Predictive Modelling of Kidney Failure using Advanced Machine Learning Techniques: A Comprehensive Review

Aditya P. Pethkar[1], Prof. Malvika Saraf[2]

[1]Research Scholar, Electronics & Telecommunication Engineering, Wainganga College of Engineering and Management, Nagpur, India

[2,]Assistant Professor, Electronics & Telecommunication Engineering, Wainganga College of Engineering and Management, Nagpur, India

How to Cite this Article:

Pethkar, A. P. (2026). Predictive Modelling of Kidney Failure using Advanced Machine Learning Techniques: A Comprehensive Review. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(02). https://doi.org/10.55041/ijcope.v2i2.005

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.

https://doi.org/10.55041/ijcope.v2i2.005

Abstract

Chronic Kidney Disease (CKD) is a growing global health concern characterized by progressive loss of kidney function and late-stage diagnosis due to its asymptomatic early progression. Recent advancements in artificial intelligence (AI) and machine learning (ML) have introduced data-driven approaches capable of improving early detection, risk prediction, and disease management. This review paper examines current research trends in CKD prediction using advanced machine learning techniques, including traditional algorithms, ensemble models, deep learning, and multimodal data integration. The study evaluates data preprocessing strategies, feature selection methods, and model evaluation techniques used in existing literature. Findings indicate that ensemble methods such as Random Forest and Gradient Boosting consistently achieve high predictive accuracy, while deep learning approaches show promise in imaging-based diagnosis. However, challenges such as limited dataset diversity, lack of external validation, model interpretability issues, and ethical concerns remain barriers to clinical adoption. The review highlights the importance of explainable AI, standardized datasets, and multimodal integration to enhance reliability. Overall, AI-driven CKD prediction systems have strong potential to support clinical decision-making and enable early intervention strategies.

Keywords: Chronic Kidney Disease, Machine Learning, Early Diagnosis, Predictive Modeling, Healthcare Analytics etc.

1. Introduction

Chronic Kidney Disease (CKD) is a progressive medical condition marked by the gradual loss of kidney function, ultimately leading to kidney failure if not detected and managed early. It has emerged as a significant global public health challenge due to its high prevalence, economic burden, and association with comorbid conditions such as diabetes, hypertension, and cardiovascular disease. One of the most concerning aspects of CKD is its asymptomatic progression during the early stages, resulting in delayed diagnosis and treatment. Studies indicate that nearly 90% of individuals with early-stage CKD remain unaware of their condition, increasing the risk of complications and mortality [1].

The kidneys play a crucial role in maintaining homeostasis by filtering waste products, regulating fluid balance, and controlling electrolyte levels. When kidney function deteriorates, harmful toxins accumulate in the body, causing systemic complications. CKD progression is categorized into stages based on the estimated Glomerular Filtration

Rate (eGFR), with advanced stages requiring dialysis or kidney transplantation for survival. Early detection and timely intervention can slow disease progression and significantly improve patient outcomes [2].

Traditional CKD diagnosis relies on laboratory tests such as serum creatinine levels, urine protein analysis, blood pressure monitoring, and imaging techniques. While these diagnostic tools are essential, they often fail to detect early abnormalities or subtle patterns hidden within clinical data. Furthermore, clinical decision-making frequently depends on physician expertise and manual interpretation, which can introduce variability and delays in diagnosis

[3].With the rapid digitization of healthcare records and the availability of large-scale clinical datasets, artificial intelligence (AI) and machine learning (ML) techniques have gained prominence in medical diagnostics. Machine learning algorithms can analyze complex and high-dimensional data, identify hidden patterns, and generate predictive models capable of supporting clinical decision-making. These computational approaches provide opportunities for early CKD detection, risk stratification, and disease progression prediction [4].

Recent research demonstrates that traditional machine learning models such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, and Convolutional Neural Network (CNN) have been widely applied to CKD prediction. However, ensemble methods, including Random Forest and Gradient Boosting, often outperform single models due to their robustness and ability to handle nonlinear relationships among variables [5]. Additionally, deep learning techniques have shown promising results in analyzing medical imaging data to detect structural kidney abnormalities [6].

Despite these advancements, several challenges hinder the clinical implementation of AI-based CKD prediction systems. Many studies rely on small or region-specific datasets, limiting generalizability across diverse populations. Lack of external validation and standardized evaluation metrics further raises concerns about real-world reliability. Moreover, complex machine learning models often operate as “black boxes,” making it difficult for clinicians to interpret predictions and trust automated systems [7].

Recent literature emphasizes the importance of explainable AI, multimodal data integration, and ethical governance in building reliable healthcare AI systems. Combining laboratory data, imaging findings, and electronic health records can significantly improve predictive accuracy and enable precision medicine approaches. Additionally, transparent models and regulatory compliance are essential to ensure safe and equitable deployment in clinical settings [8].

This review paper aims to analyze current research trends, methodologies, and challenges in CKD prediction using advanced machine learning techniques. By synthesizing findings from recent studies, the paper highlights technological advancements, identifies research gaps, and proposes future directions for developing reliable, interpretable, and clinically applicable CKD prediction systems.

2. Problem Identification

Chronic Kidney Disease (CKD) represents a major global health challenge due to its silent progression and high rate of late-stage diagnosis. Early stages of CKD often present no noticeable symptoms, resulting in nearly 90% of affected individuals remaining unaware of their condition until significant kidney damage has occurred [1]. Traditional diagnostic approaches rely on laboratory tests and physician interpretation, which may fail to detect subtle abnormalities and early risk patterns hidden within clinical data [2]. Furthermore, healthcare systems in resource-limited regions often lack access to advanced diagnostic technologies, contributing to delayed detection and increased mortality rates [3].

Existing CKD prediction studies using machine learning show promising accuracy; however, many rely on small, region-specific datasets and lack external validation, limiting real-world applicability [4]. In addition, issues such as missing data, class imbalance, and inconsistent preprocessing methods reduce model reliability [5]. Limited model transparency and interpretability further hinder clinician trust and adoption of AI-based diagnostic systems [6]. These challenges highlight the need for reliable, explainable, and data-driven CKD prediction solutions.

3. Literature Survey

A) Literature Review

Francesco Sanmarchi et al., 2023, This systematic review examines machine learning applications in predicting, diagnosing, and managing Chronic Kidney Disease. The authors evaluate a wide range of ML models, highlighting their capability to improve early detection, risk stratification, and treatment decision-making. The study emphasizes that ML significantly enhances diagnostic accuracy compared to traditional clinical assessment. However, it also identifies challenges such as dataset bias, inadequate external validation, and limited deployment in real-world healthcare environments. The review concludes that while ML holds promise for CKD management, large-scale clinical trials and robust validation frameworks are essential for safe and effective clinical integration.

Fizza Khalid et al., 2024, This systematic review investigates machine learning and artificial intelligence techniques for predicting CKD progression. The authors analyze various models, noting that ML algorithms demonstrate strong discriminatory capability in identifying disease progression patterns. However, they highlight major limitations, including small sample sizes, single-center datasets, inconsistent evaluation metrics, and heterogeneity in outcome definitions. These issues hinder the generalizability of results across diverse populations. The study stresses the need for standardized reporting, large-scale multicenter cohorts, and improved methodological quality. Overall, the findings indicate promising ML potential but call for stronger, more consistent study designs for reliable clinical adoption.

I. I. Iliyas et al., 2025, This literature review explores traditional machine learning, deep learning, and hybrid models used for CKD prediction. The authors find that ensemble algorithms—particularly Random Forest and Gradient Boosting—outperform conventional models due to their robustness and feature-handling abilities. Deep learning approaches also show potential but require larger datasets for optimal performance. The study emphasizes the need for standardized and high-quality datasets, improved feature selection methods, and enhanced explainability to build trust among clinicians. The review concludes that hybrid and ensemble techniques represent the most promising direction but must be supported by transparent and interpretable model architectures.

Charumathi Sabanayagam et al., 2025, This scoping review maps the broad applications of artificial intelligence in CKD management, covering early disease detection, risk stratification, patient communication, and treatment optimization. The authors highlight that AI tools can significantly reduce diagnostic delays and enhance patient monitoring. However, they identify challenges such as poor data quality, inconsistent integration with clinical workflows, lack of interpretability, and regulatory uncertainties. The review stresses that AI systems should be designed with clinician involvement and validated across multiple healthcare settings. It concludes that AI has transformative potential but requires structured implementation strategies to ensure reliability and real-world effectiveness.

Alessia Nicosia et al., 2025, This comprehensive review covers AI applications across nephrology, including CKD, dialysis management, and kidney transplantation. The authors highlight the growing relevance of multimodal data—omics, imaging, and electronic health records—in improving diagnostic accuracy and treatment personalization. They emphasize the importance of machine learning for early risk prediction and patient stratification. However, the study identifies challenges such as data heterogeneity, lack of prospective clinical trials, and limited interpretability of complex models. The authors conclude that while AI offers significant advancements in nephrology, future research must focus on transparent models and robust clinical validation to ensure safe adoption.

Pushkala Jayaraman et al., 2025, This review discusses ML, deep learning, and natural language processing applications in detecting and managing kidney diseases. The authors emphasize the role of multimodal intelligence—combining structured data, medical imaging, and clinical notes—for improved diagnostic accuracy. They caution that issues such as dataset bias, ethical concerns, and infrastructure limitations in low-resource settings hinder AI adoption. The review highlights the need for better data governance, fairness-aware algorithms, and training for healthcare professionals. It concludes that precision nephrology can be achieved through integrated AI systems, provided these challenges are addressed through policy and technological development.

Dan Zhao et al., 2023, This literature review focuses on AI-assisted medical imaging for CKD analysis, involving modalities such as ultrasound, CT, and MRI. The authors show that deep learning and radiomics can detect subtle renal structural changes that are difficult for clinicians to identify manually. These techniques also support prognosis by analyzing tissue textures and pathological patterns. However, the study notes that variations in imaging protocols, inadequate dataset sizes, and lack of multicenter validation restrict clinical deployment. The review concludes that AI imaging is highly promising but requires standardized imaging pipelines and extensive benchmarking.

Navdeep Tangri & Charumathi Sabanayagam, 2025, This commentary discusses AI-based approaches for early CKD detection and risk prediction. The authors highlight that machine learning models using laboratory markers and imaging data can significantly improve early diagnosis accuracy and patient risk stratification. They note that many AI tools are approaching readiness for clinical deployment but require rigorous validation, transparency, and regulatory approval. The study emphasizes the importance of integrating AI with existing nephrology workflows to improve usability. They conclude that AI has strong potential to transform early CKD detection, provided model reliability and clinical interpretability are ensured.

Yudi Kurniawan et al., 2025, This bibliometric and narrative review analyzes research trends in AI for CKD early detection and prognosis from 2022 to 2025. The authors find rapid growth in the use of deep learning models, particularly CNNs and hybrid architectures. The review identifies gaps such as limited multimodal datasets, lack of representation from low-resource regions, and minimal external validation. It emphasizes that integrating imaging, lab values, and clinical notes can significantly improve prediction accuracy. The authors conclude that while AI in CKD research is expanding rapidly, methodological inconsistencies must be addressed for real-world integration.

Rajesh Yadav, 2024, This review examines various machine learning models, including SVM, Random Forest, KNN, Artificial Neural Networks, and ensemble methods for CKD prediction. The author highlights that ensemble models and tree-based algorithms consistently achieve superior accuracy and feature-handling capability. However, the study stresses the importance of preprocessing steps such as missing value treatment, normalization, and feature selection. It also notes the need for balanced datasets to avoid biased predictions. The review concludes that ML can significantly enhance CKD detection, provided preprocessing and model selection are carefully optimized.

B) Literature Summary

Recent studies highlight the growing role of artificial intelligence (AI) and machine learning (ML) in improving the early detection and prediction of Chronic Kidney Disease (CKD). Traditional ML algorithms such as Logistic Regression, Support Vector Machines, Decision Trees, and K-Nearest Neighbors have been widely applied for CKD classification. However, ensemble techniques like Random Forest and Gradient Boosting consistently demonstrate superior accuracy and robustness due to their ability to manage nonlinear relationships and complex feature interactions. Deep learning methods are increasingly used for imaging-based diagnosis, enabling detection of structural kidney abnormalities. Researchers emphasize the importance of data preprocessing, feature selection, and handling missing values to improve prediction performance. Several studies also highlight the benefits of integrating multimodal data, including laboratory results, electronic health records, and medical imaging. Despite promising results, many studies report limitations related to dataset size, lack of external validation, and model interpretability, which affect clinical reliability and real-world implementation.

C) Research Gap

Although machine learning techniques have shown promising results in CKD prediction, several research gaps remain. Many studies rely on small, single-center, or region-specific datasets, limiting the generalizability of predictive models across diverse populations. External validation using independent datasets is often missing, raising concerns about model robustness in real-world clinical environments. Inadequate handling of missing and inconsistent medical data continues to affect model accuracy and reliability. Furthermore, limited integration of multimodal healthcare data—such as laboratory results, imaging, and clinical histories—restricts comprehensive disease prediction. Another significant gap is the lack of explainable AI models, which reduces clinician trust and hinders adoption in healthcare settings. Ethical concerns, including algorithmic bias, fairness, and patient data privacy, are also insufficiently addressed. Additionally, the absence of standardized evaluation protocols and reporting guidelines creates inconsistencies in performance comparison across studies, highlighting the need for more transparent, validated, and clinically interpretable CKD prediction systems.

4. Research Methodology

A) Criteria for selecting this study:

Chronic Kidney Disease (CKD) is a significant public health concern due to its silent progression, high prevalence, and severe complications when diagnosed at advanced stages. The selection of this study is based on the urgent need for early detection methods that can improve patient outcomes and reduce healthcare burdens. Traditional diagnostic approaches rely heavily on laboratory tests and physician expertise, which may not detect early-stage abnormalities or hidden patterns in clinical data. With the increasing availability of electronic health records and medical datasets, machine learning techniques offer a promising data-driven approach for improving diagnostic accuracy and supporting clinical decision-making. This study is selected to explore how advanced machine learning models can enhance early CKD prediction and provide reliable, automated classification. Furthermore, the research aligns with current trends in healthcare analytics and artificial intelligence integration, aiming to develop efficient, accessible, and scalable diagnostic solutions.

- High global prevalence and silent progression of CKD.
- Need for early diagnosis to prevent kidney failure and complications.
- Limitations of traditional diagnostic methods.
- Availability of clinical datasets enabling predictive modeling.
- Potential of machine learning to improve diagnostic accuracy.
- Growing demand for automated and cost-effective healthcare solutions.
- Relevance to digital healthcare transformation and decision support systems.

B) Method of analysis:

The method of analysis involves systematic processing and evaluation of clinical data to develop an accurate machine learning model for CKD prediction. Initially, the dataset is examined to identify missing values, inconsistencies, and outliers that may affect prediction accuracy. Data preprocessing techniques such as imputation, normalization, and encoding are applied to ensure data quality and consistency. Exploratory Data Analysis (EDA) is performed to understand feature distributions, detect correlations, and identify significant variables influencing CKD. The dataset is then divided into training and testing subsets to ensure unbiased model evaluation. Multiple machine learning algorithms are implemented and compared to determine the most effective predictive model. Performance is evaluated using statistical metrics to ensure reliability and robustness. Finally, results are interpreted to assess clinical relevance and model effectiveness in supporting early diagnosis.

- Data cleaning and missing value treatment.
- Feature scaling and normalization.
- Exploratory Data Analysis and correlation study.
- Dataset splitting (training and testing).
- Model training using multiple ML algorithms.
- Performance evaluation using accuracy, precision, recall, and F1-score.
- Interpretation of results for clinical applicability.

E) *Highlighting trends, advancements, and challenges*

Trends:

- Increasing use of machine learning for early CKD detection and risk prediction.
- Growing adoption of ensemble learning models for improved accuracy.
- Integration of electronic health records and clinical datasets for predictive analytics.
- Rising interest in multimodal data combining lab results, imaging, and patient history.
- Expansion of AI-driven decision support systems in healthcare.
- Growing emphasis on preventive healthcare through predictive modeling.
- Increased research focus on personalized medicine using AI techniques.

Advancements:

- Development of high-accuracy ensemble algorithms such as Random Forest and Gradient Boosting.
- Use of deep learning for kidney imaging analysis and disease detection.
- Improved data preprocessing techniques for handling missing and inconsistent medical data.
- Advanced feature selection methods enhancing model performance.
- Cloud-based healthcare analytics enabling scalable AI deployment.
- Real-time predictive systems supporting early diagnosis and monitoring.
- Progress in explainable AI improving model transparency and clinician trust.

Challenges:

- Limited availability of large, diverse, and high-quality medical datasets.
- Lack of external validation affecting model reliability.
- Black-box nature of complex AI models reducing interpretability.
- Ethical concerns including bias, fairness, and patient data privacy.
- Integration difficulties with existing clinical workflows.
- Resource constraints in rural and low-income healthcare settings.
- Absence of standardized evaluation protocols and regulatory guidelines.

5. Discussion

A) Methodology for future research directions

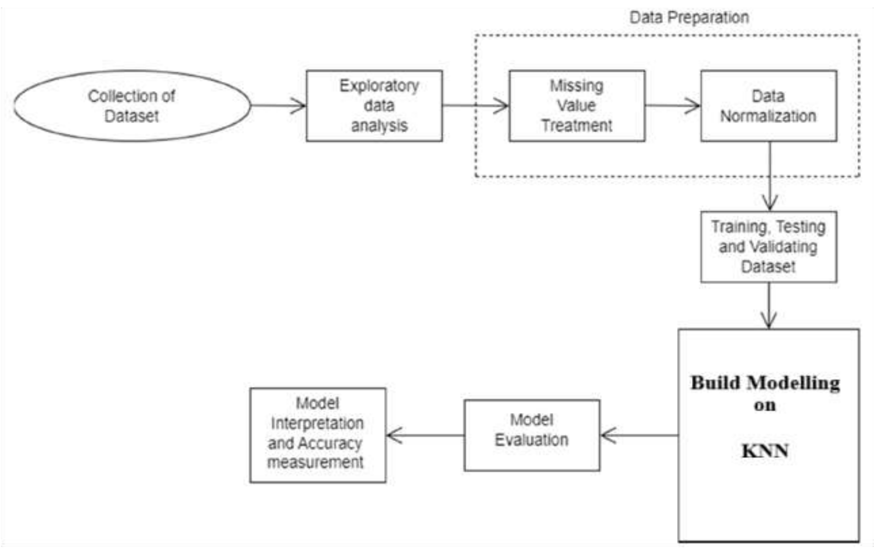

Figure 2. Proposed methodology of Chronic Kidney Disease

- The system begins with the collection of a CKD dataset, which includes clinical, demographic, and laboratory parameters essential for kidney disease prediction.
- Exploratory Data Analysis (EDA) is performed to understand data distribution, detect outliers, examine feature correlations, and identify missing or inconsistent values.
- In the data preparation phase, missing values are treated using suitable imputation techniques to ensure data completeness and prevent biased model learning.
- Data normalization is applied to scale numerical attributes, ensuring all features contribute equally to distance-based algorithms like Convolutional Neural Network (CNN).
- The cleaned and normalized dataset is then split into training, testing, and validation subsets to enable unbiased model training and performance assessment.
- Using the prepared dataset, a Convolutional Neural Network (CNN).model is built. The algorithm predicts CKD status by computing distances between data points and identifying the closest neighbors.
- During training, the value of 'K' is optimized to achieve the best classification accuracy while minimizing error.
- The trained model undergoes model evaluation, where accuracy, precision, recall, F1-score, and confusion matrix are calculated to measure performance.
- Based on evaluation results, model interpretation is performed to understand how input features influence CKD prediction and to validate the reliability of predictions.
- Finally, the system identifies the most accurate configuration of CNN, enabling effective automated classification of CKD and supporting early diagnosis.

Main Features:

- Utilizes a comprehensive CKD dataset containing clinical, demographic, and laboratory parameters for accurate disease prediction.
- Incorporates advanced Exploratory Data Analysis to identify patterns, correlations, outliers, and data inconsistencies.
- Includes robust missing value treatment methods to enhance data completeness and reliability.
- Applies data normalization techniques to standardize feature scales, especially for distance-based algorithms like CNN.
- Employs systematic dataset splitting for training, testing, and validation to ensure unbiased model performance.
- Implements Convolutional Neural Network (CNN) algorithm for classification to predict CKD or non-CKD cases effectively.
- Provides model evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix.
- Supports model interpretation for understanding feature influence and ensuring transparent decision-making.
- Enhances clinical decision support through automated, data-driven CKD risk prediction.

B) Analysis of data

- The dataset is examined for completeness, identifying missing values, inconsistent entries, and duplicate records that may affect model performance.
- Summary statistics such as mean, median, standard deviation, and range are calculated to understand the distribution of clinical parameters.
- Outliers in features like serum creatinine, blood pressure, and potassium levels are detected to assess abnormal variations related to CKD severity.
- Correlation analysis is conducted to identify strong relationships between variables such as albumin, hemoglobin, eGFR indicators, and CKD risk.
- Categorical and numerical features are visualized using histograms, boxplots, heatmaps, and pair plots to observe trends and variability.
- Class distribution is checked to ensure balance in CKD vs. non-CKD cases, as imbalance may bias machine learning results.

- Feature importance is analyzed to highlight influential factors contributing to disease prediction.
- Data preprocessing decisions are guided by analysis insights, ensuring improved accuracy, reliability, and robustness of the final prediction model.

6. Conclusion

Chronic Kidney Disease (CKD) continues to pose a major global health challenge due to its silent progression, late diagnosis, and strong association with life-threatening complications. This review highlights the growing importance of artificial intelligence and machine learning techniques in improving early detection, risk prediction, and clinical decision support for CKD management. The analysis of recent studies reveals that machine learning models, particularly ensemble approaches, provide superior predictive accuracy and robustness compared to traditional diagnostic methods. Deep learning techniques also demonstrate promising capabilities in imaging-based detection and disease monitoring.

Despite these advancements, several barriers hinder widespread clinical adoption. Limited dataset diversity, lack of external validation, model interpretability issues, and ethical concerns such as bias and data privacy remain significant challenges. Furthermore, integration of AI systems into existing clinical workflows requires regulatory support and clinician training.

Future research should focus on explainable AI, multimodal data integration, standardized evaluation frameworks, and large multicenter datasets to enhance reliability and trust. Overall, AI-driven CKD prediction systems hold substantial potential to support early intervention, improve patient outcomes, and advance preventive healthcare practices.

References

[1] F. Sanmarchi, et al., “Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review,” Journal of Nephrology, 2023.

[2] F. Khalid, et al., “Predicting the Progression of Chronic Kidney Disease: A Systematic Review of Artificial Intelligence and Machine Learning Approaches,” Cureus, 2024.

[3] I. I. Iliyas, et al., “Recent trends in prediction of chronic kidney disease using different learning approaches: a systematic literature review,” Journal of Medical Artificial Intelligence, 2025.

[4] C. Sabanayagam, et al., “Artificial intelligence in chronic kidney disease management: a scoping review,” Theranostics, 2025.

[5] A. Nicosia, et al., “Artificial Intelligence in Nephrology: From Early Detection to Clinical Management of Kidney Diseases,” Bioengineering, 2025.

[6] P. Jayaraman, et al., “Artificial Intelligence in Nephrology: Pioneering Precision with Multimodal Intelligence,” Indian Journal of Nephrology, 2025.

[7] D. Zhao, et al., “Current progress in artificial intelligence-assisted medical image analysis for chronic kidney disease: A literature review,” Computational and Structural Biotechnology Journal, 2023.

[8] N. Tangri and C. Sabanayagam, “Artificial intelligence approaches to enable early detection of CKD,” Nature Reviews Nephrology, 2025.

[9] Y. Kurniawan, et al., “Artificial Intelligence for Chronic Kidney Disease Early Detection and Prognosis,” medRxiv, 2025.

[10] R. Yadav, “Literature Review and Discussion of Machine Learning Algorithms for Predicting Chronic Kidney Disease,” Journal of Computer Technology & Applications, 2024.