IJARASEM

ISSN: 2395-7852

International Journal of Advanced Research in Arts, Science, Engineering & Management

Volume 10, Issue 3, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 6.551

+91 9940572462 +91 9940572462 ijarasem@gmail.com www.ijarasem.com

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 6.551 |Bimonthly, Peer Reviewed & Referred Journal|

| Volume 10, Issue 3, May 2023 |

A Web Application Chronic Kidney Disease Prediction Using Machine Learning

Dhanya Shenoy[1], Shreya S[2] ,Swasthik M Acharya[3], Prathik V Shenoy[4], Preran S Jathan[5]

Asst. Professor, Department of Computer Science & Engineering, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Karnataka, India[1]

Student, Department of Computer Science & Engineering, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Karnataka, India[2,3,4,5]

ABSTRACT: This paper presents a web application for predicting chronic kidney disease (CKD) using machine learning algorithms. The proposed system utilizes three popular classification algorithms, namely logistic regression,Support vector machine and k-nearest neighbors classifier(KNN) for prediction. The system is implemented using the Python programming language and deployed as a web application. The dataset used for training and testing the models is obtained from the UCI Machine Learning Repository. The performance of the models is evaluated using various evaluation metrics such as accuracy, precision, recall, and F1-score. The experimental results show that the Logistic Regression algorithm achieved the highest accuracy of 99% in CKD prediction, followed by KNN with an accuracy of 97.28%. The proposed web application provides a user-friendly interface for the prediction of CKD, which can be useful for doctors and patients to make informed decisions.

Keywords: CKD, GFR, KNN.

I.INTRODUCTION

A steady loss of kidney function over months or years is known as chronic kidney disease. Our nephrons, which are our kidneys' millions of microscopic filters, operate to remove wastes from our blood in order to keep us healthy. These nephrons start to shut down if they sustain injury. We start to experience the signs of CKD when there are eventually not enough remaining to filter our blood effectively enough to keep us healthy. However, CKD is typically far along when we first start to experience symptoms. In actuality, a person may lose up to 90% of their kidney's capacity before showing any signs or symptoms at all. Because of this, one in ten persons have CKD, yet the majority are completely unaware of it.
Without treatment, CKD stages 1 through 5 advance. End-Stage Renal Disease (ESRD), a stage 5 illness, requires continuing dialysis or a kidney transplant to survive.

- High blood pressure is one of the potential CKD symptoms and consequences.
- Low blood counts (anemia), weak bones, inadequate nourishment, nerve damage, swollen ankles, and exhaustion

II.LITERATURE REVIEW

[1] They have used a variety of data mining approaches to diagnose kidney-related ailments in this process of diagnosing chronic kidney disease, and their main goal is to make a reliable diagnosis rather than to discover the perfect cure. In this proposal, they employed two data mining techniques, Random Forest algorithm and Back Propagation Neural Network, to identify the chronic kidney diseases and analyze them to provide the best algorithm for predicting the chronic kidney diseases.

[2]In this study, feature optimization was done to discover the most advantageous feature extraction algorithm for the prediction of chronic kidney disease. Three distinct feature selection algorithms were used. In order to improve the performance of the classifier model, class balancing is required because many datasets have unbalanced classes. SMOTE was employed in this study as a class balancer. The highest degree of accuracy, 99.6%, was attained.

[3] In this paper, a data mining methodology for knowledge discovery using the CKD datasets is proposed. Datasets related to CKD are amassed in large numbers. The classic methods of data mining are used for data preparation and

preprocessing. To predict the early onset of CKD, three machine learning algorithms—Decision Tree, Random Forest, and Support Vector Machines—are employed. Each algorithm's merit is evaluated. The approach described below produces a model with a high degree of accuracy.

[4] In the current work, a supervised learning methodology is given that focuses primarily on probabilistic, tree-based, and ensemble learning-based models to build effective models for predicting the risk of developing CKD. They also assessed SVM, LR, SGD, ANN, and k-NN. The Rotation Forest model, which outperformed the other models with accuracy equal to 99.2%, was emphasized in the obtained findings

[5]They conducted a retrospective analysis during the selected observation period that looked at individuals with and without CKD diagnosis. The moment the patient received their initial CKD diagnosis is known as the index date. The index date for the non-CKD group is chosen at random. They want to anticipate the onset of CKD six and a year in advance of the index data (referred to as the lead time). Two years before the lead time, they process data to make the prediction (referred to as the observation time). Taiwan's National Health Insurance Research Database was used to conduct the study (NHIRD)

[6] The suggested method uses the Chronic Kidney Disease dataset from the UCI Machine Learning Repository, which consists of 25 attributes with 11 numerical and 14 nominal values. The dataset contains 400 events in total, of which 150 are classified as non-chronic kidney disease and 250 as chronic kidney disease (CKD) (NOTCKD). The attributes in the dataset include age, bp, sg, al, su, bc, pc, pec, ba, bgr, bu, se, sod, pot, hemo, per, we, re, htn, dm, cad, appet, pe, ane, and classification. A training group and a testing group were created from the dataset. Data are divided 70/30 for testing and training purposes.

[7] Using all available data, statistical techniques, and geographic proximity to countries with data, estimates can be calculated for countries and years with few or no primary data sources. Data collected using different case-definitions or study procedures are adjusted for each disease or risk factor to the level they would have been at if data were obtained using a reference method or case-definition set. Similar adjustments are made by assigning deaths to International Classification of Diseases and Injuries (ICD) categories that are less descriptive to cause-of-death data from vital records or verbal autopsy reports.

[8]In this work, we created and assessed several artificial intelligence-based models that took into account the bare minimum of factors like sex, age, comorbidities, and medication use. After six or twelve months, these models forecast a patient's likelihood of acquiring chronic kidney disease. Convolutional neural networks (CNN) outperformed all other models examined, with AUROC metrics of 0.957 and 0.954 for 6 and 12 months, respectively. We examined the tree-based LightGBM model to determine which properties are most important for prediction. Age, gout, diabetes mellitus, usage of sulfonamides and angiotensins, all of which are appropriate in light of CKD, were the most notable characteristics.

[9]For humans, a variety of formulae have been established to estimate GFR or CrCl. Age, sex, ethnicity, and body size are typically used as stand-ins for the creation of creatinine by muscle in GFR estimation algorithms that use creatinine. We only took into account equations that were developed using assays that could be traced to reference techniques and study populations in which SCr concentration was measured using traceable assays for our review of GFR predicting equations.

[10]In our study, data mining was used since it is a method for finding new, potentially beneficial, reliable, and ultimately understandable patterns in data. Techniques for both supervised and unsupervised learning are employed in data mining classification. Classification, statistical regression, and association rules are studied in both medical and clinical research using a "supervised" learning technique, which necessitates the creation of a model based on prior performance analysis. On the other hand, the "unsupervised" learning method does not develop a pre-analysis hypothesis and is not influenced by previous analysis. Based on the findings, a model can be created that is helpful for clustering.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 6.551 |Bimonthly, Peer Reviewed & Referred Journal|

| Volume 10, Issue 3, May 2023 |

III.METHODOLOGY OF PROPOSED SURVEY

1. **Patient data sets and parameters:**
Datasets were used for processing in the first step of the prediction process, which is where we collect medical data. The features and their descriptions utilized in the CKD prediction data sets are described in the training data-sets, which include patient details and the parameters needed for prediction. The characteristics and their measures utilised in the CKD prediction data sets..

2. **Extract and Segment Data (Data Pre-processing):**Only pertinent datasets are extracted after an analysis of the medical data. The necessary data is extracted and segmented in accordance with the need. This is done because the complete training data set is not needed for processing, and if we input all the data, processing would take too long.Clean and preprocess the patient data to remove missing values, outliers, and irrelevant features.Segment the data into training, validation, and testing sets using a suitable method (e.g., random sampling, stratified sampling).

3. **Train Data:** After the necessary data has been retrieved and segmented, the model must be trained, which entails translating the data into the necessary format, such as numerical values, binary, strings, etc. Depending on the algorithm, conversion may occur. Apply logistic regression with standard scalar to train the data.Use suitable performance metrics such as accuracy, precision, recall, and F1-score to evaluate the trained model.Tune hyperparameters (e.g., regularization parameter) to improve the model's performance.

4. **Classification Techniques:** Use the trained model to predict chronic kidney disease for new patients.Compare the performance of the logistic regression model with other classification techniques such as support vector machine,Logistic Regression and KNN.Develop a web application that integrates the trained model to provide chronic kidney disease risk assessment for patients and clinicians.

5. Utilizing accuracy, precision, recall, F-measure, and stage prediction utilizing GFR, the suggested system's output is evaluated. The "logistic regression" algorithm or "standard scalar" is used by the proposed system to forecast diseases. The confusion matrix technique is used to accurately verify the algorithm. Here, we verify the outcomes produced .

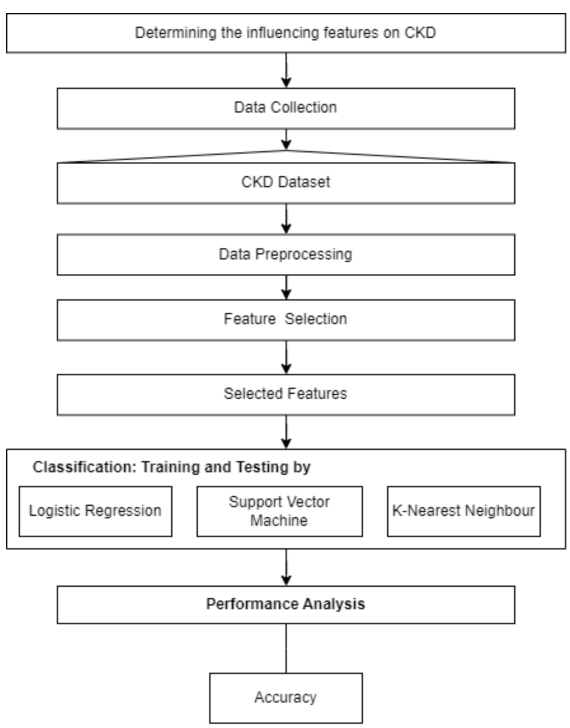

Fig 1 Methodology

IV.CONCLUSION AND FUTURE WORK

In conclusion, machine learning algorithms such as logistic regression, k-nearest neighbors (KNN), and support vector machines (SVM) have shown promising results in predicting chronic kidney disease (CKD) and its stages. These algorithms can analyze large datasets of clinical and laboratory data to identify patterns and relationships that may not be apparent to human observers.Logistic regression is a powerful algorithm for binary classification problems and has been used to predict the presence or absence of CKD. KNN, on the other hand, is a non-parametric algorithm that can be used for both binary and multi-class classification problems. SVM is another popular algorithm that can handle both linear and non-linear data and has been shown to achieve high accuracy in predicting CKD.

Overall, the use of machine learning algorithms for CKD prediction and diagnosis has the potential to improve clinical decision-making and patient outcomes. However, further research is needed to validate these algorithms and integrate them into clinical practice. Additionally, the interpretability of these algorithms needs to be improved to ensure that clinicians can understand and trust the predictions made by the algorithms.

REFERENCES

[1]Chronic Kidney Disease Prediction Using Data Mining- J.Sneha , V.Tharani, S Dhivvya Preetha

[2]Chronic Kidney Disease Prediction Using Machine Learning Models-S.Revathy,B.Bharathi,P.Jeyanthi,M.Ramesh

[3]Prediction of Chronic Kidney Disease - A Machine Learning Perspective –Pankaj Chittora, Sandeep Chaurasia, Prasun Chakrabhart,Gaurav Kumawat.Tulika Chakrabarthi, ZBIGNIEW LEONOWICZ.

[4]Machine Learning Techniques for Chronic Kidney Disease Risk Prediction Ellas Dritias and Maria Trigka .

[5]Machine Learning Prediction Models for Chronic Kidney Disease using National Health Insurance Claim Data in Taiwan-Surya Krishnamurthy,Kapellesh KS ,Erik Dovgan. .

[6]Chronic Kidney Disease Prediction using Machine Learning – Sathiya Priya S,Suresh Kumar M

[7]Global,National,Regional Burden of Chronic Kidney Disease-Dr.Theo Vos ,Dr.Boris Bikbov

[8]Machine Learning Prediction Models for Chronic Kidney Disease using National Health Insurance Claim Data in Taiwan-Anton Gradisek,Kathiravan Srinivasan.

[9]KDIGO-Clinical Practice Guideline for the evaluation and Management of Chronic Kidney Disease.

[10]Prediction of Kidney Disease Stages using Data Mining algorithms – El-Houssainy A.Randy,Ayman S Anwa