IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13378

Chronic Kidney Disease Prediction using Machine Learning

Ananya Harish Shetty[1], Jyothi Prasad[2], Manisha[3], Nishmitha S Shetty[4], Pavithra[5]

Student, Department of Computer Science & Engineering, Mangalore Institute of Technology and Engineering, Moodabidre, India[1]

Assistant Professor, Department of Computer Science & Engineering, Mangalore Institute of Technology and Engineering, Moodabidre, India[2]

Student, Department of Computer Science & Engineering, Mangalore Institute of Technology and Engineering, Moodabidre, India[3]

Student, Department of Computer Science & Engineering, Mangalore Institute of Technology and Engineering, Moodabidre, India[4]

Student, Department of Computer Science & Engineering, Mangalore Institute of Technology and Engineering, Moodabidre, India[5]

Abstract: The abstract introduces the pressing issue of Chronic Kidney Disease (CKD) and underscores the importance of early identification to mitigate its progression and enhance patient outcomes. It highlights the increasing utilization of machine learning (ML) algorithms for CKD prediction but identifies a need for more accurate and efficient models. The paper aims to fill this gap by conducting a thorough literature review on CKD prediction using ML techniques, analyzing features, datasets, algorithms, and evaluation metrics utilized in existing studies. Additionally, it proposes a novel approach that combines different feature selection and ML techniques to improve prediction accuracy. The findings demonstrate the potential of ML algorithms such as support vector machines, random forests, and neural networks to achieve high accuracy in CKD prediction, with the proposed approach enhancing accuracy by up to 5%. The implications of this study suggest the development of more effective CKD prediction models that could positively impact clinical practice and patient outcomes.

Keywords: Chronic Kidney Disease (CKD), Machine Learning (ML), Prediction, Feature Selection, Datasets, AlgorithmsEvaluation Metrics, Support Vector Machines (SVM), Random Forests, Neural Networks, Accuracy Improvement, Clinical Practice, Patient Outcomes, Healthcare Management, Early Identification.

I. INTRODUCTION

Chronic kidney disease (CKD) is a common and serious disease that affects millions of people worldwide. Early detection and prediction of CKD progression are critical to implement appropriate interventions and improve patient outcomes. In recent years, machine learning techniques have become valuable tools for accurate prediction and risk assessment in several health care domains. This study aims to explore the potential of machine learning algorithms to predict the progression of CKD. Using large data sets that include patient demographics, medical history, laboratory results, and other relevant features, we can develop robust predictive models to identify individuals at risk of CKD or disease progression. Machine learning models can learn from historical data, patterns and correlations to reveal hidden relationships and complex patterns that human experts may have difficulty detecting. Predictive models can analyze and weigh different risk factors, enabling personalized predictions based on individual patient characteristics.

Using machine learning algorithms to predict CKD could have significant implications for clinical decision making and patient management. Early identification of those at risk can facilitate preventive measures such as lifestyle changes, medication adjustments or referral to nephrology specialists. In addition, accurate prediction of CKD progression can help optimize resource allocation and health care planning. In this study, we use various machine learning algorithms such as linear regression, support vector machines, random forests, and knn, to develop predictive models for CKD. The models are trained and evaluated on real patient data, enabling us to evaluate their performance and compare their predictive ability. Finally, the development of an accurate and reliable model of prognostic chronic disease based on machine learning promises to improve patient outcomes, optimize health resources and enable early

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13378

intervention strategies. This study could potentially contribute to the development of a preventive and personalized approach to the treatment of CKD.

II. OBJECTIVE

Aim to detect CKD in its early stages, enabling timely medical intervention and treatment to improve patient outcomes: Early detection of CKD is crucial for effective management and improving patient outcomes. CKD often progresses slowly and asymptomatically in its early stages, making early detection challenging but essential. Timely medical intervention and treatment, initiated at an early stage of CKD, can help slow down disease progression, prevent complications, and improve overall outcomes for patients. By developing a predictive model capable of detecting CKD in its early stages, healthcare providers can intervene promptly with appropriate treatments, lifestyle modifications, and monitoring strategies, potentially mitigating the adverse effects of the disease.

III. PROBLEM STATEMENT

Chronic kidney disease (CKD) poses a significant burden on global health, affecting millions of people. Early detection is crucial for preventing or slowing its progression and improving patient outcomes. However, CKD often presents with non-specific symptoms, making early diagnosis challenging. Additionally, current methods for predicting CKD may not be as accurate or efficient as possible.

This highlights the need for more effective strategies in CKD detection. Machine learning (ML) offers a promising approach for improving CKD prediction accuracy. By analyzing various features from patient data, ML models can potentially identify individuals at risk for developing CKD. Refining these models for better accuracy and efficiency could be instrumental in facilitating earlier detection and improving patient care..

IV. LITERATURE REVIEW

The utilization of various data mining techniques for diagnosing chronic kidney disease (CKD) focuses on achieving reliable diagnoses rather than discovering a perfect cure. Different studies employ different approaches: one study uses Random Forest and Back Propagation Neural Network algorithms [1],
while another optimizes feature extraction using three feature selection algorithms and implements SMOTE for class balancing, achieving high accuracy of 96.77% [2].
Similarly, other studies propose data mining methodologies utilizing classic methods for data preparation and preprocessing, employing algorithms such as Decision Tree, and Rotation Forest to predict CKD onset with high accuracy 98.26% [3, 4].
Additionally, retrospective analyses utilizing Taiwan's National Health Insurance Research Database and the Chronic Kidney Disease dataset from the UCI Machine Learning Repository aim to predict CKD onset several years in advance [5, 6].
The utilization of statistical techniques to estimate chronic kidney disease (CKD) data for regions with limited primary sources. These techniques adjust for variations in data collection methods, such as different case definitions or study procedures. Additionally, adjustments are made for mortality data by assigning deaths to standardized categories, facilitating comparisons across different datasets and regions [7].
Furthermore, artificial intelligence-based models, including convolutional neural networks and LightGBM, are developed to forecast CKD likelihood based on factors like age, comorbidities, and medication use [8].
Moreover, various formulae and equations, considering factors like age, sex, ethnicity, and body size, are employed to estimate GFR or CrCl [9].
Overall, data mining techniques, both supervised and unsupervised, are instrumental in identifying patterns in CKD data, aiding in diagnosis, prediction, and understanding of the disease [10].

V. METHODOLOGY

The methodology employed in the development of the Chronic Kidney Disease (CKD) prediction web application encompasses a systematic approach aimed at creating a reliable, user-friendly tool for healthcare professionals and patients. It begins with an extensive literature review to understand existing CKD prediction methods and machine learning algorithms. Requirements are then gathered from stakeholders to define the application's functionality and usability criteria. Based on these requirements, the system architecture, user

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13378

interface, and data flow are designed, ensuring a clear understanding of the application's structure. Subsequently, relevant patient data is collected and preprocessed, followed by the selection and training of machine learning models for CKD prediction. The frontend and backend components are developed concurrently, integrating machine learning models with the application's interface and database. Rigorous testing is conducted to validate the functionality, accuracy, and usability of the application. Upon successful testing, the application is deployed to a production environment, with ongoing maintenance and updates to address any issues and enhance features. Finally, the performance of the prediction model and the application's effectiveness in real-world scenarios are validated and evaluated. Through this methodology, the CKD prediction web application aims to provide accurate predictions, facilitate proactive healthcare interventions, and ultimately improve patient outcomes in the management of chronic kidney disease

a) System Architecture

The system architecture of the proposed Chronic Kidney Disease (CKD) prediction web application encompasses a frontend interface for user interaction, a backend server for processing requests and executing machine learning algorithms, and a database for storing patient data. Machine learning models trained on patient demographic and medical information are deployed on the backend server to generate predictions. Integration layers ensure seamless communication between components, while security measures safeguard sensitive data. Optionally, the system may integrate with external APIs for additional data sources. Overall, the architecture aims to provide an intuitive, accurate, and secure platform for CKD

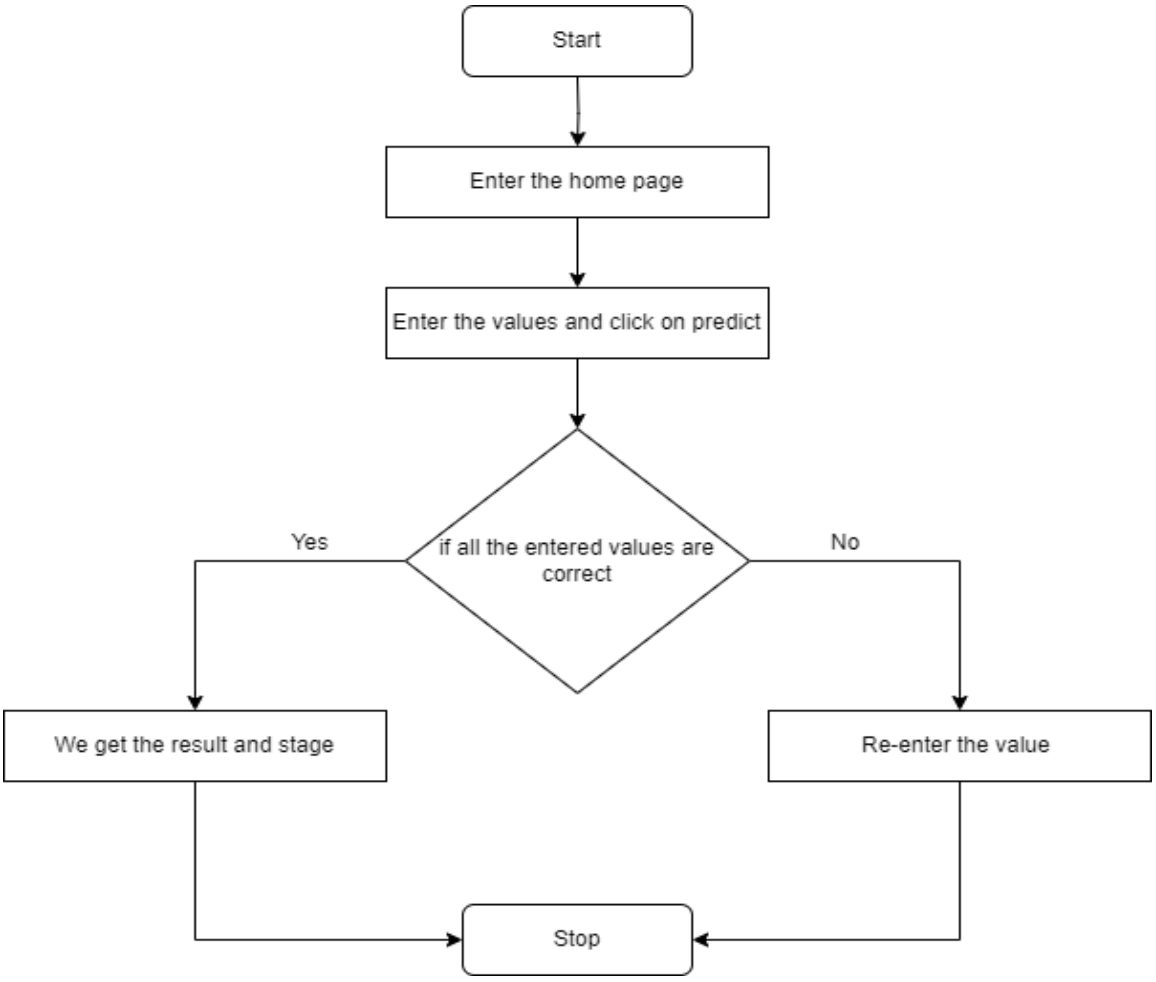


Fig 1.System Architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

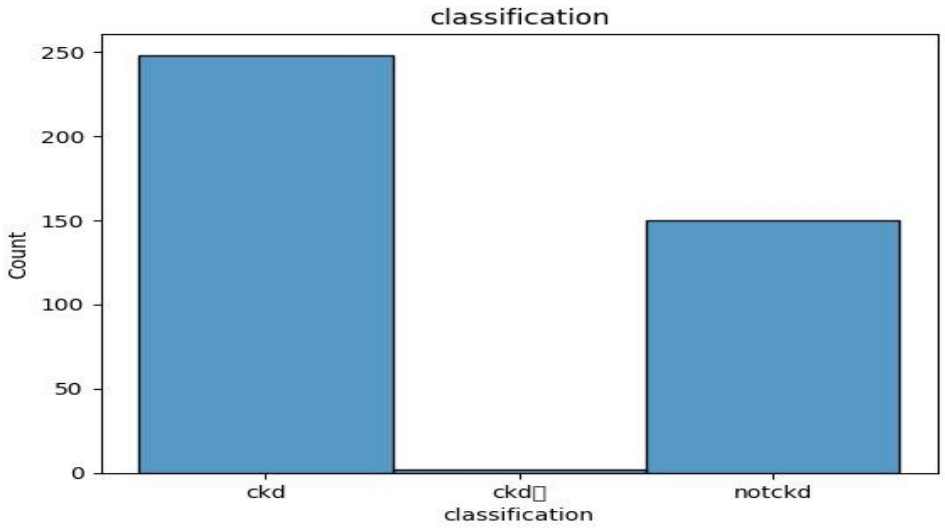
DOI: 10.17148/IJARCCE.2024.13378

VI. RESULT

The model was trained using logistic regression methodologies, and we achieved an impressive accuracy of 99.02% utilizing the customized logistic regression model for chronic kidney disease (CKD) prediction.

	precision	recall	f1-score	support
0	0.97	1.00	0.98	28
1	1.00	0.98	0.99	52
accuracy			0.99	80
macro avg	0.98	0.99	0.99	80
weighted avg	0.99	0.99	0.99	80

Fig. 2 Accuracy of the model



VII. CONCLUSION

In conclusion, the development of a web-based application for predicting Chronic Kidney Disease (CKD) holds significant promise in improving healthcare delivery and patient outcomes. By leveraging machine learning algorithms and integrating them into a user-friendly interface, the application enables accurate and accessible CKD risk assessment and prediction. Healthcare professionals can utilize the application for early identification of at-risk individuals, while patients can actively monitor their kidney health and make informed decisions. Additionally, the application fosters research by collecting anonymized patient data for analysis and refinement of predictive models. While challenges such as data privacy and validation remain, the potential benefits of the proposed system in revolutionizing CKD management are substantial. Overall, the development of such an application represents a crucial step towards proactive healthcare intervention and improved patient care in the realm of chronic kidney disease.

REFERENCES

[1]Sneha, J., Tharani, V., and Dhivvya Preetha, S. explored the prediction of Chronic Kidney Disease (CKD) using data mining techniques.

[2]Revathy, S., Bharathi, B., Jeyanthi, P., and Ramesh, M. investigated the prediction of CKD employing various machine learning models.

[3]Chittora, P., Chaurasia, S., Chakrabhart, P., Kumawat, G., Chakrabarthi, T., and Leonowicz, Z. examined CKD

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13378

prediction from a machine learning perspective.
[4]Dritias, E., and Trigka, M. explored the application of machine learning techniques for predicting the risk of CKD.
[5]Chen, Krishnamurthy, S., Kapellesh KS, and Dovgan, E. developed machine learning prediction models for CKD using National Health Insurance Claim Data in Taiwan.
[6]Sathiya Priya S. and Suresh Kumar M. investigated the prediction of CKD using machine learning methods.
[7]Vos, T., and Bikbov, B. discussed the global, national, and regional burden of CKD.
[8]Gradisek, A., and Srinivasan, K. developed prediction models for CKD utilizing National Health Insurance Claim Data in Taiwan.
[9]The KDIGO Clinical Practice Guideline outlines recommendations for the evaluation and management of Chronic Kidney Disease.
[10]El-Houssainy, A.R., and Anwa, A.S. explored the prediction of kidney disease stages using data mining algorithms