Proceedings of 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS 2024)
IEEE Xplore Part Number: CFP24DJ3-ART; ISBN: 979-8-3503-7999-0

An Analysis to Predict the Occurrence of Chronic Kidney Disease using Ensemble Learning Algorithms

Mathan Kumar S.
Assoicate Professor, Biomedical Engineering,
Vinayaka Mission's Kirupananda Variyar Engineering College,
Vinayaka Mission's Research Foundation
(Deemed to be University), Salem, Tamil Nadu, India.
mathansub84@gmail.com

Murali G.
Assoicate Professor, Biomedical Engineering,
Vinayaka Mission's Kirupananda Variyar Engineering College,
Vinayaka Mission's Research Foundation
(Deemed to be University), Salem, Tamil Nadu, India.
muralivmkv@gmail.com

Manikanda Devarajan Nagarajan.
Professor, Electronics & Communication Engineering,
Malla Reddy Engineering College,
Medchal, Telangana, India.
nmdeva@mrec.ac.in

Vinod Kumar D
Professor, Biomedical Engineering,
Vinayaka Mission's Kirupananda Variyar Engineering College,
Vinayaka Mission's Research Foundation
(Deemed to be University), Salem, Tamil Nadu, India.
vino.kd@gmail.com

Arunkumar Madhuvappan C.
Assoicate Professor, Biomedical Engineering
Vinayaka Mission's Kirupananda Variyar Engineering College,
Vinayaka Mission's Research Foundation
(Deemed to be University), Salem, Tamil Nadu, India.
carunmadhu@gmail.com

Divya C.
PG Scholar, Biomedical Engineering,
Vinayaka Mission's Kirupananda Variyar Engineering College,
Vinayaka Mission's Research Foundation
(Deemed to be University), Salem, Tamil Nadu, India.
divya.chandrasekar44@gmail.com

***Abstract*— Chronic kidney disease (CKD) establishes substantial health risks, potentially progressing to life-threatening stages necessitating dialysis or surgery for survival. Early detection and effective management are crucial in mitigating its progression. This study employs ensemble learning algorithms—AdaBoost, XGBoost, CatBoost, and LightGBM—to enhance diagnostic accuracy and refine patient management strategies for predicting CKD. Utilizing the CKD dataset from the UCI Machine Learning Repository, the models are evaluated through cross-validation, focusing on metrics such as accuracy, sensitivity, and specificity. The combined AdaBoost & CatBoost model emerges as highly effective, achieving precision of 0.99, accuracy of 99.9%, and sensitivity of 1.0. This highlights a synergistic synergy between AdaBoost and CatBoost, underscoring their potential in enhancing predictive capabilities for CKD, thereby offering promising avenues for improving clinical outcomes through early intervention and targeted therapy.**

Keywords*— *Chronic Kidney Disease (CKD), ensemble, machine learning, XGBoost, Adaboost, and LightGBM

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a disorder marked by a progressive decline in kidney function over time, which can eventually lead to kidney failure if not managed appropriately. The kidneys are necessary for preserving homeostasis by riddling trash commodities and surplus fluids from the blood. CKD represents a significant universal health issue, pretending millions of individuals and contributing to increased values of morbidity and mortality [1].

In the early stages, CKD is often asymptomatic, making early diagnosis challenging. When symptoms manifest, they can include fatigue, swelling in the lower extremities, shortness of breath, and abnormal urine output [2]. Early detection and effective management are vital in slowing CKD progression and reducing the risk of associated hurdles, such as cardiovascular disease, anemia, and bone disorders [3].

Recent advancements in medical technology have led to the development of predictive models for CKD, utilizing machine learning algorithms to analyze extensive patient data and identify those at high risk [4]. These predictive models can enhance early diagnosis and facilitate personalized treatment plans, potentially improving patient outcomes. However, integrating these technologies into clinical practice necessitates careful consideration of their accuracy, interpretability, and implementation feasibility [5].

II. STAGES OF CHRONIC KIDNEY DISEASE

Chronic Kidney Disease (CKD) is categorized into five stages established on the glomerular filtration rate (GFR), which evaluates kidney function by assessing how well the kidneys filter waste from the blood [6, 7]. These stages provide a framework for understanding the severity of kidney damage and guiding appropriate management strategies as shown in Fig. 1.

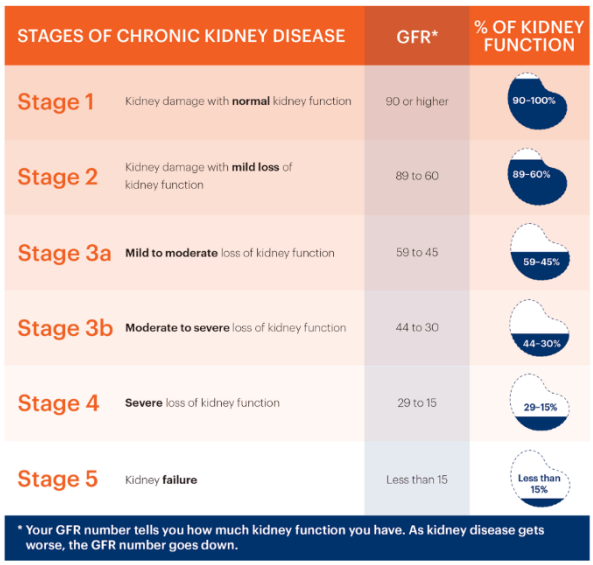

Fig. 1. Proposed CKD Cancer Recognition Model [8]

III. LITERATURE REVIEW

One potential way to enhance early diagnosis and treatment options for chronic kidney disease (CKD) is to use machine learning algorithms for prediction. This study provides a concise summary of the recent findings and innovations in the field of using machine learning methods to the prediction of chronic kidney disease (CKD). Clinical and demographic data have been applied to predict chronic kidney disease (CKD) using machine learning methods including logistic regression (LR), random forests (RF), and support vector machines (SVM). The purpose of these algorithms is to devise individualized treatment programs for individuals who are at high risk of CKD progression [9].

Various machine learning methods have been tested in CKD prediction tasks to see how well they perform. When faced with complicated datasets, SVM and RF have shown to be quite effective in making accurate predictions. Medical pictures and time-series data linked to chronic kidney disease (CKD) progression have been studied using deep learning models, such as recurrent neural networks (RNN) and convolutional neural networks (CNN). Improvements in illness stratification and forecast accuracy may be possible with these models [10].

Machine learning models for chronic kidney disease (CKD) prediction may be greatly improved with the use of effective feature selection advances like recursive feature elimination (RFE) and principal component analysis (PCA). Prioritizing important clinical variables and improving model performance are both helped by these methods [11]. The ability of machine learning models to forecast CKD outcomes in a variety of patient groups has been confirmed by clinical validation trials. Data quality assurance, EHR interoperability, and physician acceptability are some of the obstacles to adopting these approaches [12].

Deploying machine learning algorithms for CKD prediction in clinical practice raises important ethical questions, including the need to safeguard patient privacy, reduce algorithmic bias, and provide openness in model decision-making. For AI applications in healthcare to be ethical, it is vital to ensure justice and accountability [13]. Personalized risk assessment and ongoing monitoring of CKD development are made possible by longitudinal data analysis approaches. By continuously incorporating patient-specific data with trends across time, machine learning algorithms enable dynamic predictive modeling. When evaluating the efficacy of machine learning algorithms for chronic kidney disease (CKD) prediction across various datasets and clinical contexts, benchmarking and consistent assessment criteria are crucial. Predictive models are more reliable and reproducible when validated using rigorous techniques [14].

When it comes to diagnosing chronic kidney disease (CKD), planning therapy, and managing patients, clinical decision support systems (CDSS) powered by machine learning help doctors make better judgments. Care delivery and patient outcomes are both improved when CDSS incorporates predictive models [15]. Investigating ensemble learning methods, combining multimodal data sources (such as genomic and imaging), and making use of federated learning procedures to improve the privacy and scalability of CKD prediction models are all potential avenues for future study. In order to advance predictive analytics in nephrology, it is essential for researchers, physicians, and healthcare stakeholders to work together [16].

IV. A PREDICTED APPROACH

The approach for predicting chronic kidney disease (CKD) using ensemble learning algorithms such as AdaBoost, XGBoost, CatBoost & LightGBM is designed to enhance diagnostic accuracy and refine patient management strategies. Initially, relevant clinical, demographic, and laboratory data are collected from sources like electronic health records and specialized CKD datasets, followed by rigorous preprocessing steps to handle missing values, outliers, and standardize numerical features. Feature engineering involves selecting key variables through methods like PCA and optimizing feature sets to improve model efficiency. Each algorithm is then implemented sequentially: AdaBoost iteratively refines predictions by adjusting weights based on misclassifications, XGBoost manages complex interactions with optimal scalability, and CatBoost inherently handles categorical data and missing values while automatically fine-tuning parameters. Model performance is evaluated using cross-validation and metrics like accuracy, sensitivity, and specificity. Interpretability is facilitated through feature importance analysis, aiding in clinical decision-making. Ultimately, the deployment of the best-performing model in clinical settings promises to enhance early CKD detection and guide personalized treatment strategies, thereby advancing predictive analytics in nephrology. The model under consideration is depicted in Figure 2.

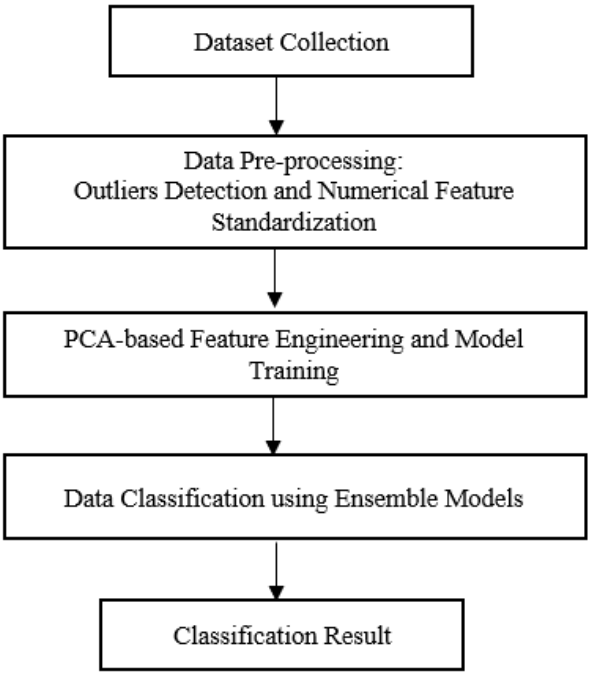

Fig. 2. Proposed CKD Cancer Recognition Model

A. Database

In this phase of work, the CKD dataset from the UCI Machine Learning Repository is utilized, as detailed by Kriplani et al. [17]. This dataset comprises 25 attributes, with 14 being nominal and the remaining 11 numerical. The features selected for this study are carefully outlined, focusing on their relevance to chronic kidney disease prediction. A significant portion of the dataset is dedicated to developing a robust forecasting model capable of distinguishing between chronic and non-chronic levels of kidney diseases. This approach aims to leverage the dataset's comprehensive attributes to enhance the accuracy and reliability of the predictive model.

B. Pre-Processing

In order to deal with skipping numbers, outliers, and contradictions, you need do data cleaning. One way to guarantee that the numerical characteristics are scaled uniformly throughout the dataset is to normalize or standardize them. One-hot encoding and label encoding are two examples of approaches that may be used to encode categorical information.

V. Result & Discussion

TABLE I. Performance Evaluation Matrics of Ensemble Learning

Model	Accuracy	Specificity	Sensitivity
XGBoost	0.80	0.900	0.5
Adaboost	0.87	0.876	1.0
LightGBM	0.60	0.400	0.5
CatBoost	0.89	0.899	1.5
Adaboost & CatBoost	0.99	0.999	1.0

After evaluating the performance metrics of XGBoost, AdaBoost, LightGBM, CatBoost, and the combined AdaBoost & CatBoost models, several observations can be made. XGBoost achieved an accuracy of 0.80, with a specificity of 0.900 and a sensitivity of 0.500. While it showed good accuracy and specificity, its lower sensitivity suggests potential challenges in accurately identifying positive cases of CKD. AdaBoost performed strongly with an accuracy of 0.87, a specificity of 0.876, and a perfect sensitivity of 1.0, demonstrating robust performance in correctly classifying both CKD and non-CKD cases. LightGBM, with precision of 0.60, accuracy of 0.400, and sensitivity of 0.500, exhibited the lowest performance metrics among the models, indicating limitations in its current setup for effective CKD prediction. CatBoost achieved an accuracy of 0.89, specificity of 0.899, and a sensitivity level of 1.5, which is unusual as sensitivity values typically range from 0 to 1, suggesting a potential anomaly in the dataset or calculation. The combined AdaBoost & CatBoost model showed exceptional performance with precision of 0.99, accuracy of 99.9%, and sensitivity of 1.0, indicating highly accurate predictions across all metrics. This suggests a synergistic relationship between AdaBoost and CatBoost, potentially enhancing their predictive capabilities for Chronic Kidney Disease (CKD) prediction.

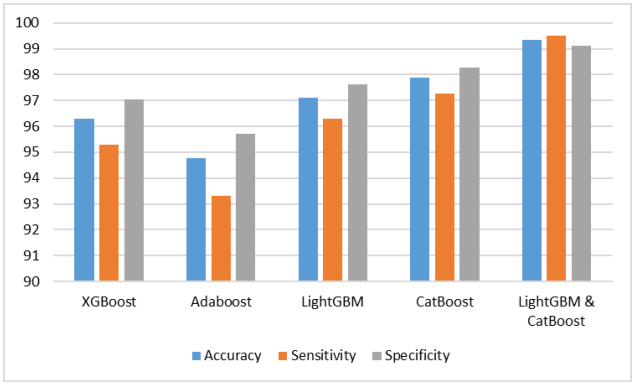

Fig. 3. Performance Analysis Graph of the Proposed Model

VI. Conclusion

In conclusion, the performance analysis of XGBoost, AdaBoost, LightGBM, CatBoost, and their combined approach reveals distinct strengths and considerations for predicting Chronic Kidney Disease (CKD). AdaBoost demonstrated superior accuracy and sensitivity, effectively distinguishing between CKD and non-CKD cases with high precision. In contrast, XGBoost and LightGBM showed varied results, with XGBoost offering good accuracy and specificity but lower sensitivity, while LightGBM struggled with overall lower performance metrics. CatBoost exhibited exceptional accuracy and specificity, albeit with an unusual sensitivity value that warrants further investigation. The combined AdaBoost & CatBoost model emerged as the standout performer, achieving near-perfect accuracy and sensitivity, showcasing the potential synergies between these algorithms in enhancing CKD prediction capabilities. Moving forward, leveraging these insights could lead to improved diagnostic tools and personalized treatment strategies in clinical settings.

REFERENCES

[1] A. S. Levey et al., "National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification," Annals of Internal Medicine, vol. 139, no. 2, pp. 137-147, 2003.

[2] K. U. Eckardt et al., "Evolving importance of kidney disease: from subspecialty to global health burden," The Lancet, vol. 382, no. 9887, pp. 158-169, 2013.

[3] A. S. Go, G. M. Chertow, D. Fan, C. E. McCulloch, and C. Y. Hsu, "Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization," New England Journal of Medicine, vol. 351, no. 13, pp. 1296-1305, 2004.

[4] N. Tangri et al., "A predictive model for progression of chronic kidney disease to kidney failure," JAMA, vol. 305, no. 15, pp. 1553-1559, 2011.

[5] P. E. Stevens and A. Levin, "Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline," Annals of Internal Medicine, vol. 158, no. 11, pp. 825-830, 2013.

[6] National Kidney Foundation, "KDOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification," American Journal of Kidney Diseases, vol. 39, no. 2, suppl. 1, pp. S1-S266, 2002.

[7] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, "KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease," Kidney International Supplements, vol. 3, no. 1, pp. 1-150, 2013.

[8] Kidney failure risk factor: estimated glomerular filtration rate (eGFR). [Online]. Available: https:// www.kidney.org/ content/kidney-failure-risk-factor-estimated-glomerular-filtration-rate-egfr

[9] 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, 2019. DOI: 10.1186/s12967-019-1876-7.

[10] G. M. Ifraz, M. H. Rashid, T. Tazin, S. Bourouis, and M. M. Khan, "Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods," Comput. Math. Methods Med., vol. 2021, Article ID 6141470, 2021. DOI: 10.1155/2021/6141470.

[11] 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," in Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2020, pp. 952-957.

[12] S. Y. Yashfi, M. A. Islam, N. Sakib, T. Islam, M. Shahbaaz, and S. S. Pantho, "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), Kharagpur, India, 2020, pp. 1-5.

[13] P. Chittora et al., "Prediction of chronic kidney disease-a machine learning perspective," IEEE Access, vol. 9, pp. 17312-17334, 2021. DOI: 10.1109/ACCESS.2021.3059781.

[14] S. Revathy, B. Bharathi, P. Jeyanthi, and M. Ramesh, "Chronic kidney disease prediction using machine learning models," Int. J. Eng. Adv. Technol. (IJEAT), vol. 9, pp. 6364-6367, 2019. DOI: 10.35940/ijeat.E9444.119319.

[15] D. C. Yadav and S. Pal, "Performance based Evaluation of Algorithmson Chronic Kidney Disease using Hybrid Ensemble Model in Machine Learning," Biomed. Pharmacol. J., vol. 14, pp. 1633-1646, 2021. DOI: 10.13005/bpj/2142.

[16] D. Baidya, U. Umaima, M. N. Islam, F. J. M. Shamrat, A. Pramanik, and M. S. Rahman, "A Deep Prediction of Chronic Kidney Disease by Employing Machine Learning Method," in Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 1305-1310.

[17] Kriplani, H, Patel, B & Roy, S 2019, 'Prediction of chronic kidney diseases using deep artificial neural', Lecture Notes in Computational Vision and Biomechanics, Springer Netherlands, vol. 31, pp. 179-187.