Smart Chronic Kidney Disease Detection and Progression Monitoring System Using Machine Learning

Thejaswini D, Abhishek C, Dinesh G, Gowtham C, Hitesh R C
Department of CSE, Don Bosco Institute of Technology, Bengaluru, Karnataka, India

thejaswinid@dbit.ac.in, abhishekurs123456@gmail.com,dineshvydvas7900@gmail.com, gowthamcgowdadbit@gmail.com, hiteshgowda3138@gmail.com

Abstract
Chronic Kidney Disease (CKD) is a long-term illness that often progresses quietly, with little to no symptoms in the early stages. Because of this, many individuals miss the opportunity for timely care. In this project, we've developed an intelligent system that not only detects the presence of CKD using clinical data but also helps determine how far the disease has progressed — from stage 1 to stage 5. Our approach combines patient metrics such as age, blood pressure, serum creatinine, and hemoglobin levels to train a set of machine learning models. Unlike many existing tools that simply flag whether CKD is present, our system goes further — offering detailed stage-wise predictions. The backend leverages an ensemble of proven algorithms including Support Vector Machines, Random Forests, and Multi-Layer Perceptrons, known for their effectiveness in medical diagnosis tasks. What sets our system apart is its focus on practical usability. It's built not just for accuracy, but also to support everyday clinical use. The interface provides clear reports, personalized insights based on disease stage, and even guidance for ongoing monitoring. This can be especially valuable in rural clinics, telemedicine settings, or primary care centers where access to specialists is limited. With the right deployment, this tool can support faster, earlier interventions and ultimately help improve patient outcomes for those living with CKD.

Keywords

Chronic Kidney Disease, Machine Learning, Early Diagnosis, Stage Prediction, GFR, Health Monitoring, Clinical data analysis, Patient Management

Introduction

Kidney-related disorders like CKD often progress in silence, with symptoms appearing only when much of the damage has already occurred. Unfortunately, this delay in detection can limit treatment choices and worsen patient outcomes. That's where technology — particularly machine learning — steps in. Unlike traditional methods, ML models can analyze large volumes of health-related data and uncover hidden patterns that may point to early signs of CKD. These models, trained on features like age, blood test results, and medical history, can help physicians predict disease onset and progression more proactively. In this way, the fusion of clinical knowledge and AI can lead to more tailored, efficient, and timely treatment strategies for patients. Machine learning has existing examples of being used to improve diagnostic speed and accuracy, yet current applications are primarily limited to binary prediction - simply

whether a patient has CKD or not - and do not predict stage of disease.

This project proposes a smart system that detects CKD, as well as identifies the stage of the disease using medically accepted methods. The smart system uses 3 types of machine learning algorithms (Random Forest, Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP)) and brings them into ensemble model (not ensemble learning) to give more accuracy and reliability. The smart system was built to be user-friendly, with specific features such as personalized health guidelines and reports suitable for patients, and is expected to perform well in clinical occupations (healthcare practitioners) and also be applied to platforms and products that relate to chronic health management .The purpose of the smart system is to contribute to a useful and interpretable solution to support healthcare providers in early intervention, especially for patients who lack access to specialist care.

Literature Review

Geetha et al. [1] (2025): in their risk prediction study of CKD, evaluated several machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP), alongside ensemble methods like hard and soft voting. Their comparative study found that SVM achieved the maximum individual accuracy of 96.25%, and ensemble models demonstrated enhanced robustness against data variability. The authors concluded that these ensemble-based models could be highly beneficial in developing reliable decision-support tools for clinicians, aiding in the early and effective management of chronic kidney disease. This research contributes to addressing CKD as a growing global health issue through advanced predictive techniques.

Pathak et al. [2] (2025): in their work on predicting chronic kidney disease, aimed to enhance early and precise detection to prevent further health deterioration. They emphasized the necessity of reliable prediction algorithms due to CKD's propensity to advance to end-stage renal disease. Their study explored various machine learning methods, such as Random Forest, k-Nearest Neighbors (KNN), Gradient Boosting, and Naive Bayes, to improve CKD prediction. This research contributes to better management of CKD by providing insights into effective predictive methodologies.

Nandi et al. [3] (2025): in their paper on predictive modeling for Chronic Kidney Disease, proposed a hybrid approach that integrates exploratory data analysis, machine learning, and Explainable AI (XAI) techniques. The objective of this integration was to enhance both the accuracy and the interpretability of CKD prediction models. This methodology aims to provide more transparent and understandable predictions, which is crucial for clinical decision-making in managing CKD. Their work highlights the benefits of combining different analytical methods to improve the robustness and clarity of disease prediction.

Chantamit-o-pas et al. [4] (2025): in their study, evaluated the effectiveness of an interactive robot-assisted avatar health education program for Chronic Kidney Disease patients. They aimed to assess the program's impact on patients' knowledge and health literacy, addressing the substantial health and economic challenges posed by CKD globally. This research indicates that innovative educational tools can play a significant role in empowering CKD patients with better understanding and self-management skills. The findings suggest a promising avenue for improving patient engagement and health outcomes in CKD care.

Dhuriya et al. [5] (2025): in their study on Chronic Kidney Disease, investigated data clustering and prediction using various machine learning algorithms. They highlighted CKD as a major worldwide health concern with high mortality and morbidity rates, often missed in early stages due to asymptomatic nature. Their research aims to

improve early detection and management of CKD by applying these algorithms, contributing to more timely interventions. This work emphasizes the importance of machine learning in addressing the global health crisis posed by CKD.

Rani et al. [6] (2024)in their study on XGBoost-driven insights for enhancing Chronic Kidney Disease detection, underscored the critical importance of early detection and treatment to halt disease progression. They noted that recent advancements in machine learning have significantly improved pattern detection and analysis in large datasets, thereby enhancing CKD diagnosis and prediction. Their work leverages XGBoost to achieve more accurate and efficient identification of CKD, contributing to better management strategies. This research emphasizes the role of advanced ML techniques in addressing CKD as a major global health issue.

Chen et al. [7] (2024) in their study, evaluated the effectiveness of an interactive robot- assisted avatar health education program for Chronic Kidney Disease patients. They aimed to assess the program's impact on patients' knowledge and health literacy, addressing the substantial health and economic challenges posed by CKD globally. This research indicates that innovative educational tools can play a significant role in empowering CKD patients with better understanding and self-management skills. The findings suggest a promising avenue for improving patient engagement and health outcomes in CKD care.

Bhuria [8] (2024): in their paper on advanced machine learning techniques for Chronic Kidney Disease prediction and management, identified CKD as one of the top five deadly diseases worldwide. The study introduced a highly accurate and sophisticated machine learning approach for CKD prediction, achieving an accuracy rate of 97.5%. The model demonstrated flawless precision and comprehensiveness (100%) for the non-CKD category, ensuring reliable identification. This research highlights the potential of machine learning to significantly improve early detection and enable timely interventions for CKD

Sirisha et al. [9] (2024): presented a predictive model for Chronic Kidney Disease stage prediction using Stochastic Gradient Descent (SGD) with ElasticNet regularization. They highlighted CKD as a progressive condition requiring accurate stage prediction for timely interventions and effective management. Their model aims to classify CKD stages efficiently, utilizing ElasticNet regularization to enhance performance by preventing overfitting and addressing multicollinearity. The findings suggest that this method achieves high accuracy and robustness in predicting CKD stages, surpassing traditional regression methods.

Sonone [10] (2024): in their report on early prediction and progression of Chronic Kidney Disease, described a groundbreaking study utilizing machine learning approaches with real-time clinical datasets. The research aimed to develop predictive models for early CKD detection and progression tracking, providing reliable insights into disease development. By combining various clinical tests and patient data, the proposed method effectively evaluates longitudinally gathered information. This study enhances the effectiveness of machine learning algorithms for CKD by incorporating ensemble approaches, improving accuracy and interpretability.

Arora et al. [11] (2024): analyzed machine learning algorithms for Chronic Kidney Disease prediction, emphasizing CKD as a critical global health concern. They highlighted the urgency of early detection due to CKD's asymptomatic early stages and its significant contribution to global mortality. Their study focused on developing a predictive model to prevent End-Stage Renal Disease (ESRD) and facilitate efficient treatment. This research contributes to improving diagnostic tools by evaluating various machine learning algorithms for robust CKD prediction

Sneha [12] (2024): in their research on developing a supervised algorithm for predicting Chronic Kidney Disease, focused on addressing the serious condition caused by renal function impairment. The study discussed the importance of early detection and intervention to prevent life-threatening complications. They employed an ensemble method, boosting, specifically focusing on the AdaBoost technique, alongside K-Nearest Neighbour (KNN) and Random Forest. This research contributes to identifying effective machine learning methods for early CKD diagnosis and avoiding fatal issues.

Sharma and Verma [13] (2024): in their analysis for predicting Chronic Kidney Disease with machine learning models, identified CKD as a major global health crisis with a high risk of progressing to end-stage renal disease. They stressed the vital importance of early detection and diagnosis for prevention and timely treatment, as unhealthy dietary habits and insufficient water consumption are significant causes. Their work applies machine learning techniques like KNN, Decision Tree, and Random Forest to accurately predict CKD. This study underscores the efficacy of these models in improving patient survival through early intervention

Makaraju and Mande [14] (2023): developed a Python-based machine learning model for Chronic Kidney Disease classification, recognizing CKD as a chronic illness with often asymptomatic early stages. They highlighted the critical need for early detection and treatment to prevent severe consequences arising from waste and fluid buildup. Their model incorporates various input features, including baseline health measurements, medication usage, medical history, and demographic data, to predict CKD risk. The primary objective is to estimate CKD risk, outputting "1" for high-risk patients in stages 3 to 5.

Anurag et al. [15] (2023): in their study on Chronic Kidney Disease prediction using a robust machine learning approach, addressed CKD as a degenerative condition impairing kidney function and causing waste buildup. They asserted that early CKD prediction and identification can halt or delay the disease's progression. Their method combines feature selection and classification algorithms using demographic, clinical, and lab data to pinpoint key risk factors and provide accurate predictions. This approach aims to facilitate effective therapies and improve outcomes for CKD patients.

Kashyap et al. [16] (2022): in their study on predicting chronic disease in kidneys using machine learning classifiers, aimed to overcome the challenges associated with late diagnosis of CKD. They noted that advanced stages of CKD often leave patients with only costly options like dialysis or kidney transplantation. Their research employed machine learning classifiers to predict CKD, striving for early detection to enable timely interventions and avoid severe consequences. This work emphasizes the potential of ML to improve the management of this widespread health concern.

Comparative Analysis

Reference	Methodology	Models Used	Key Features	Limitations
[1]	Ensemble model with RF, SVM, MLP	RF, SVM, MLP, Hard & Soft Voting	SVM accuracy: 96.25%, ensemble robustness, real-time intent	Ensemble (hard voting) lower performance, no deep learning used

[2]	Model benchmarking with six algorithms	RF, KNN, Gradient Boosting, Naive Bayes, CatBoost, AdaBoost	100% accuracy in top models, 400 UCI samples	Focused only on classification; no XAI, real-world deployment untested
[3]	Exploratory analysis + XAI (LIME)	SVM, Decision Trees, XAI (LIME)	Explainability, 17% CKD prevalence context, hybrid ML models	Model validation inconsistencies, not integrated clinically
[4]	Chatbot deployment using NLP models; clinical pilot with patients	Dialogflow, Gemini, GPT-3.5	Focus on stage 3 CKD; integration with LINE app; real-world usability tested	Only 50 patients; limited language support (Thai); accuracy varies by NLP model
[5]	Clustering, classification, imputation for missing data	Decision Tree, Random Forest, Naïve Bayes, SVM, KNN	High precision (Random Forest 99%); KNN imputation used for missing data	No new dataset; relies on UCI CKD dataset; lacks clinical integration
[6]	Feature selection, XGBoost, Confusion matrix evaluation	XGBoost	High accuracy (98.33%), F1-score 98.63%, No false positives	Limited dataset, lacks generalization analysis
[7]	Quasi-experimental study with digital avatars	Robot-assisted avatars, GEE analysis	Focus on health literacy, 12-unit program, personalized education	Small sample (n=30), limited scalability
[8]	Feature selection + classification, hybrid ML	Random Forest, Logistic Regression, Extra Trees, SVM	97.5% accuracy, hybrid techniques, no false positives	Recall slightly lower (93.75%), generalizability not discussed
[9]	SGD classifier with ElasticNet regularization	SGD, ElasticNet, Streamlit App	Deployed via Streamlit, feature selection, 85% accuracy	Lower accuracy than other models, lacks clinical dataset variability
[10]	Longitudinal analysis + real-time dataset	Ensemble Learning, Clinical Test Integration	Tracks CKD progression, real-time insights, biomarker use	Implementation details sparse, lack of tool deployment

[11]	Compared multiple ML models on a dataset of 9993 entries	Random Forest, SVM, Logistic Regression, KNN, Naïve Bayes, Decision Tree	Thorough comparative evaluation of multiple algorithms; 100% accuracy with hybrid models	No details on real-time deployment; lacks clinical trial validation
[12]	Supervised ML with training/testing split	Random Forest, AdaBoost, KNN	Ensemble methods used; practical clinical implications discussed	Dataset size limited (400 entries); doesn't include external validation
[13]	Dataset from UCI CKD, preprocessing, ML model evaluation	Decision Tree, Random Forest, KNN	High accuracy (Random Forest 97%); emphasis on metrics like precision and F1- score	Only benchmarked traditional ML models; lacks real-time system context
[14]	Feature selection + classification	ML (unspecified mix), Demographics + Labs	89% accuracy, personalized predictions, intelligent diagnosis	Accuracy lower than top models, lacks deployment or app layer
[15]	Data preprocessing, feature selection, binary classification (0/1 for CKD stage)	Python, Logistic Regression	Utilized demographic, medical, medication, and baseline health data; focuses on early	No mention of model comparison; limited
			detection	discussion of dataset performance
[16]	Comparative ML model analysis using UCI dataset	SVM, KNN, Random Forest, Decision Tree	Preprocessing with SMOTE, eGFR staging, model accuracy comparison	Lacks model explainability, small-scale testing

Methodology

The proposed system is built using a structured pipeline that begins with data preprocessing and ends with prediction and staging of Chronic Kidney Disease (CKD). The workflow involves several key steps:

1. Data Collection

Clinical data is obtained from the UCI CKD dataset, which contains relevant medical attributes such as age, blood pressure, specific gravity, albumin, serum creatinine, haemoglobin, and others. This dataset provides a reliable foundation for training and testing.

2. Data Preprocessing

Missing values are handled using techniques like mean/mode imputation. Categorical variables are encoded

numerically. Features are then normalized to ensure consistent model input.

3. Feature Selection

Techniques such as Recursive Feature Elimination (RFE) and correlation analysis are used to identify the most significant attributes contributing to CKD prediction. This step reduces noise and improves model performance.

4. Model Training

Multiple machine learning algorithms — including Support Vector Machine (SVM), Random Forest (RF), and Multi-Layer Perceptron (MLP) — are trained individually. Then, an ensemble model is created using soft voting to combine their predictions for better accuracy and reliability.

5. Stage Classification

CKD stage is determined based on the estimated Glomerular Filtration Rate (eGFR), calculated using the Modification of Diet in Renal Disease (MDRD) formula from serum creatinine, age, gender, and race.

6. Model Evaluation

The models are assessed using standard metrics: accuracy, precision, recall, and F1-score. SHAP is integrated to explain which features influenced the output, enhancing model transparency.

7. System Interface

A simple web interface allows users to enter clinical parameters. The system returns whether the patient is likely to have CKD and if so, estimates the stage, along with general lifestyle recommendations.

Expected Results

.The system should successfully classify CKD stages from **1 to 5**, with stage-wise accuracy above **85%**.

The web-based application will provide a user-friendly experience for both patients and medical professionals, including:

- Instant prediction results
- Stage-based lifestyle suggestions
- Downloadable reports for record-keeping

The overall system will offer a reliable, cost-effective, and scalable solution for **early CKD detection and stage monitoring**, especially useful in under-resourced healthcare settings.

Conclusion

Chronic Kidney Disease represents a serious and increasing health threat. Its "silent" progression and impact on patients' quality of life are significant factors of concern. With this project we have developed a machine learning-based system that identifies individuals with risk for CKD, and captures distinct clinical stages of disease if diagnosed. The experimented system built in this research relies on the combination of using several algorithms with clinically pertinent descriptive features, and aims to help healthcare professionals to provide improved and timely decisions.

What makes this solution unique is that is has a combination of accurate prediction and usability. The solution was designed with stage-wise detection, personalized recommendations, and easy to use interface to facilitate using the system in environments outside of research: hospitals, community health center, and even remote areas. Together, the solution represents an example of how machine learning research can be translated into clinical application.

In future iterations, this system can be built out with additional biomarkers, linked to electronic health records, and

paired with real-time monitoring for chronic disease management. Ultimately, this project demonstrates how data-driven tools may facilitate early detection and significantly advance preventive health.

References

[1] Geetha M, Suganthe R C, Sowrabha A, & Sudharshan P. (2025). Chronic Kidney Disease Risk Prediction Using Machine Learning. 2025 3rd International Conference on Disruptive Technologies (ICDT), DOI: 10.1109/ICDT63985.2025.10986441

[2] Pathak, A., Singh, O. P., & Biswas, U. (2025). Predicting chronic kidney disease using machine learning methodologies. 2025 8th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), DOI: 10.1109/IEMENTECH65115.2025.10959624

[3] Nandi, A., Ghosh, A., Paul, P., Singh, A., Das, A. K., & Gupta, A. (2025). Predictive Modelling for Chronic Kidney Disease: A Hybrid Approach Using Exploratory Data Analysis, Machine Learning and Explainable AI. 2025 8th International Conference on Electronics, Materials Engineering&Nano-Technology(IEMENTech),DOI: 10.1109/IEMENTECH65115.2025.10959403

[4] Chantamit-o-pas, P., Kulsuwan, S., Suflhiwichienchot, N., & Ponpinij, P. (2025). Healthcare Chatbot for slowing the Progression in Chronic Kidney Disease Stage 3 patients. 2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC),
DOI: 10.1109/ICAIIC64266.2025.10920642

[5] Dhuriya, G., Yadav, N., Saini, A., & Afliuflama. (2025). Chronic Kidney Disease: Data Clustering and Prediction Using Various Machine Learning Algorithms. 2025 3rd International Conference on Disruptive Technologies (ICDT), DOI: 10.1109/ICDT63985.2025.10986563

[6] Rani, R., Gill, K. S., Upadhyay, D., & Devliyal, S. (2024). XGBoost-Driven Insights: Enhancing Chronic Kidney Disease Detection. 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), DOI: 10.1109/ICOSEC61587.2024.10722440

[7] Chen, N.-J., Chang, C.-H., Lai, C.-L., Liu, K.-Y., & Guo, J.-L. (2024). Effectiveness of an Interactive Robot-Assisted Avatar Health Education Program on Knowledge and Health Literacy in Chronic Kidney Disease Patients. 2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom), DOI: 10.1109/HEALTHCOM60970.2024.10880787

[8] Bhuria, R. (2024). Advanced Machine Learning Techniques for Chronic Kidney Disease Prediction and Management. 2024 International Conference on Artificial Intelligence and Emerging Technology (GlobalAI Summit), DOI: 10.1109/GLOBALAISUMMIT62156.2024.10947932

[9] Sirisha, A. B. N. P., Dhana Lakshmi, N., & Siva Priyanka, S. (2024). Chronic Kidney Disease Stage Prediction Using SGD with Elastic Net Regularization: A Streamlit Application. 2024 International Conference on Computing, Sciences and Communications (ICCSC), DOI: 10.1109/ICCSC62048.2024.10830323

[10] Sonone, N. (2024). Early prediction and progression of chronic kidney disease using machine learning techniques. 2024 2nd International Conference on Networking and Communications (ICNWC), 1 DOI:

10.1109/ICNWC60771.2024.10537571

[11] Arora, A., Sehgal, C., & Agarwal, N. (2024). An Analysis of Machine Learning Algorithms for Chronic Kidney Disease Prediction. 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence), DOI: 10.1109/CONFLUENCE60223.2024.10463505

[12] Sneha, A. (2024). Development of Supervised Algorithm for Predicting Chronic Kidney Disease. 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), DOI: 10.1109/IC2PCT60090.2024.10486505

[13] Sharma, T., & Verma, N. (2024). Analysis for Predicting Chronic Kidney Disease with the Application of Machine Learning Models. 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), DOI: 10.1109/ICDICI62993.2024.10810875

[14] Makaraju, K. L., & Mande, N. V. S. (2023). Machine Learning Predictive Model for Chronic Kidney Disease Classification Using Python. 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI),
DOI:10.1109/ICAIIHI57871.2023.10489064

[15] Anurag, Vyas, N., & Balla, D. (2023). Chronic Kidney Disease Prediction Using Robust Approach in Machine Learning. 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT), DOI: 10.1109/CISCT57197.2023.10351277

[16] Kashyap, C. P., Reddy, G. S. D., & M, B. (2022). Prediction of chronic disease in kidneys using machine learning classifiers. 2022 1st International Conference on Computational Science and Technology (ICCST), 2 DOI:10.1109/ICCST55948.2022.10040329

Copyright & License:
© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.