iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: X Month of publication: October 2025

DOI: https://doi.org/10.22214/ijraset.2025.73769

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

Chronic Kidney Disease Early Detection using Machine Learning: Review and Critical Insights

Varalakshmi S S[1], Shushma R[2], Kusuma K[3], Mohamadi Ghousiya Kousar[4]
[1]AIDS, Vijaya Vittala Institute of Technology, Kannur Visveswarayya Technological University, Belagavi, Karnataka
[2, 3, 4]CS & E, Vijaya Vittala Institute of Technology, Kannur Visveswarayya Technological University, Belgavi, Karnataka

Abstract: Chronic Kidney Disease (CKD) is a significant global health concern reported to be affecting 14% of the world population, which is a serious burden to the country since it has no symptoms, and the worsening of kidney functioning occurs faster over time. The delay or prevention of development to end-stage renal disease depends on early detection and accuracy, which improves the survival and lives of patients. This review critically assesses the recent use of machine learning in the early diagnosis of CKD on the basis of their diagnostic performance, consistency, and whether they can be placed in clinical practice. The discussion of studies that have utilized algorithms such as Random Forest, XGBoost, Support Vector Machines, Convolutional Neural Networks(CNN), and hybrid deep learning models demonstrates the superiority of these algorithms over their conventional diagnostic counterparts. Conclusions of the reviewed literature demonstrate that the principles of a machine learning approach make a better result in terms of classification accuracy and predictive certainty than traditional methods can offer. When feature selection is optimized, Random Forest models achieved 100% accuracy, XGBoost showed 94-95%, hybrid CNN-SVM showed 96.18% and advanced GRU-BiLSTM architectures showed 96.5%. Incorporation of explainable AI techniques such as SHAP analysis and an interpretable hybrid framework added to the clinical trust factor by having clear and explainable information with respect to model predictions. The results highlight the paradigm-shift possibilities of machine learning to enhance the diagnosis of CKD and mainstream it into daily clinical routine.
Keywords: CKD, Random Forest, CNN, XGBoost, SVM.

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive chronic disease that is caused by a progressive deterioration of kidney function, commonly resulting in end-stage renal disease when not treated at the earliest stage. It is a major international health issue with a prevalence of about 10-15 % of the adult population and has been associated with both morbidity and mortality as well as the spending of healthcare. Comorbidities like diabetes, hypertension, and cardiovascular disease are also very strongly linked to CKD and boost its progression, and make it difficult to treat. The disease burden is also compounded by the fact that it is usually asymptomatic leading to late diagnosis in the advanced stages where treatment becomes farfetched and the prognosis worse. In addition to its direct clinical consequences, CKD is associated with significant and economically relevant social costs and healthcare resources burdens, decimating the quality of life of patients.

A. Clinical Background

CKD is a slowly developing chronic disease and refers to the situation when the kidney function gets diminished gradually. The disease is stipulated by occurrence of damage or reduction of functions of the kidneys over three months with health consequences. CKD is one of the most critical global health problems, and it is still on the increase in different parts of the globe because of aging population, and higher levels of diabetes and hypertension. Clinical presentation CKD varies a lot as per the development stage of the disease. Early cases may produce no symptoms, so that the patient cannot be detected unless they undergo systematic screening. With each stage of the disease, it is possible to progress to such symptoms as fatigue, swelling, urination changes, and cardiovascular problems. The asymptomatic stage of CKD makes the process of developing effective screening methods and early detection strategies extremely important.

B. Epidemiology

The existing epidemiological statistics show that CKD has an impact on about 14% of the world population, and this number varies greatly depending on the geographical area and demographics. It is especially common with people who have diabetes mellitus and hypertension, which is the major cause of CKD all over the world.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

In the US, more than 37 million adults have CKD, and in many cases, they do not even assume that they have this disease. The costs of CKD are very high, with health expenditure in the United States alone exceeding $120 billion per year. Treating the diseases becomes extremely expensive with the advance of the disease since end stage renal disease (ESRD) entailing dialysis or transplantation of the kidney is the most costly type of treatment. This economic observation underlines the possible usefulness of early detection and intervention schemes.

C. Significance of Early Detection

Early diagnosis of CKD is most important because of the following reasons. To start with, it will allow implementing measures that may slow the disease process, such as blood pressure regulation, treatment of diabetes, and lifestyle changes. Second, at an early stage, it is possible to avoid cardiovascular complications that are the main cause of death in patients with CKD. Third, early identification helps in proper preparation to undertake renal replacement therapy, which helps to improve the condition of patients who reach end-stage renal disease. The current practice of diagnosing is centered mainly on the use of serum creatinine and calculation of estimated glomerular filtration rate (eGFR). Such markers, however, have their serious limitations, the fact that they do not detect other early kidney damage as well as they are vulnerable to many confounders, like muscle mass, age and race. With the advancement of technology, machine learning methods have the possibility of eliminating these limitations because they can analyze the intricate trends in several biomarkers and clinical variables.

CKD is a growing burden on global health, the disease has silent manifestations in the asymptomatic early stages thus complicating early detection and subsequent incurring of expensive treatments. Diagnosis as early as possible is needed to reduce progression, complications, and help the patients. Taking into consideration the drawbacks of the traditional markers, machine learning holds a future prospect of early diagnosis of CKD in a more precise and personalized manner.

The rest of the paper is organized as follows. section II covers pathophysiology and clinical process in CKD, which involves the nature of biological events and stages of the disease. Section III shows new biomarkers, which can be used in early identification of CKD, and their benefits compared to the traditional indicators. In Section IV, related work is reviewed, and the main limitations and gaps in research are identified. In Section V, both current and potential biomarkers on issues relating to early diagnosis and disease management are addressed. In section VI, multiple machine learning algorithms used to predict CKD are benchmarked and their performance, advantages, and disadvantages compared. In section VII, the issues regarding early detection and implementation of these strategies are discussed, and new areas of research are offered. Finally, in section VIII highlights the important findings of the various perspectives. of combining advanced biomarkers and machine learning in order to better detect CKD and manage it.

II. UNDERSTANDING THE PATHOPHYSIOLOGY AND CLINICAL PROGRESSION OF CKD

Firstly, CKD refers to the inexorable and reversible upside-down activity in the kidney and can significantly affect the health of multiple individuals around the world. It is associated with high morbidity, mortality as well as the burden on healthcare costs. CKD has a multi-factorial pathophysiology that is a complex process of cascading hemodynamic, metabolic, inflammatory and fibrotic processes degrading nephron structure and reducing their functionality. Compensatory hyperfiltration of surviving nephrons is accompanied by the early injurious process of kidney damage, and eventually results in glomerular sclerosis, tubulointerstitial fibrosis and renal irreversible mass loss. Several causes such as hypertension, diabetes mellitus, dyslipidemia and oxidative stress make an important role to fasten the process of the diseases. Learning the underlying mechanisms and the stepwise process of CKD development to ESRD will help make an earlier diagnosis and save a patient by intervening and providing better outcomes through effective treatments.

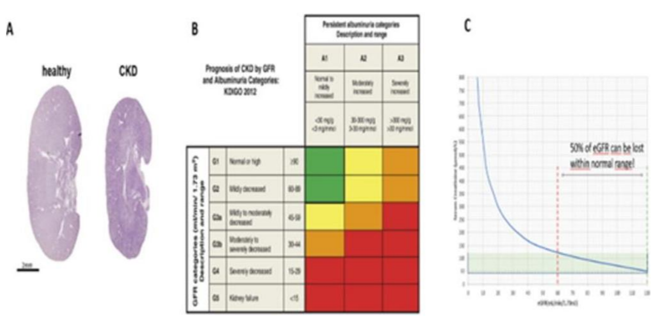

Figure 1 : Morphology to Metrics of CKD Progression and Diagnosis

The image is composed of three frames (A, B, C) which simulate structural, prognostic, and functional views of CKD:

Panel A: Comparison of histologies illustrates sections of kidney tissue of a healthy kidney and a kidney with CKD. The normal kidney is morphologically intact, showing the presence of clear cortex and medulla. The CKD kidney is associated with the development of evident tissue damage, fibrosis, and shrinkage, which is a sign of gradual nephron loss and structural degeneration because of progressive loss of CKD.

Panel B: CKD Prognosis by GFR and Albuminuria (KDIGO 2012) Panel B The diagnoses of CKD in this table are categorized using GFR (G1-G5) and albuminuria (A1-A3) categories: G1 (>90ml/min/$1.73m^2$): Normal GFR / high GFR. Decreased by moderate levels of 60 to 89: G2. G3a: 45 59, G3b: 30 44: Mild to moderate decrease and moderate to severe decrease. G4<sub (15-29)</sub>: Markedly ablated. G5 (<15):Kidney failure. PAlbuminuria levels: A1 (<30 mg/g): Normal to modest elevated. A2(30-300mg /g) Modestly elevated. A3 (>300 mg/g): Extremely exaggerated. Risk stratification is presented by the color-coded grid with green = low risk, yellow = moderate risk, orange = high risk, and red = very high risk of CKD progression.

Panel C: The Relationship with eGFR and serum Creatinine It was plotted as that of serum creatinine (mol/L versus estimated GFR (eGFR, mL/min/1.73 m 2). It shows that serum creatinine increases exponentially as eGFR reduces implying that loss in kidney function becomes overtly clinical only after severe destruction of the kidney nephrons. One of the most crucial points was outlined: 50 percent eGFR may be lost and remain in the normal range, which shows that serum creatinine is an insensitive early bias of CKD.

A. Disease Mechanisms

The pathophysiology of CKD is multi-factorial, and the mechanisms are interrelated, resulting in the gradual loss of nephrons and their functional deterioration. The initial affront, be it diabetes, hypertension, glomerulonephritis, or any other cause, leads to another chain of experiences such as glomerular hyperfiltration, proteinuria, inflammation, and fibrosis. These actions generate a self-sustaining loop of kidney harm, unnecessary to cease when the preconcerted source becomes solved. Within the cells, the pathophysiology of CKD relates to oxidative stress, liberation of inflammatory mediators and activation of the renin- renin-angiotensin-aldosterone system. The mechanisms lead to glomerulosclerosis or tubular atrophy and interstitial fibrosis, which eventually result in loss of kidney function that cannot be reversed. Such knowledge about the pathophysiological processes is essential in the design of specific interventions and selection of pertinent biomarkers to the use of machine learning models.

B. Staging and Progression

The stages of kidney disease are divided into five levels and assigned depending on the rate of estimated glomerular filtration rate (eGFR) and the presence of signs of kidney damage. Stages 1(eGFR to 90mL/min/1.73m 2) and 2 (eGFR 60-89 mL/min/1.73m 2) involve early stages of the disease and the presence of kidney damage despite having intact or mildly impaired kidney ability to filter wastes and excess fluid. Stage 3 gets further divided into 3a (eGFR 45-59 mL/min/1.73m 2) and 3b (eGFR 30-44 mL/min/1.73m 2) with a moderate decline in the functioning of the kidney. Stages 4 (eGFR 15-29 mL/min/1.73m 2) and 5 (eGFR <15 mL/min/ 1.73 m 2) occur in case of severe reduction and renal failure, respectively. The progression rate differs widely across individuals and is influenced by a number of factors, such as the cause, the occurrence of proteinuria, the levels of blood pressure, and any genetic considerations. The analysis of these multiple variables can potentially achieve an outcome where the progression rates can be predicted through the use of machine learning models and given as assessments specific to individual patients.

III. NOVEL BIOMARKERS IN THE EARLY IDENTIFICATION OF CKD

This section discusses about how CKD can be prevented by timely diagnosis and better outcomes, but existing conventional measures, such as serum creatinine, eGFR, and proteinuria, fail to identify the early-stage damage. The developments in molecular biology allow the appearance of new biomarkers that are more sensitive and specific and can capture glomerular injury, signaling tubular damage, detecting inflammation, and fibrosis even earlier, when major impairment has not taken place yet. In this area, these novel biomarkers are mentioned and can be used as an opportunity to make a prompt diagnosis and provide personalized intervention.

A. Traditional Biomarkers

The small sensitivity of the traditional markers, such as serum creatinine and eGFR, usually causes (CKD) to go unnoticed at its initial stages. The timely interventions are also important and can be done by the detection at an early period to slow down the progression of the disease. New biomarkers KIM-1, NGAL, cystatin C and microRNAs are more specific and sensitive to locate early renal injury. Although CKD has been diagnosed and managed over the years, the integration of these biomarkers in clinical practice may revolutionize the determination and management of CKD.

B. Novel Biomarkers

Recent studies have identified a number of new biomarkers that are likely to enhance the detection of kidney health at early stages. These are neutrophil gelatinase associated lipocalyin (NGAL) kidney injury molecule-1 (KIM-1) cystatin c and other inflammatory indicators. A machine learning analysis of these biomarkers has produced higher diagnostic accuracy than in conventional methods. The top important features highlighted by HAP analysis in consistently finding CKD status are hemoglobin, serum creatinine, specific gravity, albumin and diabetes status. The results give congruence with clinical knowledge but also quantification of feature importance.

TABLE 1: KEY BIOMARKERS IDENTIFIED BY MACHINE LEARNING ANALYSIS

Biomarker	Clinical Significance	SHAP Importance Ranking	Threshold Values
Serum Creatinine	Kidney function marker	1-2	>1.15 mg/dL (risk indicator)
Hemoglobin	Anemia indicator	1-3	<13.05 g/dL (risk indicator)
Specific Gravity	Urine concentration ability	2-4	Abnormal values indicate dysfunction
Albumin	Protein leak indicator	3-5	>0.5 (significant proteinuria)
Diabetes Status	Primary risk factor	1-2	Present/Absent
Blood Pressure	Cardiovascular risk	4-6	>140/90 mmHg

IV. RELATED WORK AND RESEARCH GAP

This section discusses CKD as a significant global health burden due to its progressive nature and late-stage diagnosis in many patients. Early detection and timely intervention are crucial for enhancing patient outcomes and minimizing associated healthcare costs. With the rapid progression in artificial intelligence (AI), machine learning (ML), and deep learning (DL), numerous studies have explored the potential of these technologies to enhance CKD prediction, diagnosis, and management. This survey aims to consolidate recent research efforts in applying AI-based methods to CKD detection, highlighting the key methodologies, tools, and findings. By reviewing and comparing diverse approaches, this section provides a comprehensive understanding of current trends, practical implementations, and areas requiring further research in AI-driven CKD analysis**.**

The study in [1] identified a number of new biomarkers that are likely to enhance the detection of kidney health at early stages. These are neutrophil gelatinase-associated lipocalin (NGAL), kidney injury molecule-1 (KIM-1), cystatin c, and other inflammatory indicators. A machine learning analysis of these biomarkers has produced higher diagnostic accuracy than conventional methods. The top important features highlighted by HAP analysis in consistently finding CKD status are haemoglobin, serum creatinine, specific gravity, albumin, and diabetes status. The results align with clinical knowledge and also provide quantification of feature importance. The work in [2] uses machine learning methods by consistently delivering high precision in detecting chronic CKD. Research documents accuracy values of 98-100 percent with the application of algorithms like the Random Forest, SVM, CNN, and the ensemble models. Combined classifiers and feature selection techniques also provide additional benefits in optimizing the performance of the prediction. These results highlight the high possibilities using machine learning in early detection of CKD, its risk assessment, and useful management to ensure timely intervention and better patient outcomes.

The literature review in [3] emphasizes high performance of the advanced machine learning algorithms in the context of the CKD diagnosis and prognosis as compared to traditional statistics-based algorithms. It also stresses the importance of model choice, feature selection, and model parameter tuning to get good predictive accuracy. Such algorithms as Random Forest and logistic regression algorithms are presented as being rather effective when predicting CKD. The review also highlights both the transformational potential of machine learning in providing the abilities to diagnose and treat early as well as the need to validate the machine learning models continuously to ascertain their clinical applicability.

The review in [4] highlights the global burden of CKD and the shortcomings of traditional detection methods. It explores advancements in AI and machine learning for CKD diagnosis and monitoring, with numerous studies demonstrating the effectiveness of predictive models using clinical and demographic data. The review underscores the value of integrating multimodal datasets to enhance predictive accuracy and identifies existing gaps in real-time data integration, model interpretability, and ethical considerations. Evidence suggests that AI-driven approaches can improve CKD diagnosis accuracy by 12%–15% over conventional techniques.

The study in [5] draws on comprehensive searches of both internal and external references to ensure thorough coverage. It highlights the rapid growth in CKD research and emphasizes effective ML strategies for CKD diagnosis. Various classification methods are examined for their effectiveness across different disease stages, providing insights into their clinical applicability. The review offers a holistic analysis of ML applications in CKD, addressing notable gaps in the existing literature. Additionally, it proposes two frameworks—one for categorizing learning algorithms and another for mapping medical subfields related to ML in CKD—to guide future research and practical implementations.

Subsequently, the review in [6] underscores AI's transformative potential in healthcare, particularly for improving chronic disease management. It highlights successful applications in diabetes, cardiovascular disorders, and hypertension, demonstrating the effectiveness of predictive analytics. The review stresses the need to integrate AI into routine clinical practice while addressing challenges such as data quality, interoperability, and ethical concerns. It advocates for developing robust AI models that can be seamlessly incorporated into existing healthcare systems to enhance patient outcomes.

The paper in [7] addresses challenges in early kidney cancer detection and the limitations of conventional clinical methods, highlighting the need for automated diagnostic tools to improve patient survival. It emphasizes the role of accurate classification methods and quality datasets in enhancing system performance, introducing an Adaptive Hybridized Deep Convolutional Neural Network (AHDCNN) for efficient kidney disease identification. Factors such as hypertension and specific diagnostic tests are noted as significant in CKD detection. The Random Forest model demonstrated superior predictive accuracy and sensitivity, supporting the application of data mining algorithms for improved chronic disease detection and patient outcomes.

The paper in [8] reviews AI-driven mobile applications for early chronic disease detection and management, evaluating the effectiveness of machine learning techniques in healthcare. It examines the integration of AI technologies to enhance health monitoring and diagnosis, outlining both the benefits and challenges of their medical application. The review underscores AI's potential to transform chronic disease treatment and improve overall health outcomes.

The review in [9] evaluates various machine learning and deep learning methods for CKD prediction, identifying shortcomings in past studies and areas for further research. Reported results include 94.6% accuracy with Naïve Bayes, high predictive feature selection using Cramer's V and ANOVA, 99.1% F-measure with Gradient Boosting, 97.12% accuracy with Random Forest after feature extraction, and 99.16% accuracy with PCA-applied models. The review stresses the need for diverse datasets to improve model generalizability and proposes a hybrid deep learning approach to enhance CKD diagnosis beyond traditional methods.

Subsequently, the review in [10] discusses that CKD affects over 800 million people worldwide, with the greatest burden in low- and middle-income countries due to limited treatment options. Machine learning (ML) offers potential for early detection by uncovering patterns in clinical data, while Explainable Artificial Intelligence (XAI) improves model interpretability, fostering clinical trust. However, XAI applications in CKD prediction remain limited, often focusing only on basic feature selection. Literature highlights that both feature selection and hyperparameter optimization are key to enhancing model performance, with various ML algorithms benefiting from these techniques to improve diagnostic accuracy.

A review in [11] addresses CKD is a major global health concern, often linked to diabetes mellitus and associated with severe complications. Early detection is critical, yet conventional methods frequently fall short due to the disease's gradual onset. Machine Learning (ML), supported by Explainable AI (XAI), enhances detection accuracy while ensuring model interpretability. Models such as Logistic Regression and Random Forest have demonstrated predictive effectiveness, with hybrid AI approaches showing promise for improving early-stage CKD identification and enabling personalized treatment recommendations.

The review in [12] examines various CKD detection methods, noting their strengths and limitations. Different strategies of machine learning with use towards the early detection of CKD indicate that boosting algorithms, explainable AI, and deep learning are effective. Random Forest, ANN, and optimized CNN models reached a high level of accuracy, which reminds the significance of the importance of preprocessing and feature selection. The results also indicate the importance of an external validation requirement to generalize the models.

The review in [13] highlights AI techniques namely Support Vector Machines (SVM) and Artificial Neural Networks (ANN) for predicting CKD with high accuracy. It emphasizes the role of feature selection, ensemble learning, and clustering algorithms like k-means for patient grouping. Data preprocessing and feature engineering are identified as essential for robust predictive models, while interpretability tools like SHAP and LIME enhance clinical trust. The integration of genetic data and NLP methods is also explored for improved CKD risk assessment. The review concludes that AI-driven approaches hold significant potential for advancing CKD diagnosis and management.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

The paper in [14] highlights CKD affects 17% of India's population and is often diagnosed at advanced stages, resulting in high treatment costs. AI and machine learning offer opportunities for early detection and progression prediction, improving patient outcomes. However, previous research has faced challenges with inconsistent feature selection and reliance on black-box models, limiting generalizability. This study emphasizes the need for interpretable models, employing Decision Trees and Explainable AI tools like LIME for transparency. The literature further underscores integrating AI into healthcare to reduce CKD-related costs and enhance clinical decision-making.

The researchers in [15] integrated neural networks with robotic process automation (RPA), achieving 98.3% training accuracy and reducing waste by 20.4%. the work emphasizes the benefits of AI–RPA integration for productivity and cost efficiency across sectors. Also proposed a cloud-integrated RPA framework for enhancing social robots, attaining 97.3% accuracy in behavioral recognition. The literature notes challenges in healthcare data management and proposes analytical techniques for better patient surveillance. Deep learning algorithms in biomedical predictive systems are highlighted for their transformative potential, with cognitive systems identified as key to optimizing operations and improving healthcare quality.

The review section summarizes AI, ML, and DL technologies have great prospects in early detection of CKD, with 94-100 % detection accuracy with models such as Random Forests, SVMs, and combinations of ANN modules. Biomarkers and important clinical features to know can enhance prediction, and explain AI helps to promote transparency. In spite of the advancements, data quality, generalizability, and incorporation into clinical workflows are the key challenges.

A. Research gap

The section represents the most common research gaps in terms of the application of AI to CKD detection that has been mentioned in the course of the review of the research. The data quality and availability become the most serious challenge and model interpretability and rigorous validation follow suit. There are also concerns of generalizability, integration into clinical practice, and real-time implementation restrictions. These gaps need to be addressed so that AI models can be realized as useful diagnostic tools of CKD that can be implemented in clinical practice. The chart in figure 1 illustrates the most typical research gaps in AI-based CKD detection, where the y-axis contains the lists of various categories of gaps and the x-axis contains the number of articles pointing out each gap.

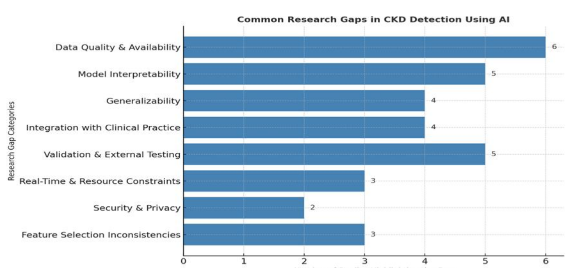

Figure 1:Barchart representing research gap in CKD detection using AI

The table bar chart above illustrates the frequency of typical gaps in research found in recent articles on the study of CKD detection using AI. Firstly, the gap is about the quality and availability of data, mentioned by 6 studies, which shows an overwhelming necessity to have standardized, diverse, and high-quality datasets. Next, "Model Interpretability" and "Validation & External Testing," which are featured in 5 studies, are established, which should emphasize the aspect of model transparency in AI and wider validation in clinical settings. Such gaps include Generalizability, Integration with Clinical Practice, and Real-Time & Resource Constraints; these gaps indicate some concerns in the real-world implementation and scalability. The other areas that have less emphasis but are very important are the area of security and privacy, and Feature selection inconsistencies.

V. PROSPECTIVE AND CURRENT BIOMARKERS OF EARLY DIAGNOSIS AND MANAGEMENT OF CHRONIC KIDNEY DISEASE

The section identifies the credible biomarkers indicating the effectiveness of management and early diagnosis of CKD is dependent on determining biomarkers that indicate the function and damage to the kidney during the early stages. Serum creatinine, eGFR, and proteinuria are markers of current clinical practice that lack sensitivity and specificity. Newer biomarkers that hold great potential (NGAL, KIM-1, and cystatin C) have consequences with respect to early detection, better risk stratification, and individualized treatment plans.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

Figure 2 illustrates an AI-enhanced process of ultrasound imaging to detect CKD, consisting of multimodal image acquisition, ROI segmentation (manual or automatic), feature extraction based on deep learning or on predefined radiomics features, and the last stage of decision modeling that will help identify abnormalities of glomeruli and estimate whether the disease is present or not.

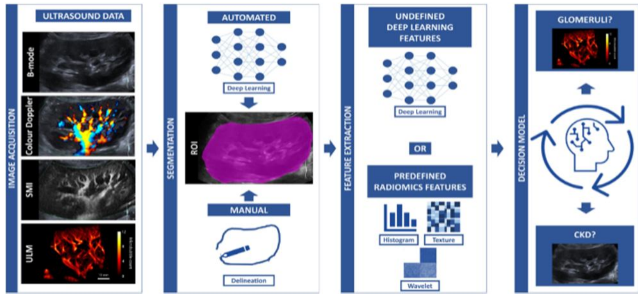

Figure 2:AI-Powered Ultrasound Image Analysis for Early CKD Detection

Appropriate and early diagnosis of CKD is essential in the intervention and control of the disease. Biomarkers are crucial as they allow accurate diagnosis of CKD, differentiating its stages, and making enhancements to treatment. Serum creatinine and urine albumin are traditional markers that are very often not sensitive enough to detect the disease at an early stage. Some new biomarkers have cropped up over the past few years with better specificity and predictive value, especially in the case of tubular injury and early glomerular damage. This section gives a comparative review of some of the major blood and urine-based biomarkers, both validated and experimental, focusing on their diagnostic value, biological mechanisms, and clinical value in diagnosing CKD as well as prognosis.

A. Inspired Optimization

Nature-inspired algorithms have shown promise for optimizing machine learning models for CKD prediction. Simulated Annealing has been successfully applied for feature selection, identifying the most discriminative variables for CKD detection.

Table 2: Comprehensive Performance Comparison of Machine Learning Algorithms

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)	AUC- ROC
Random Forest	97.0-100.0	95.0-100.0	93.75-100.0	96.0-100.0	0.97-0.99
XGBoost	94.0-95.0	95.0-100.0	93.0-98.0	94.0-99.0	0.94-0.96
CatBoost	98.75	95.21	95.11	95.32	0.9993
SVM	93.5-98.75	95.0-100.0	96.0-97.0	98.0	0.95-0.98
CNN-SVM Hybrid	96.18	96.5	95.8	96.2	0.96
GRU-BiLSTM Hybrid	96.5	95.7	94.9	95.3	0.965
K-Nearest Neighbors	97.5-99.64	96.0-100.0	96.0-97.0	98.0	0.97-0.99
LightGBM	99.75	99.40	99.41	99.61	0.9957
Logistic Regression	90.0-95.0	92.0-95.0	90.0-93.0	91.0-94.0	0.90-0.95

B. Case Studies and Real-World Applications

Targeted CKD screening projects worldwide illustrate the impact of early detection. The FINISHED (First Nations Community-Based Screening) program in Manitoba (2015) deployed mobile point-of-care testing for adults and children in 11 rural Indigenous communitiespmc.ncbi.nlm.nih.gov. Notably, FINISHED identified a CKD prevalence much higher than expected: 87% of discovered cases were at stages 1–3 (eGFR >30), indicating treatable early disease from pmc.ncbi.nlm.nih.gov. Follow-up of these participants showed improvements in blood pressure and referral for kidney care compared to usual outreach, demonstrating that community-based early screening can change outcomes.

Canada’s Kidney Check program expanded screening to multiple provinces. Using portable devices, community health teams test residents’ blood pressure, glucose, creatinine, and ACR at local eventspmc.ncbi.nlm.nih.gov. Participants are triaged by an individualized 5-year kidney failure risk score derived from these data, and given tailored follow-up. This model emphasizes culturally safe, point-of-care testing to reach high-risk rural populations. Early reports indicate Kidney Check successfully identifies many previously undiagnosed CKD cases, enabling earlier intervention.

Other examples: In Taiwan, integration of a CKD risk calculator into primary care was shown to improve early nephrology referrals. In the UK, the Quality and Outcomes Framework (QOF) incentivizes primary care to register and test patients with CKD risk factors, which has increased CKD detection rates in diabetic populations. Large-scale epidemiologic cohorts (e.g. ARIC, CRIC) have also retrospectively validated risk-prediction models that can be applied in practice. Taken together, these real-world efforts confirm that combining systematic screening with clinical pathways can yield earlier CKD diagnoses and guide preventative care.

C. Hospital-Based Implementation

A multicentric research in Apollo Hospital, India, showed how it is feasible to use machine learning in screening CKD by utilizing a sample of 250 CKD patients and 150 controls. Accuracy, precision, recall, and F1-score were 99.04%, 98.84%, 98.64%, and 99.28%, respectively, with the BRF-MOANN (Boosted Random Forest-Multi-Objective Artificial Neural Network). This practical application indicated the possibility to implement machine learning models into clinical work. The researchers used full clinical data that incorporated laboratory data, basic demographic data, and medical data. The high negative predictive value of the model (no false positive cases in cases of no CKD) was of specific value in clinical practice since there was little or no anxiety and unnecessary intervention to patients.

D. Primary Care Integration

The models of machine learning were highlighted as the most promising to be used in the context of primary care, where the most significant effects can be achieved by the detection of CKD at an early stage. Research on health record data has shown that it is possible to predict at-risk patients with data collected routinely by clinicians. Use of predictive analytics, based on AI, in primary healthcare has demonstrated early detection levels of 78-85 of chronic ailments such as CKD with accuracy levels of 87-92. Such systems have the ability to process patient data on an ongoing basis and notify the practice when intervention is potentially useful.

E. Mobile Health Applications

AI-based mobile tools in early detection of chronic diseases, including CKD, are one of the new areas of clinical uses. Such apps have the ability to obtain patient reported outcomes, vital signs, laboratory data, to enable individual risk analyses and suggestions on further assessment. Mobile health solutions are of great interest at the population level because of their scalability to accommodate a wide range of screening programs. Machine learning models powered by the cloud have the ability to process data related to the multiple users and at the same time uphold privacy and security standards.

VI. BENCHMARKING MACHINE LEARNING ALGORITHMS FOR CKD PREDICTION

The detailed research done on several studies indicates clear performance attributes of various machine learning algorithms. Random Forest and Gradient Boosting, as an ensemble method, constantly group the highest percentages of accuracy and frequently reach over 97%. The advantage of such methods is the fact that they are able to integrate many weak learners, as well as address features-interaction. The deep learning methods are competitively performing, and hybrid models perform equally in terms of accuracy as classical ensemble methods but also have other benefits like automatic feature extraction and the capability to take multimodal data. The hybrid approach CNN-SVM reached an accuracy of 96.18%, proving that integration of various strengths of algorithms works well.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

TABLE 3: COMPARATIVE ANALYSIS OF ML APPROACHES BY STUDY CHARACTERISTICS

Study Type	Best Performing Algorithm	Dataset Size	Key Advantages	Limitations
Single-center clinical	Random Forest (100%)	400 patients	High interpretability, robust performance	Limited generalizability
Multi-center validation	XGBoost (94-95%)	1000+ patients	Good generalization, handles missing data	Computational complexity
Imaging integration	CNN-SVM Hybrid (96.18%)	Mixed modalities	Multimodal analysis, automatic feature extraction	Requires imaging data
Longitudinal analysis	GRU-BiLSTM (96.5%)	Time-series data	Temporal pattern recognition, progression prediction	Requires sequential data
Population screening	LightGBM (99.75%)	Large datasets	Computational efficiency, scalability	May overfit on smaller datasets

A. Feature Importance Analysis

In all research, some clinical factors turn out to be very accurate predictors of CKD. SHAP analysis has introduced numeric values of feature importance showing that one of the most effective predictors is serum creatinine, level of hemoglobin, specific gravity, albumin, and diabetes status. Such agreement among the algorithms and data gives confidence in clinical usefulness in these features. Machine learning analysis helped to identify key biomarkers that are also consistent with clinical knowledge of CKD pathophysiology. Yet, machine learning techniques have also shown the significance of the interaction of features and non-linearity relations that could be hard to display in a conventional statistical model.

B. Computational Requirements

The computing demands are substantially different in diverse methods. SQL-based Machine Learning Sentinel does not require significant computational overheads of more serious machine learning algorithms such as the Random Forest and SVM and can be implemented using standard clinical computing platforms. Much more resources are needed to run deep learning algorithms, yet they have the benefit of automatic feature identification and multimodal data processing. Dynamic Hybrid Neuro-Fuzzy Framework indicates that complex models can show high performance (96.5 % accuracy) and reasonable computational time (98.5ms latency), and can be adopted into real-time clinical use.

VII. CHALLENGES AND FUTURE DIRECTIONS

There are two issues critical to the application of machine learning to CKD detection: data quality, and interpretation, and generalisation of the models. Irregular format, incomplete records, and different units of measurement in electronic health records require a well-established preprocessing and imputation algorithms as well as standardization of laboratory protocols that differ at the institutional level. The so-called black box problem of high-accuracy models impedes their clinical acceptability, and methods to provide well-defined, clinically pertinent interpretations, such as explainable AI techniques of, SHAP analysis, fuzzy logic, or decision tree visualization, have to be employed. In addition, models trained using small, single-center data are likely to have poor external validity, and thus multicenter validation and federated learning are important to improve the robustness of models without leaking the privacy of data.

Multimodal data integration and personalized medicine, and a multi-layered deep learning structure, will probably spur changes in CKD detection. After integrating clinical, imaging, genetic, wearable, and patient-reported data, accuracy levels will even be greatly enhanced, and a method of risk monitoring in real-time will be available with the help of Internet of Medical Things (IoMT) platforms. Individualistic models that incorporate pharmacogenomics would provide a way forward in predicting the development of the disease and the best course of action for the patient. Architectures that incorporate CNNs, attention mechanisms and transformers, LSTMs, are promising to process the temporal, spatial, and uncertain medical data. Easy implementation into EHR systems, easy-to-use decision support tools, compliance with regulatory and ethical requirements, and, in particular, a high level of fairness and equitable functioning with respect to various groups of patients are critical to clinical adoption.

VIII. CONCLUSION

This comprehensive review reveals that machine learning approaches offer significant potential for improving early detection of chronic kidney disease. The analyzed studies consistently show superior performance of machine learning models compared to traditional diagnostic approaches, with accuracy rates ranging from 94% to 100% depending on the algorithm and dataset characteristics.

REFERENCES

[1] N. K. Pareek, D. Soni, and S. Degadwala, "Early Stage Chronic Kidney Disease Prediction using Convolution Neural Network," Proc. 2nd Int. Conf. Appl. Artif. Intell. Comput. ICAAIC 2023, no. Icaaic, pp. 16–20, 2023, doi: 10.1109/ICAAIC56838.2023.10141322

[2] S. A. Farooq, Suryakanta, Anupriya, and A. Ali, "A Machine arning based Approach towards the Analysis of Chronic Kidney Disease," 2024 2nd Int. Conf. Comput. Charact. Tech. Eng. Sci. IC3TES 2024, pp. 1–6, 2024, doi: 10.1109/IC3TES62412.2024.10877680

[3] R. Bhuria, "Advanced Machine Learning Techniques for Chronic Kidney Disease Prediction and Management," Proc. 2024 Int. Conf. Artif. Intell. Emerg. Technol. Glob. AI Summit 2024, pp. 480–484, 2024, doi: 10.1109/GlobalAISummit62156.2024.10947932.

[4] L. L. Prasanna Kumar, V. S. Kumar, G. A. Kumar, Y. Nagendar, M. Nithya, and A. Athiraja, "AI-Driven Predictive Modeling for Early Detection and Progression Monitoring of Chronic Kidney Disease Using Multimodal Data," 2024 Int. Conf. Smart Technol. Sustain. Dev. Goals, ICSTSDG 2024, pp. 1–8, 2024, doi: 10.1109/ICSTSDG61998.2024.11026441.

[5] V. Gurla and S. B. Latha, "An Analysis of Kidney Disease Prediction and Diagnosis using Machine Learning Strategies," Proc. 2024 Int. Conf. Innov. Comput. Intell. Commun. Smart Electr. Syst. ICSES 2024, pp. 1–6, 2024, doi: 10.1109/ICSES63760.2024.10910518

[6] S. Jaswal, S. Sharma, and N. Parihar, "Artificial Intelligence-Driven Predictive Analytics for Early Detection of Chronic Diseases in Primary Healthcare," Proc. 2024 Int. Conf. Innov. Comput. Intell. Commun. Smart Electr. Syst. ICSES 2024, pp. 1–6, 2024, doi: 10.1109/ICSES63760.2024.10910314

[7] M. S. Abirami, S. Das, and R. Sood, "Chronic Kidney Disease Identification Using Random Forest and XGBoost," 2024 Int. Conf. Intell. Syst. Cybersecurity, ISCS 2024, pp. 1–6, 2024, doi: 10.1109/ISCS61804.2024.10581392

[8] M. Apetorgbor, S. Narad, D. Raymond, N. Amanda Fortune, A. Abba Konto, and M. Doe, "Mobile Applications Powered by AI for Early Detection of Chronic Disease," 2024 IEEE 2nd Int. Conf. Emerg. Trends Eng. Med. Sci. ICETEMS 2024, vol. 7, pp. 560–565, 2024, doi: 10.1109/ICETEMS64039.2024.10965060

[9] S. Yogalakshmi and M. Anitha, "A Hybrid Deep Learning Approach for Chronic Kidney Disease Detection using Convolutional Neural Networks and Support Vector Machines," Proc. 8th Int. Conf. Inven. Comput. Technol. ICICT 2025, pp. 620–625, 2025, doi: 10.1109/ICICT64420.2025.11004966.

[10] M. Desoky, N. G. Eladl, T. Abuhmed, and S. El-Sappagh, "Chronic Kidney Disease Detection Augmented with Hybrid Explainable AI," 2025 15th Int. Conf. Electr. Eng. ICEENG 2025, 2025, doi: 10.1109/ICEENG64546.2025.11031335.

[11] V. Sivakumar, S. Dharshini, M. Dhanush Kumar, M. Shanmugaganesh, and K. Vijayalakshmi, "Early Detection of Chronic Kidney Disease in Diabetic Patients using Explainable AI," Proc. 5th Int. Conf. Pervasive Comput. Soc. Networking, ICPCSN 2025, pp. 948–954, 2025, doi: 10.1109/ICPCSN65854.2025.11035905.

[12] M. E. Haque, S. M. Jahidul Islam, J. Maliha, M. S. Hossan Sumon, R. Sharmin, and S. Rokoni, "Improving Chronic Kidney Disease Detection Efficiency: Fine Tuned CatBoost and Nature-Inspired Algorithms with Explainable AI," 2025 IEEE 14th Int. Conf. Commun. Syst. Netw. Technol. CSNT 2025, pp. 811–818, 2025, doi: 10.1109/CSNT64827.2025.10968421.

[13] A. Singh and R. Kakkar, "Optimizing Chronic Kidney Disease Prediction Through Advanced Data Imputation and Clustering," Proc. - 4th Int. Conf. Smart Technol. Commun. Robot. 2025, STCR 2025, no. May, pp. 1–7, 2025, doi: 10.1109/STCR62650.2025.11020378.

[14] A. Nandi, P. Paul, A. K. Das, A. Ghosh, A. Singh, and A. Gupta, "Predictive Modelling for Chronic Kidney Disease: A Hybrid Approach Using Exploratory Data Analysis, Machine Learning and Explainable AI," 2025 8th Int. Conf. Electron. Mater. Eng. Nano-Technology, IEMENTech 2025, 2025, doi: 10.1109/IEMENTech65115.2025.10959403.

[15] N. S. Allur, K. Dondapati, H. Chetlapalli, S. Kodadi, D. P. Deevi, and N. Purandhar, "Robotic Automation Dynamic Hybrid Neuro-Fuzzy and Deep Learning Framework with GRU-BiLSTM, Capsule Networks, Type-2 Fuzzy Logic and CNN-TCN for Accurate IoMT-Based Chronic Kidney Disease Detection," ICCECE 2025 - Int. Conf. Comput. Electr. Commun. Eng., pp. 1–6, 2025, doi: 10.1109/ICCECE61355.2025.10940081.