International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Early Detection of Chronic Kidney Disease Using Advanced Machine Learning Models

A.Vaishnovi Anuhya[1], Ayyala Ganesh[2], Nallabathini Poojitha[3], Amandeep Singh [4] , Amitha S K[5], Dr. DHANANJAYA. V [6]

[1,2,3,4]Student, Dept of Computer Science and Engineering, Impact College of Engineering and Applied Science, Affiliated to Visvesvaraya Technological University, Bengaluru, Karnataka, India
[5]Asst. Professor, Dept of Computer Science and Engineering, Impact College of Engineering and Applied Science, Affiliated to Visvesvaraya Technological University, Bengaluru, Karnataka, India
[6]HOD, Dept of Computer Science and Engineering, Impact College of Engineering and Applied Science, Affiliated to Visvesvaraya Technological University, Bengaluru, Karnataka, India

---

Abstract - *Every year, an increasing number of patients are diagnosed with late stages of renal disease. Chronic Kidney Disease, also known as Chronic Renal Disease, is characterized by abnormal kidney function or a breakdown of renal function that progresses over months or years. Chronic kidney disease is often found during screening of persons who are known to be at risk for kidney issues, such as those with high blood pressure or diabetes, and those with a blood family who has chronic kidney disease (CKD). As a result, early prognosis is critical in battling the disease and providing effective therapy. Only early identification and continuous monitoring can avoid serious kidney damage or renal failure. Machine Learning (ML) plays a significant part in the healthcare system, and it may efficiently aid and help with decision support in medical institutions. The primary goals of this research are to design and suggest a machine learning method for predicting CKD. Support Vector Machine (SVR), Random Forest (LR), Artificial Neural Network (ANN), and Decision Tree are four master teaching methodologies investigated (DT). The components are built using chronic kidney disease datasets, and the outcomes of these models are compared to select the optimal model for prediction.*

Keywords: Chronic Kidney Disease (CKD), Machine Learning (ML), Support Vector Machine (SVR), Random Forest (LR), Artificial Neural Network (ANN), Decision Tree (DT).

1. INTRODUCTION

A kind of artificial intelligence is machine learning (ML) (AI). Its heart is algorithmic procedures, which allow the machine to solve issues without the need for specialist computer programming. The widespread use of ML in the medical industry promotes medical innovation, lowers medical expenses, and improves medical quality. However, further research on using ML to solve clinical problems in nephrology is needed. Understanding the aim and technique of ML application, as well as the present state of its use in nephrology, is required to properly address and overcome these issues. Machine learning has previously been used to identify human body state, evaluate disease-related aspects, and diagnose a variety of disorders. The term machine learning (ML) is very popular these days, and a lot of clinical prediction model studies have employed this type of technology. While the capacity to capture vast volumes of information on individual patients is transforming the healthcare business, the enormous volume of data being gathered is impractical for humans to comprehend. Machine learning allows healthcare practitioners to advance toward individualized care, often known as precision medicine, by automatically finding patterns and reasoning about data. The integration of machine learning, health informatics, and predictive analytics provides prospects to alter clinical decision support systems and assist improve patient outcomes.

Chronic Kidney Disease refers to the kidneys' inability to fulfil their normal blood filtration role and other functions (CKD). The term "chronic" refers to the progressive deterioration of kidney cells over time. This is a severe renal failure in which the kidney no longer filters blood and there is a significant fluid accumulation in the body. This causes an abnormally high level of potassium and calcium salts in the body. High quantities of these salts in the body cause a variety of additional problems. The primary function of the kidneys is to filter excess water and wastes from the blood. This mechanism must work properly to balance the salts and minerals in our bodies. The proper salt balance is required to manage blood pressure, activate hormones, and create red blood cells, among other things. A high calcium concentration causes bone problems and cystic ovaries in women. CKD can also cause a sudden sickness or an allergy to specific medications. Acute is the medical term for this condition.

2. LITERATURE SURVEY

The kidneys are positioned in the abdominal cavity, on each side of the spine. They generally weigh around 5 times their body weight yet receives only 20% of the blood flow from the heart. The urine generated by each kidney drains into the urinary bladder, which is positioned in the pelvic area, via a distinct urethra. The kidney is the most essential organ in the human body because it manages fluid levels, electrolyte balance, and other elements that maintain the body's internal environment stable and comfortable. Kidney diseases are conditions that impact the kidney's functioning. Renal disorders can lead to kidney failure in its advanced stages. Kidney diseases are conditions that impact the kidney's functioning. Kidneys can be injured, which means they can't accomplish what they should. This is known as chronic kidney disease (CKD). Anyone can get chronic kidney disease.

In medical research, nephrologists primarily employ two primary tests to identify CKD. A blood test to determine glomerular filtration rate (GFR) and a urine test to determine albumin [1]. Genetics, hypertension, diabetes, obesity, age, and other factors can all have an impact on CKD. Important facts and advancements concerning CKD are described in worldwide kidney disease development guidelines and standard foundations such as the US National Kidney Foundation Kidney Disease Outcomes Quality Initiative (KDOQI) and KDIGO (Kidney Disease Improving Global Outcome). The renal patient is detected by two tests, according to the KDIGO CKD and English National Institute for Health and Care Excellence (NICE) CKD recommendations, these are blood tests to examine how well the kidneys filter the blood to remove creatinine, a normal muscle breakdown byproduct. In comparison, a urine test will show that protein is still present in the urine. Protein (albumin) is a blood component that is normally not transferred into the urine by the kidney filter. When albumin is detected in the urine, it shows that the kidney filters are compromised and may indicate chronic renal disease. Chronic Kidney Disease (CKD) is defined as kidney damage or a glomerular filtration rate (GFR) of less than 60 ml/min/1.73 m2 for more than 3 months with serious health consequences.

Prognosis of CKD by GFR and Albuminuria Categories: KDIGO 2012

GFR categories (ml/min/ 1.73 m²) Description and range			Persistent albuminuria categories Description and range		
			A1	A2	A3
			Normal to mildly increased	Moderately increased	Severely increased
			< 30 mg/g < 3 mg/mmol	30-300 mg/g 3-30 mg/mmol	> 300 mg/g > 30 mg/mmol
G1	Normal or high	≥ 90			
G2	Mildly decreased	60-89			
G3a	Mildly to moderately decreased	45-59			
G3b	Moderately to severely decreased	30-44			
G4	Severely decreased	15-29			
G5	Kidney failure	< 15			

Fig- 1 : GFR and Albuminurea categories KDIGO 2012.

Nephrologists in medical research generally use two major tests to diagnose CKD. GFR is determined by a blood test, whereas albumin is determined by a urine test. CKD can be influenced by genetics, hypertension, diabetes, obesity, age, and other variables. Important CKD facts and developments are given in international kidney disease development recommendations and standard foundations such as the US National Kidney Foundation Kidney Disease Outcomes Quality Initiative (KDOQI) and KDIGO (Kidney Disease Improving Global Outcome). According to the KDIGO CKD and English National Institute for Health and Care Excellence (NICE) CKD guidelines, the renal patient is recognized by two blood tests that measure how well the kidneys filter the blood to eliminate creatinine. a natural consequence of muscle breakdown A urine test, on the other hand, will reveal that protein is still present in the urine. Protein (albumin) is a blood component that is generally not excreted by the kidney filter. When albumin is found in the urine, it indicates that the kidney filters are faulty and may suggest chronic renal illness. Chronic Kidney Disease (CKD) is defined as kidney damage or a glomerular filtration rate (GFR) of less than 60 ml/min/1.73 m2 for more than 3 months, which has major health repercussions.

N.A. Almansour et al. intend to aid in the prevention of CKD by using machine learning techniques to detect CKD at an early stage, focusing on applying multiple machine learning classification algorithms to a dataset of 400 patients and 24 CKD-related variables. As classification algorithms, they employ Artificial Neural Network and support vector machine, and any missing values in the dataset are replaced by the relevant characteristics. The best collected parameters and characteristics were used to construct the final model of the two suggested strategies. Experiment results showed that ANN performed better than SVM, with accuracy of 99.75 percent and 97.75 percent, respectively.

3. Problem Statement

Noncommunicable illnesses are the leading cause of early death, and CKD is the leading noncommunicable disease. Chronic Kidney Disease is a major concern for the global health care system. People with CKD must focus on implementing proven, cost-effective therapies to as many people as possible while taking into consideration restricted needs, human and financial resources. Chronic kidney disease (CKD) is now wreaking havoc on society and is spreading at an alarming rate. Various efforts have been undertaken to advance early therapy to prevent the condition from progressing to chronic disease. Recent research suggests that some of the negative outcomes can be avoided with early identification and treatment.

4. SYSTEM ARCHITECTURE

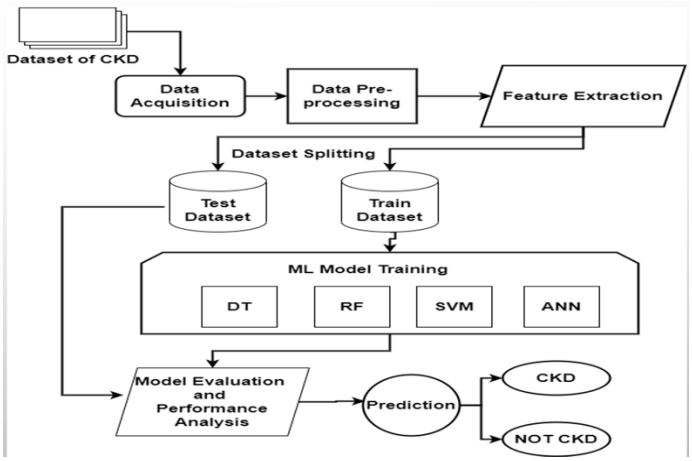


Fig- 2 : Block Diagram of Proposed System

Models of Machine Learning

To predict chronic kidney disease, this study employs Decision Tree Classifiers, Random Forest Classifiers, Support Vector Machines, and Artificial Neural Networks. Among these algorithms, we attempt to construct our prediction model, and we choose the best performance by assessing their accuracy.

Support Vector Machine (SVM)

This is the most well-known and useful supervised machine-learning method, which works on classification and regression issues but is most used for classification. To segregate labelled data, SVM employed a kernel function. One of the benefits of employing kernels in SVM is that SVM applies kernel dentitions to non-vector inputs, and kernels may be built using a variety of data types. The SVM algorithm divides data into two data points when the hyperplane sits between two branches. SVM generates a distinct hyper plane in the training data's signifier space, and compounds are categorized based on which side of the hyper plane they are found on. SVM has been used in a variety of applications, including bioinformatics and handwriting recognition. Other uses for SVM include medical diagnosis, weather prediction, financial market analysis, and image processing. SVM, like all other machine learning approaches, is a computer algorithm that assigns labels to objects based on experience and examples.

The Decision Tree

The decision tree algorithm is a supervised machine learning technique that can handle both classification and regression issues, however it is most employed to tackle classification problems. The Decision Tree approach solves the classification issue by turning the dataset into a tree representation through feature value sorting. Every node in a decision tree represents a feature of an instance to be categorized, and every leaf node represents a class label to which the instance belongs.

Random Forest

The random forest technique is a simple and adaptable supervised machine learning approach that uses diverse collections of decision trees to solve classification and regression problems. Random forest works by constructing the forest, which is an ensemble of decision trees that are typically trained using the bagging approach. Bagging techniques are a method of combining learning models to improve overall outcomes. This model consists of numerous decision trees and outputs the class target that is the target with the highest selection outcomes from each tree.

Neural Network Artificial

Artificial Neural Networks, or ANN, are essentially computational models. These computations Models are composed of a sophisticated network of fundamental components or nodes known as neurons. The nodes are connected to one another. Each node-to-node link has a weight associated with it. A neural network's fundamental structure consists of three layers. The first layer is referred to as the input layer, the intermediate layer as the concealed layer, and the last layer as the output layer. All three layers of a neural network will almost certainly include one or more nodes. A rudimentary neural network will only have one hidden layer. A sophisticated neural network, on the other hand, may include numerous hidden layers. The network's inputs are routed through the basic input layers. The calculation takes place in the hidden layers, and the resulting output is sent through the output layers. There are several types of neural networks accessible. Nowadays, neural networks with several hidden layers are common and serve an important part in a variety of large computing tasks. The Multi-Layer

Preceptor is an excellent example of a neural network with numerous hidden layers (MLP). Even a basic neural network cannot address a complicated issue that an MLP can.

5. METHODOLOGY

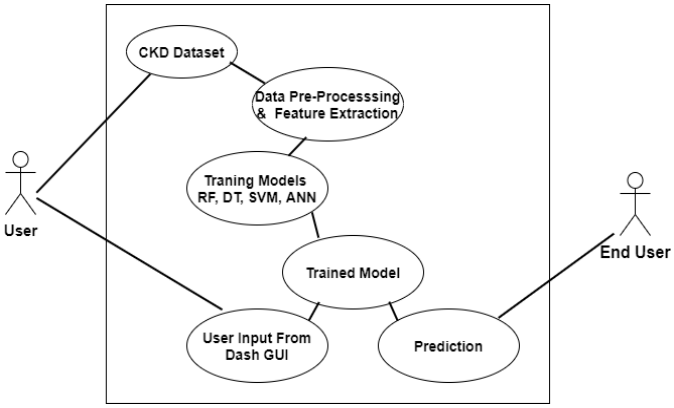


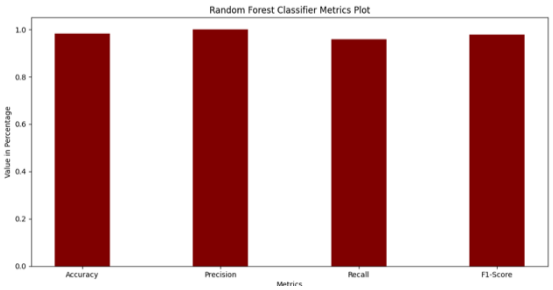
Fig- 3 : Use Case Diagram of Proposed System

This section outlines the system's methodology. The study process begins with gathering publicly available data sets from the internet. The data utilized in the study came from the UCI repository. The data collection contains a total of twenty-five characteristics, including the chronic diagnostic class. Following successful data collection, the data is put into the python script for exploratory data analysis, followed by machine learning for CKD diagnosis. The comma-separated data is read as a data frame by pandas and subjected to pre-processing, pre-processing, and feature selection. During the preparation, data of various data types are aligned as integer type data and verified for null values and duplicate values. The null values in the data are replaced with the attribute's mean.

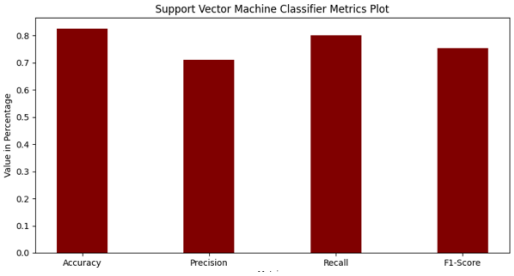
To transform strings to numerical values, the prepared data is subjected to label encoding. The processed data is subjected to feature and label separation. Apart from the class element, which serves as the label, the characteristics are the attributes given in the table. The separated features are induced for optimum feature selection using the chi-square technique, and the top twenty features from the specified characteristics are chosen. The chosen characteristics are utilized to train two different machine learning algorithms, Support vector machine and Decision tree classifier. Metrics such as accuracy, precision, recall, and f1-score are used to evaluate the chosen machine learning models. Aside from individual classifiers, the combination of the SVM and DT is used to test the feasibility of merging two classifiers. Using the pickle module, all models are stored locally.

Knowledge discovery is a significant datamining application that comprises several phases of processing. Preprocessing data acquired from numerous sources makes it easier to use datamining methods. Data preparation, also known as preprocessing, is cleaning, extracting, and converting data into appropriate forms. A bigger feature set is used to identify the essential aspects of knowledge representation. Following that, several categorization or pattern assessment methods are used to aid in knowledge discovery. Bharathi demonstrates a broad illness prediction model based on machine learning. The research attempts to develop a datamining methodology for CKD dataset knowledge discovery. Many CKD datasets are being gathered. The standard methods of data mining process are used for data preparation and preprocessing. To predict the early onset of CKD, three machine learning techniques are used: Decision Tree, Random Forest, and Support Vector machines. Each algorithm's effectiveness is evaluated.

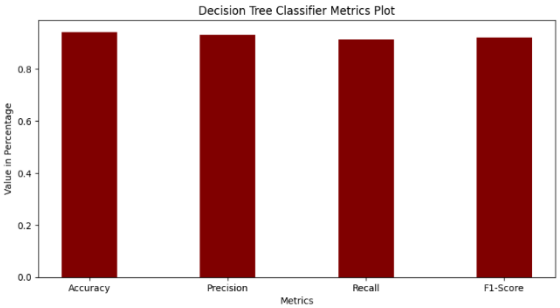
6. RESULTS



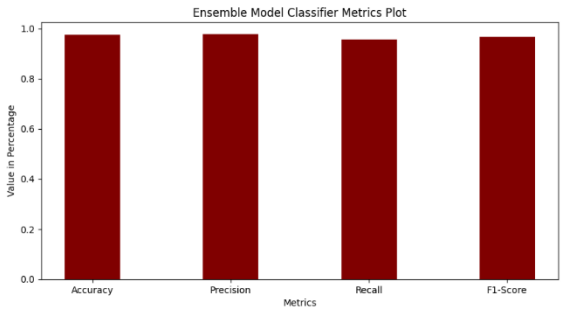
Graph- 1: Plotting of Random Forest Classifier Metrics



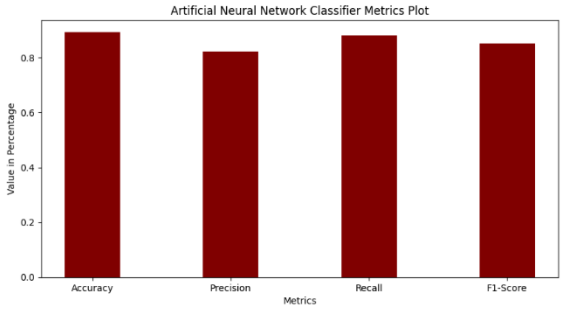
Graph- 2: Plotting of Support Vector Machine Classifier Metrics



Graph- 3: Plotting of Decision Tree Classifier Metrics



Graph- 4: Plotting of Ensemble Model Classifier Metrics



Graph- 5: Plotting of Artificial Neural Network Classifier Metrics

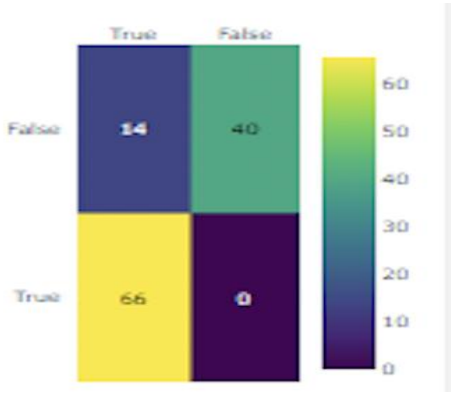


Fig - 4 : Confusion Metrix

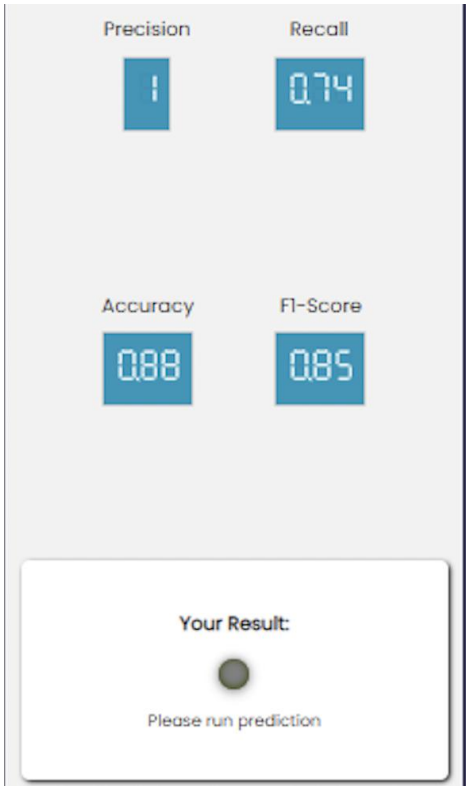


Fig - 5 : Representation of Precision & Accuracy

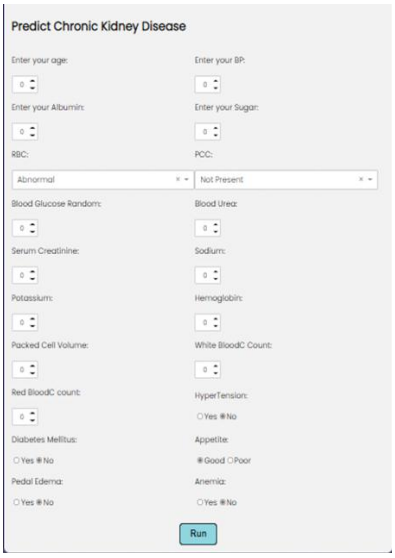


Fig - 6 : Final O/p in Web Application

7. CONCLUSION

This study developed an algorithm for predicting CKD at an early stage. The dataset contains input parameters obtained from CKD patients, and the models are trained and validated using the valid parameters. To diagnose CKD, decision tree, random forest, and support vector machine learning models are built. The accuracy of

prediction is used to assess the performance of the models. The study's findings revealed that the Random Forest Classifier model outperforms Decision Trees and Support Vector Machines in predicting CKD. As an extension of this research, the comparison may also be done depending on the duration of execution and feature set selection.

REFERENCES

1) G. Chen et al., "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," IEEE Access, vol. 8, pp. 100497–100508, 2020, doi: 10.1109/ACCESS.2020.2995310.

2) P. T. Coates et al., "KDIGO 2020 Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease," Kidney Int., vol. 98, no. 4, pp. S1–S115, 2020, doi: 10.1016/j.kint.2020.06.019.

3) L. Chen, "Overview of clinical prediction models," Ann. Transl. Med., vol. 8, no. 4, pp. 71–71, 2020, doi: 10.21037/atm.2019.11.121.

4) H. Kriplani, B. Patel, and S. Roy, Prediction of chronic kidney diseases using deep artificial neural network technique, vol. 31. Springer International Publishing, 2019.

5) T. O. Ayodele, "Atherosclerotic Cardiovascular Disease," Atheroscler. Cardiovasc. Dis., 2012, doi: 10.5772/711.

6) Abdi, "Three types of Machine Learning Algorithms List of Common Machine Learning Algorithms," no. November 2016, doi: 10.13140/RG.2.2.26209.10088.

7) S. Y. Yashfi et al., "Risk Prediction of Chronic Kidney Disease Using Machine Learning Algorithms," 2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020, 2020, doi: 10.1109/ICCCNT49239.2020.9225548.

8) N. A. Almansour et al., "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," Comput. Biol. Med., vol. 109, no. October 2018, pp. 101–111, 2019, doi:10.1016/j.compbiomed.2019.04.017.

9) Sobrinho, A. C. M. D. S. Queiroz, L. Dias Da Silva, E. De Barros Costa, M. Eliete Pinheiro, and A. Perkusich, "Computer-Aided Diagnosis of Chronic Kidney Disease in Developing Countries: A Comparative Analysis of Machine Learning Techniques," IEEE Access, vol. 8, pp. 25407–25419, 2020, doi: 10.1109/ACCESS.2020.2971208.

10) N. V. Ganapathi Raju, K. Prasanna Lakshmi, K. G. Praharshitha, and C. Likhitha, "Prediction of chronic kidney disease (CKD) using Data Science," 2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019, no. Iciccs, pp. 642–647, 2019, doi: 10.1109/ICCS45141.2019.9065309.

11) P. Kotturu, V. V. S. Sasank, G. Supriya, C. S. Manoj, and M. V. Maheshwarredy, "Prediction of chronic kidney disease using machine learning techniques," Int. J. Adv. Sci. Technol., vol. 28, no. 16, pp. 1436–1443, 2019, doi: 10.17148/IJARCCE.2018.71021.

12) DeepLearning," 2018 Int. Conf. Innov. Eng. Technol. ICIET 2018, pp. 1–6, 2019, doi: 10.1109/CIET.2018.8660844.