Grenze International Journal of Engineering and Technology, Jan Issue

A Efficient Method to Predicting Chronic Kidney Disease using Machine Learning

Nitin L. Shelake[1], Kene Swati[2], Yogesh S. Deshmukh[3], Maher Priyanka[4] and Shingade Sakshi[5]
[1-5]Department of Information Technology, Sanjivani College of Enginerring, Kopargoan, India
Email: nitinshelake2011@gmail.com, keneswati05@gmail.com, yogesh.d151@gmail.com, maherpriyanka2000@gmail.com, sshingade138@gmail.com

***Abstract*— chronic kidney disease (CKD), commonly referred to as chronic renal disease, is a group of disorders that affect your kidneys and reduce their ability to keep you healthy. Nerve harm, high blood pressure, anemic (a low blood count), weak bones, and inadequate nutrition could be symptoms. Early detection and treatment often prevents chronic renal disease from getting worse. Data mining is the technique used to gather knowledge from huge datasets. Data mining, which aims to use previous data to find recurring patterns and enhance decision-making moving forward, is a result of the convergence of several current trends: the decreasing price of massive data storage devices and the growing simplicity of data collection across networks; the development of reliable and effective machine learning algorithms to handle this data, as well as the declining cost of computing power, which makes it possible to apply computationally expensive techniques for data analysis. Machine learning has already produced useful applications in fields like analyzing results from medical research, spotting fraud, spotting bogus users, etc. For the purpose of predicting chronic diseases, various data mining and categorization methodologies and machine learning algorithms are used. The goal of this study is to develop a new decision-support system for forecasting chronic renal disease. This study compares support vector machine (SVM) classifier performance for CKD prediction based on its accuracy, precision, and execution time**

***Index Terms*— Chronic kidney disease, Prediction, Machine Learning, Decision Tree, Random Forest and Logistic Regression.**

I. INTRODUCTION

Chronic kidney disease (CKD) is a global public health concern, affecting a substantial portion of the population and imposing a significant burden on healthcare systems. CKD is characterized by the gradual loss of kidney function over time, leading to various complications and an increased risk of end-stage renal disease (ESRD) and cardiovascular events. Early identification of CKD is crucial for implementing timely interventions, including lifestyle modifications, medication management, and potential renal replacement therapy.

Traditional methods for CKD diagnosis and prediction rely on laboratory tests, medical history, and clinical assessments. However, these approaches often have limitations in terms of accuracy, efficiency, and cost-effectiveness. With the increasing availability of electronic health records (EHR) and the accumulation of large-scale datasets, machine learning techniques have gained prominence in the field of healthcare for their potential to leverage the power of data-driven approaches in predicting and identifying complex diseases. Machine learning algorithms offer the ability to automatically discover intricate patterns and relationships within vast amounts of

Grenze ID: 01.GIJET.10.1.807

data, providing insights and predictive capabilities that can enhance CKD management. By training on historical patient data, these algorithms can learn from the features associated with CKD, such as demographic information, laboratory test results, medical history, and comorbidities, to generate predictive models that estimate an individual's risk of developing CKD.

Kidney disease, also known as CKD, prevents the kidneys from filtering blood properly. Because of this, extra fluids and blood waste build up in the body and can lead to various health problems like heart disease and stroke. In the US, kidney illnesses are one of the main causes of death. Adults in the US are thought to have CKD in about 37 million cases, most of which go untreated. 40% of individuals with significant renal impairment (not receiving dialysis) are unaware that they have CKD. 360 people start receiving dialysis for renal failure every 24 hours. 3 out of every 4 new occurrences of renal failure in the US are caused by diabetes and excessive blood pressure. Medicare enrollees with CKD required $87.2 billion in treatment in 2019, while those with ESRD required an extra $37.3 billion. Depending on a patient's chronic condition, from Stage 1 to Stage 5, kidney disease requires appropriate medical treatment. The steps will change depending on the cause. Treatment often consists of steps to lessen consequences, control symptoms, and delay the disease's course. Doctors do many laboratory tests on a patient, such as blood, urine, etc., to diagnose them. Doctors will use those tests to assess the patients' conditions and plan their care. We suggest employing support vector machines to diagnose a decision support system for chronic renal disease. The classification and clustering algorithm used is called Support Vector Machine (SVM). The Support Vector Machine algorithm's fundamental premise is to find the best restriction for dividing the two components of a class known as the hyper plane. SVM has been used by researchers to do classification tasks in a variety of fields, including affective computing. The creation of this system is intended to serve as a decision-support tool for diagnosing kidney disease and determining whether or not patients with renal disease have progressed to a stage of chronic kidney disease.

A. Contribution

1. To develop a system which can able to identify whether the person is infected with the disease or not.
2. To present the detail analysis of the kidney disease
3. To implement the chronic kidney disease application with highest possible accuracy.
4. To develop a user friendly interface so that user can easily navigate through it.
5. To develop the system based on real data

II. LITERATURE REVIEW

[1] Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Disease Shamima Akter, Ahsan Habid, Md. Ashiqul Islam, Md. Sagar Hossen, Wasik Ahmmed Fahim, Puza Rani Sarkar, Manik Ahmed 19 November 2021. In this paper the incidence of chronic kidney disease (CKD) is rising rapidly around the globe. Asymptomatic CKD is common and guideline directed monitoring to predict CKD by various factors is underutilized. Computer-aided automated diagnostic (CAD) can play a major role to predict CKD. This will assist predictive analytics to advance CKD prediction by using deep learning more efficiently and effectively. With an average AUC and a precision above 0.87 and 0.8, respectively, the linear models including Elastic Net, lasso regression, ridge regression, and logistic regression had the strongest overall predictive power. First place went to logistic regression, which had an AUC of 0.873 and AUCs of 0.83 and 0.82, respectively. Elastic Net (0.85) had the highest sensitivity, while XGBoost had the highest specificity (0.83). We found that ALB, Scr, TG, LDL, and EGFR had significant effects on the predictability of the models using effect size analyses, whereas other predictors including CRP, HDL, and SNA had less of an impact.

[2] Prediction of Chronic Kidney Disease - A Machine Learning Perspective Pankaj Chittora, Sandeep Chaurasia, (senior member, IEEE), Prasun Chakrabarti (senior member, IEEE), Gaurav Kumawat, Tulika Chakrabarti, Zbigniew Leonowicz (senior member, IEEE) 22 January 2021 In this paper in chronic kidney disease is one of the most critical illness nowadays and proper diagnosis is required as soon as possible. Machine learning technique has become reliable for medical treatment. With the help of a machine learning classifier algorithms, the doctor can detect the disease on time. For this perspective, Chronic Kidney Disease prediction has been discussed in this article. It was discovered that particular parameters with a higher information gain, such as specific gravity, albumin, serum creatinine, hemoglobin, packed cell volume, red blood cell count, and hypertension, were more significant than the other features. LOG, NB, ANN, and RF were able to perform better with KNN Imputation than they did with random imputation, removing rows with missing values, or mean and mode imputation.

[3] Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform Guozhen Chen , Chenguang Ding , Yang Li , Xiaojun Hu , Xiao Li , Li Ren

, Xiaoming Ding , Puxun Tian , And Wujun Xue 18 May 2020 In this research the Chronic Kidney disease is a severe lifelong condition caused either by renal disease or by impaired functions of the kidneys. In the present area of research, Kidney cancer is one of the deadliest and crucial importance for the survival of the patients' diagnosis and classification. Early diagnosis and proper therapy can stop or delay the development of this chronic disease into the final stage where dialysis or renal transplantation is the only way of saving the life of the patient.

[4] A Neural Network based Model for Predicting Chronic Kidney Diseases Chalumuru Suresh, Assistant Professor, B. Chakra Pani, C. Swatisri, R. Priya, R. Rohith 01 September 2020 the research paper of Chronic kidney infection is a huge and rising problem among developing populations. It is one of the critical known health concerns with rising demand for early detection for prosperous and sustainable care. Many factors (such as Haemoglobin, Blood Pressure, diabetes, obesity) contribute to the continuous decline of kidney function over time. Chronic kidney disease has many stages for the criticality of the disease. Hence, a better diagnosis of chronic kidney disease is required to prevent continuous progression.

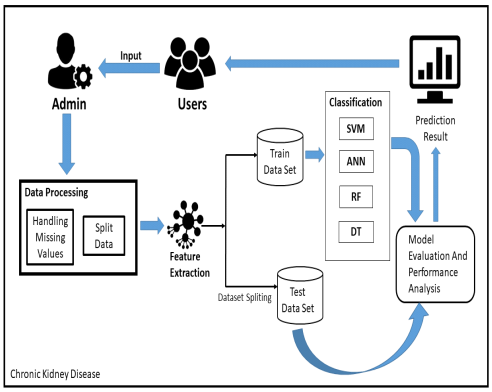
[5] Diagnostic Decision Support System of Chronic Kidney Disease Using Support Vector Machine Mubarik Ahmad, Vitri Tundjungsari, Dini Widianti, Peny Amalia, Ummi Azizah Rachmawati 05 February 2018. This paper of Kidney disease or commonly known as kidney failure is a condition when the renal function is declining that could result in the inability of the kidneys to perform their duties. Kidney disease patients have the potential to get into the chronic phase. The renal function is declining that could result in the inability of the kidneys to perform their duties. Kidney disease patients have the potential to get into the chronic phase. Chronic kidney disease is a decrease in kidney function gradually during the three months which resulted in the cessation of kidney function in total.

The findings of this study have the potential to contribute significantly to the early detection and management of CKD. By leveraging machine learning techniques, healthcare providers can enhance risk assessment, optimize resource allocation, and facilitate personalized interventions, ultimately improving patient outcomes and reducing the burden of CKD on individuals and healthcare systems

III. Proposed System

Chronic Kidney Infection is proved to be fatal when it is not diagnosed or treated properly. Diagnosing a chronic Kidney disease at the correct time is more important in order to stop its progression. The proposed system will predict the chronic kidney disease with better classification of the CKD stage, accuracy, recommendations and with less over fitting problems. For that, the proposed research work has developed a machine learning model with support vector machine. There were several steps involved in creating the desired machine learning model which include:

1. Data Pre-processing,
2. Feature Extraction,
3. Splitting the data,
4. Applying the support vector machine algorithm



The proposed system's architecture is made up of various blocks, as shown in fig (1) as follows.

A. Data Processing

Preparing raw data to be acceptable for a machine learning model is known as data preparation. In order to build a machine learning model, this is the first and most important stage. As the available data is real-world data, it

contains inconsistent data, missing values, and other noisy data. The proposed system must clean the raw data in order to remove this inconsistent data from the dataset. Data processing is a way to transform noisy and huge data into relevant and clean data. This is necessary to finish the prediction model. It decreases dimensionality and aids the machine in producing better outcomes.

1) Handling Missing Values

It was discovered that the CKD dataset had a significant amount of missing values for all of the different variables. It is simply not possible to eliminate samples or features that have missing values because doing so could result in the loss of important data. Backfilling is the process of filling in missing data by using the value that is next to a missing or unavailable number

2)Data Spliting

The CKD dataset was divided into two halves, with 70% of the dataset (280 instances) used to train the model and the remaining 30% (120 instances) used to assess how well data mining algorithms performed on the CKD dataset.

3) Data Scaling

Since not all of the characteristics in the dataset correspond to entities that were photographed and measured on the same scale, a data scaling methodology was employed to bring the features onto a scale that matched their unique intended values

B. Feature Extraction

The process of turning raw data into process able numerical features while keeping the original data set's content intact Feature extraction in chronic kidney disease (CKD) is a process used to identify and extract relevant characteristics or features from patient data, such as clinical, laboratory, and imaging data, to aid in the diagnosis, prognosis, and management of CKD. CKD is a long-term condition where the kidneys do not function properly; leading to a gradual loss of kidney function over time. The employment of feature selection models has a significant negative impact on each model's performance. As they eliminate qualities that have a low or negative impact on the data, they significantly alter the attributes that will be used to train the model. This fundamental principle is the foundation of machine learning. The initial step in model design is expected to be feature selection and data cleaning

C. Spliting the data

In machine learning, splitting is frequently used to prevent overfitting. In this case, the model fits the training data too well and is unable to accurately fit new data. We divide the data in our system into two categories: training data and testing data. We categorized our data using various algorithms in the training data set.

D. Applying machine learning algorithm

In our model we add some following algorithm:
a)SVM: A deep learning technique known as a support vector machine performs supervised learning for the categorization or regression of data groupings. SVM is used in our model to classify trained data. SVMs in CKD applications depends on the quality and relevance of the input features, the availability of labeled data for training, and the choice of the appropriate SVM variant (linear, non-linear, etc.). Collaboration between medical experts, data scientists, and domain specialists is crucial to develop effective SVM-based models for CKD-related tasks. The SVM decision function is used to classify new instances into different classes based on a decision boundary. For a binary classification problem, the decision function is:

$$f(x)=sign(w^T*x+b)$$

'f(x)' is the predicted class label for input instance 'x'
b)ANN: Artificial Neural Networks (ANNs) can be utilized for chronic kidney disease (CKD) identification by training a neural network model on relevant data. Applications for artificial neural networks include speech recognition, machine translation, picture recognition, and medical diagnosis, correleation analysis, feature importance to identify the most informative features for CKD identification. An important advantage of ANN is that it learns from the sample datasets.
c) RF : Leo Breiman and Adele Cutler are the creators of the widely used machine learning technique known as random forest, which mixes the output of various decision trees to produce a single outcome. Its widespread use is motivated by its adaptability and usability because it can solve classification and regression issues. Also here the use of random forest to measure the number of trees and depth in RF.

d) DT- Decision trees are a supervised learning method that is mostly used to solve classification algorithms but may also be applied to regression problems. In a decision tree, this approach compares the values of the root property with the record attribute and then jumps to the branch that is based on the comparison. to the next node to forecast the class of the provided dataset.. The algorithm checks the attribute value with the other sub-nodes once more for the next node before proceeding. Continues doing this until it reaches the tree's leaf node. Here decision treee splitting criterion used to determine the best attribute and threshold for identification of the chronic symptoms.

IV. METHODOLOGY

1. In this system, we created a model to anticipate patients' CKD disease. The model's performance with regard to all qualities and particular aspects was examined.
2. On the chosen features, the performance of classifier methods was evaluated. The model is created with the IBM SPSS tool.
3. The model was trained using a variety of machine learning classifiers, including the artificial neural network (ANN), C5.0, logistic regression, linear support vector machine (LSVM), K-closest neighbours (KNN), and random forest.

Here are some common symptoms associated with chronic kidney disease like fatigue, Swelling (edema) in the-legs, ankles, feet or hands, Shortness of breath, frequent urination, blood in urine, foamy, difficulty concentrating, decreased appetite, nausea, vomiting, itching, muscles cramps, persistent itching, unexplained weight loss, sleep problems, High blood pressure, bone pain, hypertension and so on, experiencing these symptoms does not necessarily mean someone has CKD, as they can also be caused by other conditions so this system will diagnosis properly like health care professional by using certain attribute along with above symptoms. Chronic kidney disease (CKD) is a prevalent and progressive health condition that affects millions of individuals worldwide. Early detection and accurate prediction of CKD can significantly improve patient outcomes and enable timely interventions. Machine learning techniques have emerged as promising tools for predicting and identifying CKD by leveraging the patterns and relationships within large and complex datasets

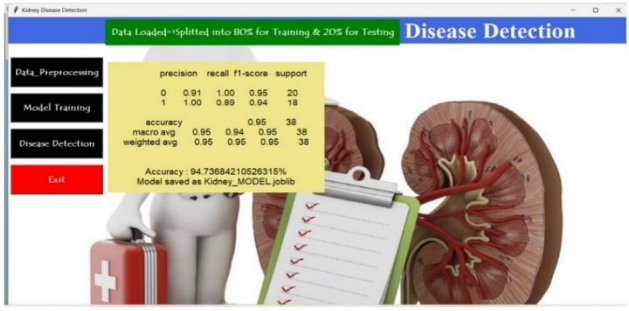
V. DESCRIPTION OF DATASET

The dataset used in this study is a chronic is Kidney disease dataset Collected from Kaggle. Table 1 shows that the dataset contains 400 instances of CKD patients with different symptoms and 25 attributes, with 11 numeric and 14 nominal values. The dataset is made of two classes' non CKD and CKD relating to the patients without Chronic Kidney disease and with Chronic Kidney Disease.

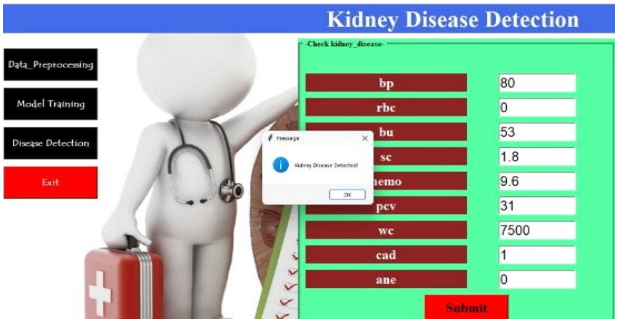
Sr. No	Attribute Name	Description
1	Age	Patient age (It is in years)
2	Bp	Patient blood pressure (It is in mm/HG)
3	Sg	Patient urine specific gravity
4	Al	Patient albumin ranges from 0-5
5	Su	Patient sugar ranges from 0-5
6	Rbc	Patient red blood cells two value normal and abnormal
7	Pc	Patient pus cell two value normal and abnormal
8	Pcc	Patient pus cell clumps two values present and not present
9	Ba	Patient bacteria two values present and not present
10	Bgr	Patient blood glucose random in mg/dl
11	Bu	Patient blood urea in mg/dl
12	Sc	Patient serum creatinine
13	Sod	Patient sodium
14	Pot	Patient potassium
15	Hemo	Patient hemoglobin (protein molecule in red blood cells)
16	Pcv	Patient packed cell volume % of red blood cells in circulating blood
17	Wc	Patient white blood cell counts in per microliter
18	Rc	Patient red blood cell count in million cells per microliter
19	Htn	Patient hypertension two value Yes and No
20	Dm	Patient diabetes mellitus two value Yes and No
21	Cad	Patient coronary artery disease two value Yes and No
22	Appet	Patient appetite two value good and poor
23	Pe	Patient pedal edema two value Yes and No
24	Ane	Patient anemia two value Yes and No
25	Class	Target Variable (CKD or Not)

VI. Project Experimental outcomes

In the interface designed to facilitate the training of machine learning models specifically tailored for CKD prediction, risk assessment, or treatment optimization.



This page plays a crucial role in developing accurate and reliable predictive models by leveraging CKD-specific data and advanced machine learning techniques
In this experiment, anticipating the CKD count will help determine whether or not a person has the disease.



With a generally user-friendly interface, the data is provided via a flask web page and is powered by efficient machine learning technologies and algorithms. To compare new data with trained data, one might use training datasets that have been gathered. As a result, after the inputs, the application displays the outcome. The project's suggested model will employ the supervised machine learning technique known as decision tree regression. To improve accuracy, it has been trained on a variety of criteria, including hypertension, high blood pressure, obesity, and diabetes.

VII. Prediction And Recommendation Outcomes

Machine learning predictive algorithms have highly optimized estimation that is likely to be a result based on trained data. Predictive analysis addresses the use of data, statistical algorithms and machine learning techniques to predict the probability of outcomes by using historical data. The objective is to go beyond the knowledge of what has happened and to provide a better evaluation of what will happen in the future. Our system utilizes a supervised machine learning algorithm that has subcategories for classification and regression. Classification algorithm will be the most suitable for our system. We used a Support vector machine algorithm for prediction of this system to contain some other features such as displaying the count if the person is infected with CKD or not. Those details will help users or doctors to handle huge data in a smart way.

VIII. Acknowledgements

Author would like to thank management and staff of Sanjivani College of engineering, kopargaon, India for their kind support and guiding us time to time.

IX. Conclusion

In this system, an algorithm for early CKD prediction was created. The models are trained and verified using valid input parameters, which were collected from patients with chronic kidney disease. Decision tree, random forest, and support vector machine learning models are constructed to diagnose CKD. The performance of the models is

evaluated using prediction accuracy. According to the results, the random forest classifier model performs better in predicting chronic kidney disease than decision trees and support vector machines. Depending on the execution time and feature set choice, the comparison may alternatively be conducted as a continuation of this project. A diagnostic decision support system is required to enhance doctors' clinical judgment when determining a patient's chronic renal disease state. We put forth a system that will assist physicians in caring for individuals with kidney illnesses. The support vector machine (SVM) system was created using machine learning methods.

The two primary stages of this study's technique are system creation and categorization modelling. We discovered guidelines and a model for categorizing kidney disease using classification modelling. This technique is anticipated to assist the physician in accurately determining the chronic state of kidney disease patients. The future scope are highly experienced in the management of older multi-morbidity subjects at risk for renal impairment and related clinical complication, including in functional status.

REFERENCES

[1] J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun, S. Zhu, and Z. Ye, ‘‘Comparison and development of machine learning tools in the prediction of chronic kidney disease progression,’’ J. Transl. Med., vol. 17, p. 119, Dec. 2019.

[2] Janani J, Sathyaraj R, "Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques", Turkish Journal of Computer and Mathematics Education, vol. 12, pp. 6383 – 6390, 2021.

[3] F. Ma, T. Sun, L. Liu and H. Jing, "Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network", Future Gener. Comput. Syst., vol. 111, pp. 17-26, Oct. 2020.

[4] M. A. Islam, S. Akter, M. S. Hossen, S. A. Keya, S. A. Tisha and S. Hossain, "Risk factor prediction of chronic kidney disease based on machine learning algorithms", Proc. 3rd Int. Conf. Intell. Sustain. Syst. (ICISS), pp. 952-957, Dec. 2020.

[5] I. U. Ekanayake and D. Herath, "Chronic kidney disease prediction using machine learning methods", Proc. Moratuwa Eng. Res. Conf. (MERCon), pp. 260-265, Jul. 2020.

[6] J. Coresh, "Update on the burden of CKD", J. Amer. Soc. Nephrol., vol. 28, no. 4, pp. 1020-1022, Apr. 2017

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Scien B. Khan, R. Naseem, F. Muhammad, G. Abbas and S. Kim, "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy", IEEE Access, vol. 8, pp. 55012-55022, 2020.

[8] S. Aljahdali, S.N. Hussain "Comparative prediction performance with support vector machine and random forest classification techniques", Int J Comput Appl, vol. 69, pp. 11, 2013.

[9] M. Almasoud and T. E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors", Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 8, pp. 1-9, 2019

[10] El-Houssainy A. Radya , Ayman S. Anwar, "Prediction of kidney disease stages using data mining algorithms" ScienceDirect, vol. 15, pp. 100178, 2019.

[11] A.Vaishnovi Anuhya, Ayyala Ganesh, Nallabathini Poojitha, Amandeep Singh, Amitha S K, Dr. DHANANJAYA. V "Early Detection of Chronic Kidney Disease Using Advanced Machine Learning Models ",International Research Journal of Engineering and Technology (IRJET), vol 09, pp. 3484 – 3489

[12] C. Ho, T. Pai, Y. Peng, C. Lee, Y. Chen, Y. Chen, “Ultrasonography mage Analysis for Detection and Classification of Chronic Kidney Disease,” IEEE Complex, Intelligent and Software Intensive Systems, pp. 624 – 629, July 2012

[13] Miguel A. Estudillo-Valderrama; Alejandro Talaminos-Barroso; Laura M. Roa;David Naranjo-Hernández; Javier Reina-Tosina; Nuria Aresté Fosalba; José A. Milán-Martín, “A Distributed Approach to Alarm Management in Chronic Kidney Disease ,” IEEE Transl. Biomedical and Health Informatics , vol. 18, pp. 1796 – 1803, November 2014.

[14] R. G. Brereton, and G. R. Lloyd, “Support Vector Machines for classification and regression,” Analyst, vol. 135, no. 2, pp. 230-267, 2010