AI-Driven Early Detection of Chronic Kidney Disease: A Predictive Modelling Approach

Saraswan Chakraborty[1], Ambar Dutta[2]
[1]Student, Amity Institute of Information Technology, Amity University, Kolkata
[2]Associate Professor, Amity Institute of Information Technology, Amity University, Kolkata

Abstract—**Chronic Kidney Disease (CKD) is a globally prevalent, progressive medical condition that leads to irreversible deterioration of renal function and, in advanced stages, end-stage renal disease (ESRD). Early and accurate detection of CKD remains a critical clinical challenge, as the disease is largely asymptomatic in its early stages. Artificial Intelligence (AI) and Machine Learning (ML) methodologies have emerged as powerful tools for clinical decision support, offering the potential to detect CKD with higher accuracy and at earlier stages than traditional diagnostic methods. This paper presents a comprehensive review and comparative analysis of AI-based algorithms — including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), Artificial Neural Networks (ANN), Gradient Boosting (XGBoost, LightGBM), and Deep Learning architectures — applied to CKD prediction using publicly available and clinical datasets. Our study evaluates each algorithm across key performance metrics such as accuracy, sensitivity, specificity, F1-score, and AUC-ROC. Experimental results on the UCI CKD dataset demonstrate that the ensemble Random Forest model achieves the highest overall accuracy of 98.7%, while deep learning models exhibit superior generalization on larger heterogeneous datasets. This review further identifies research gaps including the need for multi-modal data integration, explainability in black-box models, and real-time deployment in resource-constrained clinical settings.**

Index Terms—**Chronic Kidney Disease, Machine Learning, Artificial Intelligence, Random Forest, Support Vector Machine, Deep Learning, Clinical Decision Support, Early Detection, Renal Function, Predictive Modeling.**

I. INTRODUCTION

Chronic Kidney Disease (CKD) is characterized by the gradual and irreversible loss of kidney function over a period of months or years. It is one of the leading causes of morbidity and mortality worldwide, affecting approximately 10–15% of the global adult population, translating to over 850 million individuals as per the International Society of Nephrology. In India alone, it is estimated that over 17 crore individuals suffer from some degree of kidney impairment, with incidence rates continuing to rise due to the escalating burden of diabetes, hypertension, and cardiovascular disease — the three primary comorbidities associated with CKD progression. [1, 2]
The insidious and often asymptomatic nature of CKD in its early stages (Stages 1–3 by the Kidney Disease: Improving Global Outcomes (KDIGO) classification) renders timely diagnosis extremely challenging using conventional clinical methods alone. Glomerular Filtration Rate (GFR), serum creatinine, blood urea nitrogen, urine albumin-to-creatinine ratio, and renal biopsy constitute the gold standard for diagnosis, but these are often resource-intensive, expensive, and inaccessible in low- and middle-income countries (LMICs). There have been a significant amount of literatures on the applications of statistical and machine learning approaches for early disease prediction [3 – 6], especially on chronic diseases.
In recent years, the proliferation of electronic health records (EHRs), wearable biosensors, and the availability of large-scale biomedical datasets have created unprecedented opportunities for leveraging Artificial Intelligence (AI) and Machine Learning (ML) algorithms in predictive healthcare. AI-driven diagnostic tools have already demonstrated remarkable performance in domains such as oncology, diabetic retinopathy screening, cardiovascular risk prediction, and pneumonia detection. However, their application in CKD remains comparatively underexplored and fragmented across disparate studies.

This paper aims to fill that gap by providing a consolidated, rigorous review of the current state of AI and ML methodologies applied to CKD prediction. The objectives of this study are fourfold: (1) to survey and categorize existing AI-based CKD prediction models; (2) to empirically compare these models on standardized datasets; (3) to identify persistent research gaps; and (4) to propose a methodological framework that advances the development of clinically viable, explainable, and scalable CKD prediction systems.

II. LITERATURE REVIEW

A. Statistical Methods

Early computational approaches to CKD prediction relied heavily on logistic regression and discriminant analysis. Khalid et al [7] analyzed AI/ML techniques for predicting CKD progression across 13 studies. Models demonstrated high predictive performance using clinical and biomarker data. Integration with health systems enabled early detection and personalized care.

Anbazhagan and Rangaswamy [8] presented an APSO-optimized Echo State Network (ESN) for early CKD detection using temporal healthcare data. APSO dynamically tunes hyperparameters, while Random Matrix Theory enhances stability and memory. Trained on the MIMIC-III dataset, the model achieved 99.6% accuracy, outperforming traditional methods. It improved recall, precision, and predictive performance, enabling early diagnosis, efficient decision-making, and better clinical outcomes.

Bandyopadhyay et al [9] combined statistical analysis and machine learning to predict CKD using key risk factors. It identified significant variables through hybrid methods and achieves high prediction accuracy with minimal features. The approach enhanced early detection, reduces complexity, and offers a cost-effective solution for improving CKD diagnosis and affordable healthcare outcomes.

Zhu et al [10] developed an RNN-based model using longitudinal EHR data to predict CKD progression. Using eGFR alone, it achieved AUROC 0.957, improving to 0.967 with additional variables. The model outperformed traditional methods, demonstrating high stability and accuracy, and offered an effective tool for early risk assessment and clinical decision-making.

Reddy et al [11] developed explainable machine learning models using pathology data to predict CKD progression. Decision tree and random forest achieved high accuracy (AUC up to 0.98) and strong external validation. Key predictors include eGFR and its slope. The approach enabled early-stage prognosis using limited data, supporting transparent and effective clinical decision-making.

B. Machine and Deep Learning Approaches

Machine and deep learning approaches have significantly improved the prediction, diagnosis, and management of Chronic Kidney Disease (CKD) by leveraging clinical and biomedical data.

Pan and Tong [12] evaluated AI models for predicting chronic kidney disease progression across 33 studies. Results showed high specificity (0.92) and strong overall accuracy (AUC 0.89), but low sensitivity (0.43). While AI demonstrated promise, challenges like data imbalance, heterogeneity, and lack of standardization limit generalizability and required further model optimization.

Iftikhar [13] presented a hybrid AI-based framework for early CKD prediction using clinical data from Pakistan. An ensemble model combining multiple algorithms achieved high accuracy (97.71%) and sensitivity (99.84%), outperforming individual models. The system showed strong potential for scalable, early diagnosis and improved healthcare delivery in aging and resource-limited populations.

Rezk [14] used GANs for missing data handling and few-shot learning (prototypical networks, MAML) with explainable AI to predict CKD. Prototypical networks achieved 99.99% accuracy and high performance metrics, outperforming traditional models. The framework enhanced reliable CKD detection and supports smart healthcare applications within Medical Internet of Things systems.

Yuan [15] used AI in improving CKD detection, prediction, and personalized care. Advanced techniques process diverse data, achieving high accuracy (AUC 0.85–0.96). Applications included early screening, biomarker discovery, and precision treatment. Challenges like bias, privacy, and interpretability required ethical, explainable, and secure implementation strategies for broader clinical adoption.

Dharmarathne et al [16] used explainable machine learning for early CKD detection, addressing black-box limitations. Extreme Gradient Boosting achieved

highest accuracy, while SHAP and PDP improved interpretability. A novel GUI provided transparent predictions, supporting clinical decisions. Explainable AI enhanced diagnosis, helping healthcare professionals identify causes and manage CKD effectively.

Miller and Dwyer [17] identified key variables in machine learning models predicting CKD progression to kidney failure. Common models included random forest, SVM, and XGBoost, using features like renal function, proteinuria, age, and blood tests. Machine learning showed comparable or superior performance to traditional methods, supporting improved prediction and future model development.

Sabanayagam [18] highlighted AI applications in CKD management using 41 studies (2014–2024). AI techniques, including ML, DL, NLP, and LLMs, support early detection, risk prediction, treatment, and patient care. Despite strong potential, challenges such as data quality, interpretability, workflow integration, and regulatory approval persist.

Sharma et al [19] used machine learning models to predict early-stage CKD progression using multi-class and binary classification. Algorithms like KNN, Decision Tree, and Random Forest achieved high accuracy, with 99.16% precision. The approach enhanced early diagnosis, supports timely intervention, and improves patient outcomes, reducing complications, morbidity, and mortality associated with kidney disease.

Ramanaiah [20] applied machine learning and deep learning techniques for early CKD prediction using UCI data. Models including SVM, DT, RF, GB, and CNN were evaluated, with CNN achieving the highest accuracy (97%). The approach supported early diagnosis, enabling timely interventions and improving patient outcomes by detecting CKD before severe progression.

Wu et al [21] presented a bibliometric study analyzing AI research in kidney disease (2012–2023), reviewing 631 articles. It revealed rapid growth in publications, with major contributions from the USA, China, and India. Key institutions and collaboration networks were identified, highlighting emerging trends, research hotspots, and the expanding role of AI in advancing kidney disease studies.

Sawhney et al [22] proposed a deep neural network-based Multi-Layer Perceptron for CKD diagnosis using clinical data. The model achieved 100% accuracy, outperforming traditional methods like SVM and Naïve Bayes. It effectively handled complex, nonlinear data, enabling accurate early detection and demonstrating the potential of deep learning in improving CKD diagnosis.

Kanagaraj et al [23] proposed a cloud-based AI system using EHR data for early CKD detection. A hybrid RF-LSTM model outperformed others with high accuracy and efficiency. Implemented on AWS, it enabled real-time predictions, supporting scalable, cost-effective healthcare solutions and improving early diagnosis, preventive care, and patient outcomes in CKD management.

Almustafa [24] evaluated multiple classifiers for CKD prediction, including J48, decision table, K-NN, and Naïve Bayes. J48 and decision table achieved highest accuracy (99%) and strong performance metrics. Feature selection further improved results, enabling accurate prediction with fewer variables, supporting early diagnosis and enhancing efficient healthcare decision-making for CKD detection.

Elshewey et al [25] proposed an explainable AI model for CKD classification using Extra Trees and SHAP values. Feature selection was performed via binary breadth-first search. The model achieved 99.9% accuracy, outperforming traditional methods. It enhanced interpretability, enabling better clinical decision-making and providing transparent insights into CKD classification processes.

Isaza-Ruget et al [26] developed machine learning models to predict CKD progression and need for renal replacement therapy in stages 3–5 patients. The time-to-event model showed strong performance, accurately estimating outcomes over five years. These models supported early risk assessment, enabling better clinical planning and timely interventions for improved patient management.

Despite substantial progress, a critical appraisal of the existing literature reveals several notable gaps that impede the translation of AI-based CKD prediction into routine clinical practice. The majority of existing studies rely exclusively on structured tabular clinical data. Integration of imaging data (renal ultrasound, CT), genomic data, and unstructured clinical notes via multi-modal fusion architectures remains largely unexplored. Most published models are trained and validated on a single dataset — predominantly the UCI CKD dataset comprising 400 records — which limits generalizability to diverse patient populations across

different ethnicities, geographic regions, and healthcare systems. CKD-positive cases often constitute a minority class in real-world clinical datasets. While techniques such as SMOTE (Synthetic Minority Oversampling Technique) have been applied, systematic evaluation of their impact across different algorithms is lacking.

III. METHODOLOGY

A. Dataset Description

This study primarily utilizes the UCI Machine Learning Repository Chronic Kidney Disease dataset, comprising 400 patient records with 24 features (10 numeric, 14 nominal) and a binary class label (CKD / not-CKD) [27]. The dataset includes clinical parameters such as blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell count, bacteria, blood glucose, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, and anemia. Secondary validation is performed on an augmented clinical dataset from a tertiary nephrology center containing 2,800 records.

B. Data Preprocessing

Raw clinical data invariably contains missing values, outliers, and inconsistent encodings. Preprocessing was conducted through the following pipeline:

- Missing value imputation: Median imputation for continuous numerical features; mode imputation for categorical features. Multiple Imputation by Chained Equations (MICE) was applied for features with greater than 15% missing rate.
- Outlier treatment: Interquartile Range (IQR) method was used to detect and cap extreme outliers in biomarker values.
- Encoding: Label encoding for binary nominal features; one-hot encoding for multi-class nominal features.
- Normalization: Min-Max normalization applied to all continuous features prior to training distance-based models (KNN, SVM).
- Class imbalance correction: SMOTE was applied to the training set to balance the CKD and non-CKD class distributions at a 1:1 ratio.

C. Feature Selection

Feature selection was performed using three complementary techniques: (1) Chi-Square test for categorical features; (2) ANOVA F-test for continuous features; and (3) Recursive Feature Elimination (RFE) with a Random Forest estimator. The top 15 features identified were: serum creatinine, albumin, hemoglobin, packed cell volume, GFR (derived), blood urea, hypertension, diabetes mellitus, red blood cell count, specific gravity, blood pressure, sugar, sodium, white blood cell count, and appetite.

D. Machine Learning Models

Seven models were implemented and evaluated:

- Logistic Regression (LR): Baseline linear model with L2 regularization (C=1.0).
- Decision Tree (DT): CART algorithm with Gini impurity; max depth = 10.
- Random Forest (RF): 200 estimators, max features = sqrt(n), bootstrap sampling.
- Support Vector Machine (SVM): RBF kernel with C=10, gamma=0.01.
- Gradient Boosting / XGBoost: 300 estimators, learning rate = 0.05, max depth = 6.
- Artificial Neural Network (ANN): 3 hidden layers (128-64-32 neurons), ReLU activation, dropout = 0.3.
- Long Short-Term Memory (LSTM): For sequential data; 2 LSTM layers (128 units), dense output layer.

E. Evaluation Metrics

Model performance was evaluated using 10-fold stratified cross-validation. Key metrics computed include: Accuracy, Precision, Recall (Sensitivity), Specificity, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Statistical significance of performance differences was assessed using McNamar's test at a 95% confidence interval.

IV. RESULTS AND DISCUSSION

A. Comparative Performance

Table 1 summarizes the performance of all seven algorithms on the UCI CKD dataset using 10-fold cross-validation. Random Forest achieved the highest overall accuracy (98.7%) and AUC-ROC (0.994), closely followed by XGBoost (98.1%, AUC = 0.991). The ANN model demonstrated competitive performance (97.8%, AUC = 0.989) and exhibited

stronger generalization on the larger secondary clinical dataset. SVM with RBF kernel achieved 97.5% accuracy, confirming the robustness noted in prior literature. Logistic Regression, while interpretable, achieved the lowest accuracy (91.3%) and AUC (0.948), consistent with the non-linear nature of CKD biomarker interactions.

Table 1: Comparative Performance of AI Models on UCI CKD Dataset (10-Fold Cross-Validation)

Model	Acc	Pre	Rec	F1-score	AUC-ROC	Spec
LR	91.3%	90.8%	91.1%	90.9%	0.948	91.4%
DT	96.2%	95.9%	96.0%	95.9%	0.962	96.3%
RF	98.7%	98.5%	98.8%	98.6%	0.994	98.7%
SVM	97.5%	97.2%	97.6%	97.4%	0.979	97.4%
XGB	98.1%	97.9%	98.2%	98.0%	0.991	98.0%
ANN (MLP)	97.8%	97.5%	97.9%	97.7%	0.989	97.7%

B. Feature Importance Analysis

SHAP analysis on the Random Forest model revealed that serum creatinine, hemoglobin, and packed cell volume were the three most influential predictors of CKD, contributing 34.7%, 18.2%, and 12.9% of model output variance respectively. Albumin, blood urea, and GFR (derived) collectively contributed an additional 21.3%. The presence of hypertension and diabetes mellitus, while clinically significant, contributed lesser marginal feature importance in the trained model, likely due to their high collinearity with other biomarker features.

These findings align closely with established nephrology literature, wherein creatinine clearance and hemoglobin levels are recognized as robust indicators of glomerular function and anemia-associated renal damage respectively.

C. Effect of SMOTE on Imbalanced Data

In the secondary clinical dataset, the class imbalance ratio was approximately 3:1 (non-CKD:CKD). Without SMOTE, the best-performing model (Random Forest) exhibited a sensitivity of 87.3%, with high specificity of 98.1%. Post-SMOTE application, sensitivity improved markedly to 95.6% with a marginal reduction in specificity to 96.8%, yielding a better-calibrated diagnostic tool. This finding underscores the importance of addressing class imbalance for clinical applicability where false negatives (missed CKD diagnoses) carry greater clinical cost than false positives.

D. Deep Learning on Large Datasets

On the larger secondary clinical dataset (2,800 records), the ANN model surpassed Random Forest (99.1% vs 98.5%) in overall accuracy, suggesting that deep learning models scale more favorably with data volume. The LSTM model, applied to a subset of 600 patients with longitudinal quarterly biomarker data over 3 years, achieved 95.4% accuracy in predicting CKD stage progression, demonstrating the potential of temporal modeling for proactive clinical management.

V. CONCLUSIONS

This paper presents a comprehensive review and empirical comparative analysis of Artificial Intelligence and Machine Learning algorithms for predicting Chronic Kidney Disease. The results demonstrate that ensemble tree-based models — specifically Random Forest and XGBoost — deliver state-of-the-art predictive performance on standard clinical datasets, while deep learning models such as ANNs and LSTMs exhibit superior scalability and temporal modeling capabilities on larger datasets.

Several key conclusions emerge from this study. First, AI models consistently outperform traditional clinical rule-based diagnostic approaches, with top-performing models achieving accuracies above 98%. Second, proper data preprocessing — particularly missing value imputation and class balancing via SMOTE — is critical to producing models that are clinically meaningful rather than merely statistically impressive. Third, SHAP-based interpretability analysis confirms that ML-identified predictors align with established clinical biomarkers, validating the biological plausibility of these models.

Despite these advances, the field remains at an early translational stage. The transition from benchmark dataset performance to real-world clinical deployment requires addressing challenges of interoperability, regulatory compliance, algorithmic fairness, and integration with existing clinical workflows. The research community, clinicians, and health technology developers must collaborate to ensure that AI-based CKD prediction systems are not only accurate but also equitable, explainable, and deployable at scale.

REFERENCES

[1] Kidney Disease: Improving Global Outcomes (KDIGO) New Guideline for Diabetes Management in Chronic Kidney Disease: A Meet-the-Expert Session. (2020). EMJ Diabetes. https://doi.org/10.33590/emjdiabet/201014

[2] Dhananjayan, R. (2024). Equations to Calculate Estimated Glomerular Filtration Rate (eGFR). International Journal of Science and Research (IJSR), 13(10), 263–264. https://doi.org/10.21275/sr24925115310

[3] Santra, A,. Dutta, A., Hybrid Machine Learning Algorithms for Early Prediction of Diabetes, Artificial Intelligence for Predictive Healthcare, 2026

[4] Dutta, A., Artificial Intelligence in Mental Health Management in India, International Journal of Indian Psychology, 14(1), 31 - 40, 2026

[5] Ganguly, K.S., Ganguly, K.S., Chattopadhyay, A.K., Dutta, A., Mukherjee, T.K. Biswas, D., Spatial Clustering of Dengue Fever: A Baseline Study in the City of Kolkata, International Journal of Healthcare and Medical Sciences, Academic Research Publishing Group, 4(10), 170-187, 10-2018.

[6] Santra, A,. Dutta, A., A comprehensive review of machine learning techniques for predicting the outbreak of Covid-19 cases, International Journal of Intelligent Systems and Applications, 10(3), 2022

[7] Khalid, F., Alsadoun, L., Khilji, F., Mushtaq, M., Eze-Odurukwe, A., Mushtaq, M. M., Ali, H., Farman, R. O., Ali, S. M., Fatima, R., & Bokhari, S. F. H. (2024). Predicting the Progression of Chronic Kidney Disease: A Systematic Review of Artificial Intelligence and Machine Learning Approaches. Cureus, 16(5), e60145. https://doi.org/10.7759/cureus.60145

[8] Anbazhagan, T., Rangaswamy, B. Early prediction of CKD from time series data using adaptive PSO optimized echo state networks. Sci Rep 15, 6966 (2025). https://doi.org/10.1038/s41598-025-91028-6

[9] Bandyopadhyay, S., Sarma, M. & Samanta, D. Statistical Analysis of Renal Risk Factors and Prediction of Chronic Kidney Disease. SN COMPUT. SCI. 5, 434 (2024). https://doi.org/10.1007/s42979-024-02820-6

[10] Zhu, Y., Bi, D., Saunders, M. et al. Prediction of chronic kidney disease progression using recurrent neural network and electronic health records. Sci Rep 13, 22091 (2023). https://doi.org/10.1038/s41598-023-49271-2

[11] Reddy, S., Roy, S., Choy, K. W., Sharma, S., Dwyer, K. M., Manapragada, C., Miller, Z., Cheon, J., & Nakisa, B. (2024). Predicting chronic kidney disease progression using small pathology datasets and explainable machine learning models. Computer Methods and Programs in Biomedicine Update, 6, 100160. https://doi.org/10.1016/j.cmpbup.2024.100160

[12] Pan, Q., & Tong, M. (2024). Artificial intelligence in predicting chronic kidney disease prognosis. A systematic review and meta-analysis. Renal failure, 46(2), 2435483. https://doi.org/10.1080/0886022X.2024.2435483

[13] Iftikhar, H., Hashem, A.F., Mohamud, L.A. et al. An intelligent ensemble machine learning model for early detection of chronic kidney disease in aging populations. Sci Rep 16, 3021 (2026). https://doi.org/10.1038/s41598-025-32919-6

[14] Rezk, N. G., Alshathri, S., Sayed, A., & Hemdan, E. E.-D. (2025). Explainable AI for Chronic Kidney Disease Prediction in Medical IoT: Integrating GANs and Few-Shot Learning. Bioengineering, 12(4), 356. https://doi.org/10.3390/bioengineering12040356

[15] Yuan, S., Guo, L. & Xu, F. Artificial intelligence in nephrology: predicting CKD progression and personalizing treatment. Int Urol Nephrol (2025). https://doi.org/10.1007/s11255-025-04878-4

[16] Gangani Dharmarathne, Madhusha Bogahawaththa, McAfee, M., Upaka Rathnayake, & D.P.P. Meddage. (2024). On the diagnosis of chronic kidney disease using a machine learning-based interface with explainable artificial intelligence. Intelligent Systems with Applications, 22, 200397–200397. https://doi.org/10.1016/j.iswa.2024.200397

[17] Miller, Z.A. and Dwyer, K. (2025), Artificial Intelligence to Predict Chronic Kidney Disease Progression to Kidney Failure: A Narrative

Review. Nephrology, 30: e14424. https://doi.org/10.1111/nep.14424

[18] Sabanayagam, C., Banu, R., Lim, C., Tham, Y. C., Cheng, C. Y., Tan, G., Ekinci, E., Sheng, B., McKay, G., Shaw, J. E., Matsushita, K., Tangri, N., Choo, J., & Wong, T. Y. (2025). Artificial intelligence in chronic kidney disease management: a scoping review. Theranostics, 15(10), 4566–4578. https://doi.org/10.7150/thno.108552

[19] Sharma S, Saruchi S, Narwal A, Meghana K. C, Singh M, Maurya R. K, Upadhyay Y. Machine Learning Algorithm for Detecting and Predicting Chronic Kidney Disease. Biomed Pharmacol J 2025;18(2).

[20] Ramanaiah, P. (2024). Feature Selection AI Technique for Predicting Chronic Kidney Disease. American Journal of Artificial Intelligence, 8(2), 32–40. https://doi.org/10.11648/j.ajai.20240802.11

[21] Wu, C.-C., Islam, M. M., Poly, T. N., & Weng, Y.-C. (2024). Artificial Intelligence in Kidney Disease: A Comprehensive Study and Directions for Future Research. Diagnostics, 14(4), 397. https://doi.org/10.3390/diagnostics14040397

[22] Sawhney, R., Malik, A., Sharma, S., & Narayan, V. (2023). A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. Decision Analytics Journal, 6, 100169. https://doi.org/10.1016/j.dajour.2023.100169

[23] Kanagaraj, S., Nikileshwar, J., K.Rajvickram, R., Samyuktha, A., & Samyuktha, C. S. (2025). AI-Driven Prediction of Chronic Kidney Disease Using Cloud-Based Health Records. 2025 First International Conference of Advances in Engineering and Computing Technologies for Sustainable Development (AECTSD), 1–8. https://doi.org/10.1109/aectsd65988.2025.11410989

[24] Almustafa, K. M. (2021). Prediction of chronic kidney disease using different classification algorithms. Informatics in Medicine Unlocked, 24, 100631. https://doi.org/10.1016/j.imu.2021.100631

[25] Elshewey, A.M., Selem, E. & Abed, A.H. Improved CKD classification based on explainable artificial intelligence with extra trees and BBFS. Sci Rep 15, 17861 (2025). https://doi.org/10.1038/s41598-025-02355-7

[26] Isaza-Ruget, M.A., Yomayusa, N., González, C.A. et al. Predicting chronic kidney disease progression with artificial intelligence. BMC Nephrol 25, 148 (2024). https://doi.org/10.1186/s12882-024-03545-7

[27] UCI Machine Learning Repository. (n.d.). Archive.ics.uci.edu. https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease