



Chronic Kidney Disease Prediction System Using Machine Learning

Savitha. D¹, Haritha. V P², Mathisha. N³

Assistant Professor, Department of Computer Applications, Vellalar College for Women (Autonomous), Erode, India¹

II-MCA, Department of Computer Applications, Vellalar College for Women (Autonomous), Erode, India^{2,3}

Abstract: Chronic kidney disease (CKD) is a progressive condition that often remains unnoticed until a large part of kidney function has been lost, which makes early screening one of the most effective ways of delaying dialysis and transplantation. Machine learning (ML) can learn the relationship between routine clinical and laboratory measurements and the presence of CKD, and can turn it into a fast, low-cost decision-support tool. This paper presents a structured survey of ten peer-reviewed studies published between 2020 and 2026 on ML-based CKD prediction, covering classical and ensemble classifiers, feature selection, class-imbalance handling, explainable artificial intelligence and progression prediction. The studies are compared in a summary table and the common gaps are identified: reliance on small public datasets, near-perfect but weakly validated accuracy, limited external validation, explanations that never reach the user, and few complete systems that connect prediction to patient records and reports. Based on these gaps, the motivation, application areas and future directions for a practical web-based CKD prediction system built on a Random Forest classifier are discussed.

Keywords: Chronic kidney disease, Machine learning, Randomforest, Early prediction, Explainable AI, Feature selection, Clinical decision support.

I.INTRODUCTION

Chronic kidney disease (CKD) is a long-term, progressive loss of kidney function that is now recognised as a major worldwide health burden (Debal & Sitote, 2022; Qureshi et al., 2026). In its early stages the disease usually causes no noticeable symptoms, so many patients are diagnosed only after serious damage has occurred, when dialysis or kidney transplantation is the only remaining option. Early diagnosis and prompt clinical management are therefore the most practical way to limit complications, delay progression and improve survival.

Diagnosis relies on routine clinical and laboratory measurements such as blood pressure, serum creatinine, blood urea, hemoglobin, albumin and urine findings. Reading these values together requires experience, and specialists are scarce in many regions, particularly in low- and middle-income countries. Machine learning (ML) offers a way to learn the combined pattern of these measurements from historical records. Compared with single-threshold rules, ML models can capture nonlinear relationships and interactions between variables and can return a prediction within seconds (Qureshi et al., 2026).

A large body of work has applied ML to CKD prediction in recent years. Most studies use the public UCI CKD dataset and report accuracy close to 99% with ensemble, boosting and neural models (Ebiaredoh-Mienye et al., 2022; Islam et al., 2023; Qin et al., 2020), while newer studies work with hospital data, class-imbalance handling, feature selection and explainable artificial intelligence (Debal & Sitote, 2022; Ghosh & Khandoker, 2024; Qureshi et al., 2026). Despite this progress, the step from a high-scoring model to a trustworthy tool that clinicians can actually use is still incomplete.

The problem addressed here can be stated simply. Given the routine clinical and laboratory attributes of a patient, a system should decide whether the patient has CKD, express how confident it is, and present the result in a form that supports a clinical decision. The survey therefore asks three questions: which ML approaches have been used for this task since 2020, how well do they perform and under what conditions, and what is still missing before such a model can serve as a dependable screening tool?

This paper offers a focused survey of ten peer-reviewed studies published between 2020 and 2026. The studies are discussed in Section II and compared in Table I; Section III identifies the research gaps; Sections IV and V present the motivation and the application areas of a CKD prediction system; and Section VI concludes with future directions. The survey also underpins the design of a web-based CKD prediction system that combines patient and test records, a Random Forest classifier that returns a diagnosis with its confidence and risk level, and downloadable PDF reports.



II. LITERATURE SURVEY

The reviewed studies fall into four overlapping themes: classical and ensemble classifiers on clinical data, feature selection and imbalance handling, explainable prediction, and progression prediction. The ten studies were chosen from peer-reviewed journals for their relevance to CKD prediction, their publication between 2020 and 2026, and the variety of their methods, so that classical, ensemble, explainable and prognostic approaches are all represented.

A. Classical and Ensemble Classifiers

Qin et al. (2020) proposed an ML methodology for diagnosing CKD that handled missing values before training and combined several classifiers into one integrated model, reporting accuracy above 99% on the UCI data. Islam et al. (2023) compared twelve supervised classifiers and found that XGBoost gave the best results, with an accuracy of 0.983 and precision, recall and F1-score of 0.98. Dubey et al. (2023) evaluated gradient boosting, decision tree, k-nearest neighbors, random forest, histogram boosting and XGBoost, and added a feature-significance analysis to show which clinical attributes drive the prediction. Debal and Sitote (2022) moved away from the UCI data and used a larger and more recent hospital dataset collected in Ethiopia, which makes their comparison of classifiers more representative of real clinical records.

B. Feature Selection and Class Imbalance

Clinical datasets contain redundant attributes and unequal class sizes, and several authors treat these issues explicitly. Ebiaredoh-Mienye et al. (2022) combined information-gain feature selection with a cost-sensitive AdaBoost classifier and reached 99.8% accuracy, 100% sensitivity and 99.8% specificity; the reduced feature set also improved most of the other classifiers they tested. Dritsas and Trigka (2022) addressed the small-sample problem by applying oversampling together with ensemble learners and obtained accuracy of about 99%. Qureshi et al. (2026) went further by grouping features according to their correlation with the outcome, using recursive feature elimination and LASSO, and comparing SMOTE and SMOTETomek. Their results show that resampling does not always help: KNN reached 94.74% accuracy without SMOTE, whereas C5.0 reached 92.98% with it.

C. Explainable Prediction

Accuracy alone does not convince clinicians, so recent work adds explanations. Ghosh and Khandoker (2024) trained logistic regression, random forest, decision tree, naive Bayes and XGBoost on data from 491 patients (56 with CKD) and used SHAP and LIME to explain the best model; XGBoost obtained an AUC of 0.9689 and 93.29% accuracy. Qureshi et al. (2026) applied the same two explanation methods to a hospital dataset from Pakistan and found that albumin, pus cells and blood in urine were the most influential predictors, which agrees with clinical knowledge about proteinuria and urinary abnormalities. The same clinical setting had been studied earlier by Iftikhar et al. (2025), who evaluated ML and ensemble classifiers for early-stage CKD detection.

D. Progression Prediction

Most studies label a patient as CKD or not CKD. Bai et al. (2022) instead used ML to predict end-stage kidney disease in patients who already have CKD, showing that the same family of methods can support prognosis as well as detection. Table I summarises all ten studies.

E. Datasets, Metrics and Preprocessing

The studies use two kinds of data. The first is the UCI CKD dataset, which holds 400 instances described by 24 attributes and is used in many of the reviewed works (Qureshi et al., 2026). The second is hospital data collected locally, as in the studies from Ethiopia and Pakistan (Debal & Sitote, 2022; Qureshi et al., 2026). Typical preprocessing covers imputation of missing values, for example by the median, standardisation of numeric attributes and encoding of categorical ones (Qureshi et al., 2026). Performance is reported as accuracy, precision, recall, F1-score and the area under the ROC curve. Accuracy alone can mislead when the classes are unequal: in the cohort of Ghosh and Khandoker (2024) only 56 of 491 patients had CKD, so a classifier that always answered 'no CKD' would already be right for about 89% of the patients. Recall, F1-score and AUC are therefore more informative for screening, where missing a true case is costly.

F. Comparative Discussion

Three patterns emerge from the comparison. First, tree-based and boosting models are the most frequent winners: XGBoost was the best model in the studies of Islam et al. (2023) and Ghosh and Khandoker (2024), and a boosted classifier gave the best result of Ebiaredoh-Mienye et al. (2022). Random forest appears among the compared classifiers in several studies (Dubey et al., 2023; Ghosh & Khandoker, 2024; Qin et al., 2020) and remains a strong and easily interpreted baseline. Second, simpler instance-based and linear models can be competitive when the features are well chosen, as the KNN result of Qureshi et al. (2026) shows. Third, the most accurate results are not always the most useful



ones: a model with 99.8% accuracy on a clean public dataset says little about its behaviour on the noisy and incomplete records found in daily practice. Choosing a classifier is therefore a trade-off between accuracy, robustness, interpretability and computing cost, not a search for the highest number.

TABLE I SUMMARY OF REVIEWED STUDIES ON ML-BASED CKD PREDICTION (2020–2026)

Study	Data / Setting	Methods	Key Findings
Qin et al. (2020)	UCI CKD dataset	Missing-value handling; several classifiers and an integrated model	Accuracy above 99%; integrated model gave the best result
Bai et al. (2022)	CKD patient cohort	ML models for end-stage kidney disease prediction	Predicts progression, not only the presence of CKD
Debal & Sitote (2022)	Recent, larger hospital dataset (Ethiopia)	Multiple supervised classifiers	Early prediction on real-world data newer than UCI
Dritsas & Trigka (2022)	Small public CKD dataset	Ensemble and classical ML with oversampling	Accuracy of about 99% on a small sample
Ebiaredoh-Mienye et al. (2022)	UCI CKD dataset	Information-gain feature selection; cost-sensitive AdaBoost	99.8% accuracy, 100% sensitivity, 99.8% specificity
Islam et al. (2023)	UCI CKD dataset	Twelve supervised classifiers	XGBoost best: accuracy 0.983, F1-score 0.98
Dubey et al. (2023)	CKD clinical dataset	GB, DT, KNN, RF, HB and XGB with feature-importance analysis	Identifies the most influential clinical features
Ghosh & Khandoker (2024)	491 patients (56 CKD, 435 non-CKD)	LR, RF, DT, NB, XGBoost with SHAP and LIME	XGBoost: AUC 0.9689, accuracy 93.29%, explained predictions
Iftikhar et al. (2025)	Hospital records, Pakistan	ML and ensemble classifiers for early-stage detection	Baseline on local data, extended by Qureshi et al. (2026)
Qureshi et al. (2026)	382 records, 20 attributes (Pakistan)	RFE, LASSO, correlation groups; SMOTE/SMOTETomek; SHAP and LIME	KNN 94.74% (no SMOTE); C5.0 92.98% (SMOTE); albumin and pus cells key

Two observations follow from the table. First, performance on the UCI data is almost saturated, with accuracy between 98% and 99.8%, whereas studies that use larger or locally collected data report lower values of about 93% to 95%, which suggests that the UCI results are optimistic. Second, only two of the ten studies pair their predictions with SHAP or LIME explanations, and none of the ten describes a complete system for entering records and delivering results to users.

III. RESEARCH GAP

Comparing the studies reveals six gaps that limit clinical use of ML-based CKD prediction.

- 1) **Dependence on Small Public Datasets:** Many studies still rely on the 400-record UCI dataset. Models trained on it reach near-perfect scores that are unlikely to hold on new populations, and the lower accuracy reported on locally collected data (Ghosh & Khandoker, 2024; Qureshi et al., 2026) points in the same direction.
- 2) **Weak External Validation:** Few studies test the model on data from another hospital. Qureshi et al. (2026) themselves note that their data came from a single centre, that no external validation was performed, and that preprocessing carried out before cross-validation can leak information into the test folds. Leakage-safe pipelines are rarely reported.
- 3) **Explanations Do Not Reach the User:** Where SHAP or LIME are used, they are produced as part of the research analysis rather than shown next to each prediction, so the clinician who needs the reasoning does not see it.
- 4) **Missing System View:** None of the ten studies describes a workflow in which patient records, laboratory results, prediction and a printable report are managed together with access control. The model is the end point of the paper rather than the starting point of a usable system.
- 5) **Detection Rather Than Progression:** Binary CKD versus non-CKD classification dominates, and only Bai et al. (2022) address progression to end-stage disease.
- 6) **Inconsistent Data Preparation:** Missing values, class imbalance and feature selection are handled differently in every study, and resampling can even lower accuracy (Qureshi et al., 2026), so there is no agreed and reproducible preprocessing recipe.



These gaps translate into the questions that a practical CKD prediction system should answer: how to obtain stable performance on realistic and imperfect data, how to expose the reasoning behind each prediction, and how to embed the model in a workflow that clinicians can adopt.

TABLE II RESEARCH GAPS AND THE PLANNED RESPONSE OF THE PROPOSED SYSTEM

Research Gap	Planned Response of the Proposed System
1. Small public datasets	Use the UCI data as a baseline and keep the pipeline ready for retraining on local hospital records
2. Weak external validation	Report hold-out and cross-validated performance; external validation is planned (Section VI)
3. Explanations not delivered	Return the prediction with its confidence and risk level; add feature-level explanations (Section VI)
4. Missing system view	One web application with patient, kidney-test and diabetic-test records, prediction history, PDF reports and role-based access
5. Detection only	Binary result with a risk level now; staging and progression are future work

IV. MOTIVATION

CKD is silent in its early stages, but the tests that reveal it are inexpensive and widely available. The motivation for a CKD prediction system is to make better use of these routine results: a model that screens every patient record can flag people who need confirmatory testing long before symptoms appear. This is especially valuable where nephrologists are few and patients travel long distances for care (Debal & Sitote, 2022). Early flagging allows lifestyle changes and blood-pressure and glucose control to start sooner and can postpone the need for dialysis.

There is also an economic reason. Managing established kidney failure places a heavy load on patients, families and health services, whereas CKD found early can often be controlled with medication, diet and regular follow-up. Any tool that moves even a fraction of cases from late to early detection therefore has value well beyond the accuracy figure of the underlying model.

The gaps in Section III shape the design goals. A practical system should (i) use a robust and well-understood classifier; Random Forest suits mixed numeric and categorical laboratory data, tolerates noise and provides feature importance at modest computing cost; (ii) report a confidence value and a risk level instead of a bare label; (iii) keep patient, kidney-test and diabetic-test records in one place so that the prediction can be repeated over time; and (iv) produce a report that can be given to the patient or attached to the case file. Such a system does not replace the clinician; it gives the clinician a consistent second opinion.

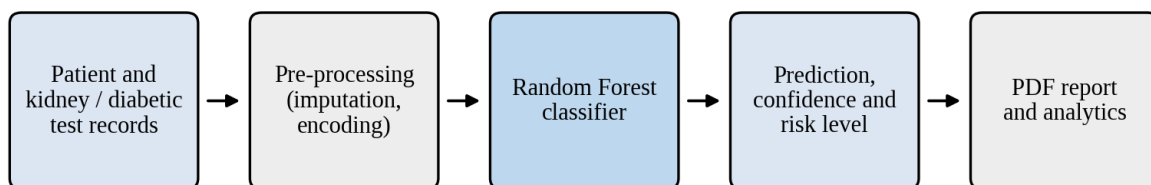


Fig. 1. Workflow of the proposed web-based CKD prediction system

Trust is a further motivation. Many advanced models behave as black boxes, which lowers the confidence of clinicians in them (Qureshi et al., 2026). A system that states its confidence, groups patients by risk level and keeps a record of every prediction makes its output easier to audit. Because the result supports and does not replace clinical judgement, the system is designed as a screening aid whose output is confirmed by standard diagnostic tests.

V. APPLICATIONS

A CKD prediction system of this kind can be used in the following settings.

Primary-care and community screening. Health workers can enter routine test results and obtain a risk flag that decides who is referred for confirmatory tests. This moves detection from the specialist to the first point of contact, where most early cases are seen.



Hospital decision support. In outpatient and nephrology clinics the system can help prioritise patients for further investigation and provide a second opinion that is consistent from one patient to the next. The printed report can be added to the case file for follow-up visits.

Monitoring of high-risk groups. Diabetes and high blood pressure are the two main causes of CKD (Dritsas & Trigka, 2022); storing repeated kidney and diabetic tests allows these patients to be re-assessed over time.

Telemedicine and rural healthcare. A web-based tool lets remote clinics use the same model without local specialists.

Public-health analytics. Aggregated risk levels and prediction trends give health managers a picture of CKD burden in a population.

Integration with health information systems. Because the inputs are routine laboratory values, the model can be connected to electronic health record and laboratory information systems so that risk is calculated as soon as a result is available (Qureshi et al., 2026).

Education and research. The system can serve as a teaching aid for students and as a test bed for comparing classifiers and explanation methods on new datasets.

VI.CONCLUSION AND FUTURE DIRECTIONS

This paper surveyed ten studies published between 2020 and 2026 on ML-based prediction of chronic kidney disease. Ensemble and boosting classifiers, often with feature selection, reach very high accuracy on the UCI data, whereas studies on larger or local data report more modest but more realistic figures. Recent work adds SHAP and LIME explanations and extends prediction to disease progression. The main gaps are small and single-source data, limited external validation, explanations that are not delivered to the user, and the lack of complete and usable systems. A web-based system built around a Random Forest classifier, with confidence, risk level and report generation, addresses part of this gap and offers a practical base for further clinical work.

The survey has limits. It covers only ten studies and is not exhaustive. The reported metrics come from different datasets, preprocessing steps and validation schemes and therefore cannot be ranked directly; Table I is meant to show trends and gaps rather than to name a best algorithm.

Future work should proceed in the following directions.

- 1) Multi-Centre Data and External Validation: Train and test on records from several hospitals and populations, followed by prospective clinical evaluation.
- 2) Leakage-Safe Pipelines: Keep imputation, scaling, feature selection and resampling inside the training folds so that reported performance reflects unseen patients.
- 3) Staging and Progression: Extend binary detection to CKD stage and to the risk of progression to end-stage disease using repeated measurements.
- 4) Integrated Explainability: Show the main contributing features beside every prediction so that the result can be questioned and verified.
- 5) Calibration and Fairness: Check that predicted probabilities are reliable and that performance holds across age, sex and other subgroups.
- 6) Privacy-Preserving Learning and Integration: Explore federated learning across hospitals and connection to electronic health record and laboratory systems and to mobile devices.

REFERENCES

- [1] Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific Reports*, 12, Article 8377. <https://doi.org/10.1038/s41598-022-12316-z>
- [2] Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, 9, Article 109. <https://doi.org/10.1186/s40537-022-00657-5>
- [3] Dritsas, E., & Trigka, M. (2022). Machine learning techniques for chronic kidney disease risk prediction. *Big Data and Cognitive Computing*, 6(3), Article 98. <https://doi.org/10.3390/bdcc6030098>
- [4] Dubey, Y., Mange, P., Barapatre, Y., Sable, B., Palsodkar, P., & Umate, R. (2023). Unlocking precision medicine for prognosis of chronic kidney disease using machine learning. *Diagnostics*, 13(19), Article 3151. <https://doi.org/10.3390/diagnostics13193151>
- [5] Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. *Bioengineering*, 9(8), Article 350. <https://doi.org/10.3390/bioengineering9080350>



- [6] Ghosh, S. K., & Khandoker, A. H. (2024). Investigation on explainable machine learning models to predict chronic kidney diseases. *Scientific Reports*, 14, Article 3687. <https://doi.org/10.1038/s41598-024-54375-4>
- [7] Iftikhar, H., Hashem, A. F., Qureshi, M., & Rodrigues, P. C. (2025). Clinical application of machine learning models for early-stage chronic kidney disease detection. *Diagnostics*, 15(20), Article 2610. <https://doi.org/10.3390/diagnostics15202610>
- [8] Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics*, 14, Article 100189. <https://doi.org/10.1016/j.jpi.2023.100189>
- [9] Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2020). A machine learning methodology for diagnosing chronic kidney disease. *IEEE Access*, 8, 20991–21002. <https://doi.org/10.1109/ACCESS.2019.2963053>
- [10] Qureshi, M., Azad, A., Iftikhar, H., & Rodrigues, P. C. (2026). Toward explainable precision nephrology: Machine learning-based chronic kidney disease prediction. *Biomedicines*, 14(7), Article 1459. <https://doi.org/10.3390/biomedicines14071459>