of decision trees are constructed, and at the end, the class that represents the mean prediction (regression) or mode of the classes (classification) of each individual tree is output. Random forests, also known as random decision forests, are an ensemble learning technique for classification, regression, and other tasks. The tendency of decision trees to overfit to their training set is corrected by random decision forests. An algorithm for supervised machine learning based on ensemble learning is called random forest. In ensemble learning, distinct algorithm types or iterations of the same algorithm are combined to create a more potent prediction model. The random forest algorithm creates a forest of trees by combining many algorithms of the same sort, such as multiple decision trees. Thus, "Random Forest" was named. The random forest algorithm is applicable to applications involving both classification and regression.

The random forest algorithm can be executed using the following fundamental steps:

- Select N records at random from the dataset.
- On the basis of these N records, create a decision tree.
- Select the desired number of trees for your algorithm, then go back and complete steps 1 and 2.

In a regression issue, every tree in the forest predicts a value for Y (output) for a new record. The average of all the values that each tree in the forest projected can be used to determine the final value. Alternatively, every tree in the forest predicts the category to which the new record belongs in the event of a classification challenge. Ultimately, the category with the majority vote is given the new record.

V. RESULTSAND DISCUSSION

Comparing Algorithm with prediction in the form of best accuracy result:

The K-fold cross validation procedure is used to evaluate each algorithm, importantly configured with the same random seed to ensure that the same splits to the training data are performed and that each algorithm is evaluated in precisely the same way. Before that comparing algorithm, Building a Machine Learning Model using install Scikit-Learn libraries. In this library package have to done preprocessing, linear model with logistic regression method, cross validating by KFold method, ensemble with random forest method and tree with decision tree classifier. Additionally, splitting the train set and test set. To predicting the result by comparing accuracy.

Prediction result by accuracy:

Logistic regression algorithm also uses a linear equation with independent predictors to predict a value. The predicted value can be anywhere between negative infinity to positive infinity. It need the output of the algorithm to be classified variable data. Higher accuracy predicting result is logistic regression model by comparing the best accuracy.

True Positive Rate(TPR) = TP / (TP + FN)

False Positive rate(FPR) = FP / (FP + TN)

Accuracy:

The Proportion of the total number of predictions that is correct otherwise overall how often the model predicts correctly defaulters and non-defaulters.

Accuracy calculation:

The easiest performance metric to understand is accuracy, which is just the ratio of properly predicted observations to total observations. If our accuracy is high, one might assume that our model is the best. Indeed, accuracy is an excellent statistic, but only in cases when the datasets are symmetric, meaning that the false positive and false negative values are almost equal.

Precision:

the percentage of optimistic forecasts that come true.

TP / (TP + FP) equals precision.

The ratio of accurately anticipated positive observations to all predicted positive observations is known as precision. How many of the passengers who were reported as have survived actually did? is the question that this measure attempts to address. Good accuracy is associated with a low false positive rate. Our precision is 0.788, which is quite good.

Recall:

The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)

Recall = TP / (TP + FN)

The ratio of accurately predicted positive observations to all observations in the actual class is known as recall (sensitivity).

F1 Score:

It is the Precision / Recall weighted average. As such, this score accounts for both false positives and false negatives. F1 is typically more valuable than accuracy, even if it is intuitively harder to understand—especially in cases when the distribution of classes is not uniform. When the cost of false positives and false negatives is comparable, accuracy performs best. It is preferable to include both Precision and Recall if the costs associated with false positives and false negatives are significantly different.

General Formula:

F- Measure = 2TP / (2TP + FP + FN)

F1-Score Formula:

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

False Positives (FP):

A payer who is anticipated to default. when the expected class is yes and the actual class is no. For example, if the actual class reports that the passenger did not survive, but the forecast class indicates that they will.

False Negatives (FN):

A defaulter who was anticipated to pay. when the anticipated class is not the actual class. For example, if the predicted class informs you that the passenger will die and the actual class result shows that this passenger survived.

True Positives (TP):

A nonpaying individual who is anticipated to default. These are the positively predicted values that were accurate, indicating that both the anticipated and actual values for the class are yes. For example, if the projected class and actual class values both suggest that this passenger survived, then you know the same thing.

True Negatives (TN):

A defaulter who was anticipated to pay. These are the accurately predicted negative values, indicating that both the actual and anticipated values for the class are zero. For example, if the anticipated class reports the same thing and the actual class reports that the passenger did not survive.

Used Python Packages:

sklearn:

- Sklearn is a machine learning library for Python that contains numerous machine learning methods.
- Here, we are using some of its modules like train_test_split, DecisionTreeClassifier or Logistic Regression and accuracy_score.

NumPy:

- It is a numeric python module which provides fast maths functions for calculations.
- It is utilised for data manipulation and reading from Numpy arrays.

Pandas:

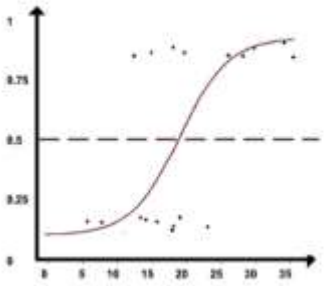
- Used to read and write different files.
- Data manipulation can be done easily with data frames.

Matplotlib:

- Finding trends in a given dataset can be aided by the use of data visualisation.
- Data manipulation can be done easily with data frames.

1.Logistic Regression

This statistical technique is employed to analyse a set of data where an outcome is determined by one or more independent factors. A dichotomous variable—one in which there are only two possible outcomes—is used to quantify the result. Finding the best fitting model to explain the relationship between a set of independent (predictor or explanatory) factors and the dichotomous feature of interest (dependent variable = response or outcome variable) is the aim of logistic regression. A machine learning classification approach called logistic regression is used to estimate the likelihood of a categorical dependent variable. The dependent variable in logistic regression is a binary variable that has information coded as 1 (yes, success, etc.) or 0(no,failure,etc.).

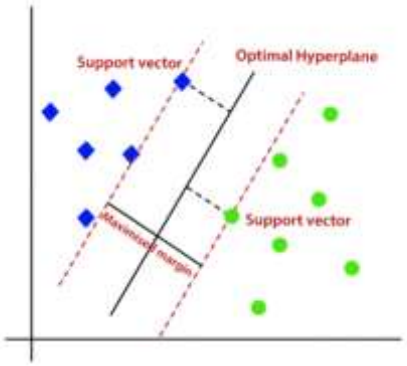


$$J(\Theta) = -\frac{1}{m}\sum_{m}^{i=1}(y^i \log(p^i) + (1 - y^i)\log(1 - p^i))$$

m is the number of samples in the training data. y^i is the label of the i-th sample, p^i i is the prediction value of the i-th sample. When the current sample's label is 1, then the second term of the formula is 0. We hope the larger the first term, the better, and vice versa. Finally, we add the loss of all samples, take the average, and add a negative sign. We want to minimize the quadratic cost function $J(\Theta)$. When $J(\Theta)$ is smaller, it means that the model fits better on the data set. There is no closed-form method to find Θ. To achieve this goal, we need to use some optimization algorithms, such as gradient descent. Since $J(\Theta)$ is a convex function, the gradient descent is guaranteed to

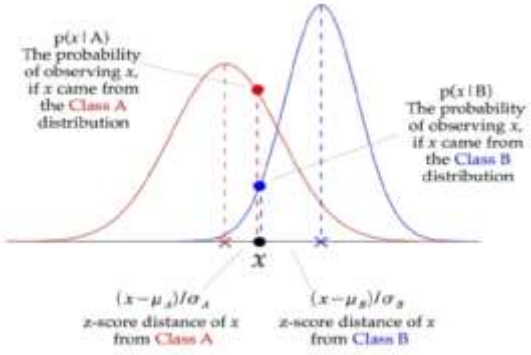
2.Support Vector Machine

A classifier that divides the data set into groups according to the best hyperplane between each group of data. I choose this classifier since it can produce a high prediction rate and is quite flexible in terms of the range of kernelling functions that may be used. Perhaps one of the most well-known and discussed machine learning methods is support vector machines



3.Gaussian Naive Bayes

The probability of each attribute belonging to each class are used by the Naive Bayes algorithm, an intuitive technique, to generate predictions. It is the supervised learning strategy you would devise in order to simulate a probabilistic predictive modelling problem.By assuming that the probability of each attribute belonging to a given class value is independent of all other attributes, Naive Bayes simplifies the calculation of probabilities. Although this is a significant assumption, the technique is quick and efficient.



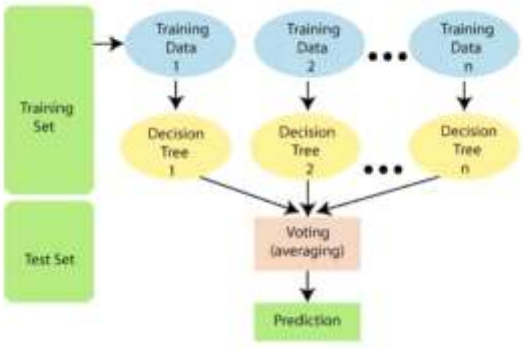
$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

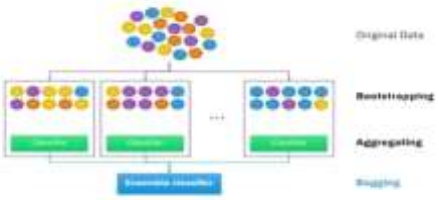
Sometimes assume variance

- is independent of Y (i.e., σi),
- or independent of Xi (i.e., σk)
- or both (i.e., σ)

4.Random Forest Classifier

During training, a large number of decision trees are constructed, and at the end, the class that represents the mean prediction (regression) or mode of the classes (classification) of each individual tree is output. Random forests, also known as random decision forests, are an ensemble learning technique for classification, regression, and other tasks. The tendency of decision trees to overfit to their training set is corrected by random decision forests. An algorithm for supervised machine learning based on ensemble learning is called random forest. In ensemble learning, distinct algorithm types or iterations of the same algorithm are combined to create a more potent prediction model. The random forest algorithm creates a forest of trees by combining many algorithms of the same sort, such as multiple decision trees. Thus, "Random Forest" was named. The random forest algorithm is applicable to applications involving both classification and regression.





ALGORITHM	LOGISTIC REGRESSION	SUPPORT VECTOR MACHINE	naïve BAYES	RANDOM FOREST CLASSIFIER
ACCURACY	94.5	62.5	92.7	98.25
PRECISION	0.96	0.39	0.96	1.0
RECALL	0.96	0.62	0.96	1.0
F1 SCORE	0.96	0.48	0.96	1.0

VI. Conclusions

Data cleaning and processing, missing value analysis, exploratory analysis, and model construction and evaluation were the first steps in the analytical process. Higher accuracy scores on public test sets indicate the best accuracy, which will be determined. The Chronic Renal Disease Prediction can be found with the aid of this application.

VII. FUTURE WORKS

- Hospitals want to automate the detecting disease persons from eligibility process (real time) based on the account detail.
- To show the forecast result in a desktop or web application to automate this process.
- To maximise the effort put into implementing the task in an AI environment.

VIII. References

[1] Haishuai Wang, Benjamin Bowe*, Zhicheng Cui Hong Yang*, S. Joshua Swamidass, Yan Xie, and Ziyad Al-Aly, "A Deep Learning Approach for the Estimation of Glomerular Filtration Rate", IEEE access,vol-10,pp.1109/TNB.2022.3147957.*

[2] S. N. Induja and C. G. Raji, "Computational methods for predicting chronic disease in healthcare communities," in Proceedings of the 2019 International Conference on Data Science and Communication (IconDSC), pp. 1–6, IEEE, Bangalore, India, March 2019.

[3] B. Manjulatha and P. Suresh, "An ensemble model for predicting chronic diseases using machine learning algorithms," in Smart Computing Techniques and Applications, pp. 337–345, Springer, New York, NY, USA, 2021.

[4] R. Ge, R. Zhang, and P Wang, "Prediction of chronic diseases with multi-label neural network," IEEE Access, vol. 8, pp. 138210–138216, 2020.

[5] I. Preethi and K. Dharmarajan, "Diagnosis of chronic disease in a predictive model using machine learning algorithm," in Proceedings of the 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), pp. 191–96, IEEE, Bengaluru, India, October 2020.

[6] S. Y. Yashfi, Md A. Islam, N. Sakib et al., "Risk prediction of chronic kidney disease using machine learning algorithms," in Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–5, IEEE, Kharagpur, India, July 2020.

[7] S. Ganiger and K. M. M. Rajashekharaiah, "Chronic diseases diagnosis using machine learning," in Proceedings of the 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), pp. 1–6, IEEE, Kottayam, India, December 2018.

[8] J. Mishra and S. Tarar, "Chronic disease prediction using deep learning," in Proceedings of the International Conference on Advances in Computing and Data Sciences, pp. 201–11, Springer, Valletta, Malta, April 2020.

[9] B. Gudeti, S. Mishra, S. Malik, T. F Fernandez, A. K. Tyagi, and S. Kumari, "A novel approach to predict chronic kidney disease using machine learning algorithms," in Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1630–1635, IEEE, Coimbatore, India, November 2020.

[10] A. Al-Sammarraee and N. Alshareeda, "The role of artificial intelligence by using automatic accounting information system in supporting the quality of financial statement," Information Sciences Letters, vol. 10, no. 2, pp. 223–254, 2021.

[11]. R. Al-Mashhadani, Gamal Alkawsi, Y. Baashar et al., "Deep learning methods for solar fault detection and classification: a review," Information Sciences Letters, vol. 10, no. 2, pp. 323–331, 2021.

[12] Mirjana Sabljar Matovinovic, "PATHOPHYSIOLOGY AND CLASSIFICATION OF KIDNEY DISEASES", DOI 20/01-2009.

[13] Muhammad SalmanI, Amer Hayat KhanII, Azreen Syazril AdnanIII, Syed Azhar Syed SulaimanIV, Khalid HussainV, Naureen ShehzadiVI, Fauziah Jummaa "Attributable causes of chronic kidney disease in adults", DOI 10.1590/1516-3180-2015.

[14] Robert Thomas MD, Abbas Kanso MD, John R. Sedor MD "Chronic Kidney Disease and Its Complications" DOI 10.1016/j.pop-2008.

[15] M. Chen, H. Yixue, K. Hwang, L Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," Ieee Access, vol. 5, pp. 8869–8879, 2017

- s from eligibility process (real time) based on the account detail.
- To show the forecast result in a desktop or web application to automate this process.
- To maximise the effort put into implementing the task in an AI environment.

VIII. References

[1] Haishuai Wang, Benjamin Bowe*, Zhicheng Cui Hong Yang*, S. Joshua Swamidass, Yan Xie, and Ziyad Al-Aly, "A Deep Learning Approach for the Estimation of Glomerular Filtration Rate", IEEE access,vol-10,pp.1109/TNB.2022.3147957.*

[2] S. N. Induja and C. G. Raji, "Computational methods for predicting chronic disease in healthcare communities," in Proceedings of the 2019 International Conference on Data Science and Communication (IconDSC), pp. 1–6, IEEE, Bangalore, India, March 2019.

[3] B. Manjulatha and P. Suresh, "An ensemble model for predicting chronic diseases using machine learning algorithms," in Smart Computing Techniques and Applications, pp. 337–345, Springer, New York, NY, USA, 2021.

[4] R. Ge, R. Zhang, and P Wang, "Prediction of chronic diseases with multi-label neural network," IEEE Access, vol. 8, pp. 138210–138216, 2020.

[5] I. Preethi and K. Dharmarajan, "Diagnosis of chronic disease in a predictive model using machine learning algorithm," in Proceedings of the 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), pp. 191–96, IEEE, Bengaluru, India, October 2020.

[6] S. Y. Yashfi, Md A. Islam, N. Sakib et al., "Risk prediction of chronic kidney disease using machine learning algorithms," in Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–5, IEEE, Kharagpur, India, July 2020.

[7] S. Ganiger and K. M. M. Rajashekharaiah, "Chronic diseases diagnosis using machine learning," in Proceedings of the 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), pp. 1–6, IEEE, Kottayam, India, December 2018.

[8] J. Mishra and S. Tarar, "Chronic disease prediction using deep learning," in Proceedings of the International Conference on Advances in Computing and Data Sciences, pp. 201–11, Springer, Valletta, Malta, April 2020.

[9] B. Gudeti, S. Mishra, S. Malik, T. F Fernandez, A. K. Tyagi, and S. Kumari, "A novel approach to predict chronic kidney disease using machine learning algorithms," in Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1630–1635, IEEE, Coimbatore, India, November 2020.

[10] A. Al-Sammarraee and N. Alshareeda, "The role of artificial intelligence by using automatic accounting information system in supporting the quality of financial statement," Information Sciences Letters, vol. 10, no. 2, pp. 223–254, 2021.

[11]. R. Al-Mashhadani, Gamal Alkawsi, Y. Baashar et al., "Deep learning methods for solar fault detection and classification: a review," Information Sciences Letters, vol. 10, no. 2, pp. 323–331, 2021.

[12] Mirjana Sabljar Matovinovic, "PATHOPHYSIOLOGY AND CLASSIFICATION OF KIDNEY DISEASES", DOI 20/01-2009.

[13] Muhammad SalmanI, Amer Hayat KhanII, Azreen Syazril AdnanIII, Syed Azhar Syed SulaimanIV, Khalid HussainV, Naureen ShehzadiVI, Fauziah Jummaa "Attributable causes of chronic kidney disease in adults", DOI 10.1590/1516-3180-2015.

[14] Robert Thomas MD, Abbas Kanso MD, John R. Sedor MD "Chronic Kidney Disease and Its Complications" DOI 10.1016/j.pop-2008.

[15] M. Chen, H. Yixue, K. Hwang, L Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," Ieee Access, vol. 5, pp. 8869–8879, 2017