A Survey Paper on an Intelligent Explainable AI Framework for Early Prediction of Chronic Kidney Disease and Cardiovascular Disease

D.Radhika[1], S.Nivenitha sri[2]
[1]*Head of the Department, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India*
[2]*PG Scholar, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamilnadu, India*

***Abstract*—Chronic Kidney Disease (CKD) and cardiovascular disease (CVD) are leading causes of death and long-term disability worldwide, and the two conditions frequently coexist and worsen one another. Machine learning (ML) models can detect both diseases early from routine clinical data, but most high-performing models behave as black boxes, which limits clinician trust and hinders adoption. Explainable artificial intelligence (XAI) addresses this gap by showing which patient attributes drive a prediction. This paper surveys recent work on explainable ML and deep learning for early prediction of CKD and CVD, with attention to SHAP, LIME, counterfactual explanations and rule-based explanations. We compare the reviewed studies by data, models, explainability technique and reported performance, and we identify recurring weaknesses: small and single-source datasets, a lack of external validation, a scarcity of joint cardiorenal models, and limited evaluation of explanation quality. Drawing on these findings, we outline a conceptual framework that combines multi-source data, imbalance-aware learning, joint CKD-CVD risk modelling, multi-level explanation and clinical validation, and we suggest directions for future research.**

***Index Terms*—Chronic Kidney Disease, Cardiovascular Disease, Explainable AI, SHAP, LIME, Early Prediction, Clinical Decision Support, Machine Learning.**

I. INTRODUCTION

Cardiovascular disease and chronic kidney disease are two of the largest and most costly chronic conditions. The Global Burden of Disease 2019 study identified CVD as the leading cause of death worldwide, with about 18.6 million deaths in 2019 [1]. CKD affected roughly 697.5 million people worldwide in 2017, around 9% of the global population, and caused about 1.2 million deaths that year [2]. Both diseases are often silent in their early stages, so many patients are diagnosed only after substantial and sometimes irreversible damage has occurred.

The two diseases are closely linked. Reduced kidney function is an independent risk factor for cardiovascular events, and people with CKD are more likely to die of CVD than to reach kidney failure [3]. In 2023 the American Heart Association formalised this interplay as cardiovascular-kidney-metabolic (CKM) syndrome, which links metabolic risk factors, CKD and CVD in a staged framework and calls for prediction tools that capture these shared exposures and outcomes [4]. Screening for one condition therefore carries information about the other, which makes a joint approach to early prediction attractive.

A large body of work has applied machine learning to these conditions, and ensemble and deep learning methods frequently report excellent results on standard datasets. Nevertheless, strong predictive performance is not sufficient for clinical adoption. Clinicians need to know why a model flags a patient, whether the reasoning agrees with clinical knowledge, and when the model should not be trusted. Rudin argues that for high-stakes decisions, opaque models should be avoided where an interpretable alternative exists, and that post-hoc explanations of black boxes can be misleading [5]. Surveys of explainability in medicine reach a similar conclusion: transparency is a prerequisite for clinical acceptance, and the field needs

methods that are faithful, stable and understandable to clinicians [6], [7].

This paper reviews explainable AI approaches for the early prediction of CKD and CVD. Section II summarises the main XAI techniques, Section III reviews recent studies, Section IV outlines a conceptual framework synthesised from the literature, Section V compares reported performance, Section VI analyses the findings and open challenges, and Section VII concludes.

II. BACKGROUND: EXPLAINABLE AI TECHNIQUES

Explainability methods are commonly grouped into intrinsically interpretable models and post-hoc methods applied to complex models [7]. Four families recur throughout the clinical prediction literature.

LIME (Local Interpretable Model-agnostic Explanations) explains one prediction at a time by fitting a simple surrogate model, such as a sparse linear model, to the black-box model's behaviour in the neighbourhood of that instance [8]. It is model-agnostic and easy to read, but its explanations can vary with the sampling procedure.

SHAP (SHapley Additive exPlanations) assigns each feature a contribution to a prediction using Shapley values from cooperative game theory, and unifies several earlier attribution methods under one framework [9]. Aggregating SHAP values gives global feature importance, while force and waterfall plots give patient-level explanations, which makes SHAP the most common choice in the studies reviewed here.

Counterfactual explanations describe the smallest change to a patient's features that would change the model's output, for example the change in a laboratory value that would move a patient from high risk to low risk [10]. They are action-oriented and map naturally to clinical conversations about modifiable risk.

Rule-based and intrinsically interpretable models, such as decision trees, logistic regression and extracted if-then rules, expose their logic directly. They are often competitive on small tabular clinical datasets, and they serve as a baseline against which black-box gains should be judged [5].

On the modelling side, gradient-boosted trees such as XGBoost [11] and random forests dominate tabular clinical prediction, and tree-based models also allow efficient exact SHAP computation. Class imbalance is usually handled by resampling, for example with SMOTE [12], although oversampling must be applied only to the training split to avoid leakage into the test data.

III. LITERATURE REVIEW

This section reviews recent studies that combine ML or deep learning with explainability for CKD, for CVD, and for settings where the two overlap.

A. Explainable Ai for Chronic Kidney Disease

[13] Ghosh and Khandoker, in Investigation on explainable machine learning models to predict chronic kidney diseases, built a CKD predictor from clinical characteristics using logistic regression, random forest, decision tree, Naive Bayes and XGBoost. The final model was selected based on its accuracy and AUC scores, and its predictions were then interpreted using SHAP and LIME. SHAP force plots visualised individual predictions, and LIME provided a second view of the same cases. The authors argue that these tools help clinicians understand the rationale behind a prediction.

[14] Singamsetty, Ghanta, Biswas and Pradhan, in Enhancing machine learning-based forecasting of chronic renal disease with explainable AI, tuned several models with GridSearchCV and reported 99.07% accuracy for random forest, extra trees, L2-penalised logistic regression and an artificial neural network. Logistic regression with an L2 penalty was reported as the most consistent performer. LIME and SHAP were used to expose key features, and the authors also released a web application for real-time prediction. Accuracy this close to ceiling on a small benchmark should be read with caution, since it suggests that the task is easy for the data rather than that the model is clinically robust.

[15] Jayarathna and Hewapathirana, in Explainable deep learning for chronic kidney disease prediction, propose an explainable deep-learning model for predicting CKD progression from historical patient data. SHAP and LIME are intended to give transparent insight into risk factors. This is a research proposal presented at a symposium rather than a completed evaluation, but it reflects the field's move from detecting existing CKD toward predicting progression over time.

[16] Nguycharoen, in Explainable Machine Learning System for Predicting Chronic Kidney Disease in

High-Risk Cardiovascular Patients, is notable because it targets patients who already have or are at high risk of CVD. The random forest model reached a sensitivity of 88.2%, chosen deliberately because screening favours few false negatives. The explainability framework goes beyond feature importance to include global and local interpretation, counterfactual explanations, scoped rules, bias inspection and safety assessment. Diabetes and ACEI/ARB medication use and initial eGFR were the leading predictors, and the extracted rules were reported to agree with the medical literature. The author also noted some bias in the relationship between initial eGFR and predictions, which is a useful reminder that explanation can expose model problems as well as justify predictions.

B. Explainable Ai for Cardiovascular Disease

[17] Wang, Xue, Zhang, Wong and Liu, in Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework, combined five public coronary artery disease datasets and used the AutoGluon AutoML framework to build an ensemble. The ensemble outperformed the individual baseline models, and SHAP was used to explain the ensemble's predictions. The study shows that AutoML can lower the barrier to building strong models, while explanation keeps the result interpretable.

[18] Al-Ataby and Attia, in Explainable hybrid models for cardiovascular disease detection and mortality prediction, propose a two-stage framework for heart disease detection and for mortality prediction in heart failure patients. Logistic regression, random forest and gradient boosting were compared, and a stacked ensemble was added in the second stage. Gradient boosting achieved an ROC AUC of 96% and accuracy of 89% on internal testing, and external validation gave an ROC AUC of 94%. SHAP supplied global rankings and patient-level attributions, and two interactive web applications displayed SHAP visualisations in real time. External validation is the strongest feature of this study and is rare in the reviewed literature.

[19] Andritsch and Dike, in XAI-Driven Explainability for Cardiovascular Diseases Prediction, compare explanation techniques on the Framingham Heart Study dataset across conventional models (logistic regression, SVM, decision tree, random forest) and deep tabular architectures (AutoInt, FT-Transformer and Category Embedding). SHAP and LIME are used to assess interpretability and feature relevance. By comparing explainers across model families, the study addresses a question many applied papers ignore, namely whether explanations stay consistent when the model changes.

C. Toward Joint Cardiorenal Prediction

Most published work treats CKD and CVD as separate problems with separate datasets. The study by Nguycharoen [16] is one of the few to connect them directly, by predicting CKD in cardiovascular patients. Clinically, the case for joint modelling is strong: CKM staging was introduced precisely because the two diseases and their metabolic drivers share risk factors, such as hypertension, diabetes and dyslipidaemia, and because prediction algorithms are needed that cover both kidney and cardiovascular outcomes [4]. A joint model that shares features and explanations across both diseases could surface common drivers, for instance serum creatinine, haemoglobin and diabetes status, and could help clinicians prioritise interventions that benefit both organs. We found few published frameworks that predict both conditions together and also explain them, and we regard this as the most significant gap in the area.

IV. CONCEPTUAL FRAMEWORK FOR EXPLAINABLE CKD-CVD PREDICTION

Synthesising the strengths of the reviewed studies, we outline a conceptual framework for an intelligent, explainable early-prediction system. This is a proposal derived from the survey and not an implemented or validated system.

1. Data layer. Electronic health records and laboratory results, including eGFR, serum creatinine, albuminuria, blood pressure, lipid profile, HbA1c, haemoglobin, age, sex, medications and comorbidities. Public benchmarks such as the UCI Chronic Kidney Disease and UCI Heart Disease datasets [20], [21] are useful for development but are small, and should be complemented by larger clinical cohorts.
2. Preprocessing layer. Imputation of missing values, scaling, outlier handling and class-imbalance treatment such as SMOTE [12],

applied inside cross-validation folds to prevent leakage.

3. Feature selection layer. Filter and embedded methods to retain clinically meaningful variables and to identify features shared between the kidney and cardiovascular outcomes.
4. Prediction layer. Interpretable baselines (logistic regression, decision trees) alongside gradient boosting [11], random forests, stacked ensembles and, where data allow, deep tabular models. A multi-task output predicts CKD and CVD risk jointly.
5. Explanation layer. SHAP for global importance and patient-level attributions [9], LIME as a cross-check [8], counterfactual explanations for actionable guidance [10], and extracted rules for clinical review.
6. Validation layer. Calibration, sensitivity at screening thresholds, subgroup and fairness analysis, external validation, and agreement checks between explanations and clinical knowledge.
7. Deployment layer. A clinician-facing dashboard that shows the risk score, the top contributing factors and suggested modifiable targets, with logging for auditability.

V. PERFORMANCE METRICS

Table I compares the reviewed studies. Values are those reported by the authors in their abstracts and are not directly comparable, because datasets, endpoints and validation protocols differ. Entries marked "Not reported" were not stated in the sources consulted.

Table I. Comparison of Reviewed Explainable Ai Studies for Ckd and Cvd

Study	Condition	Data	Models	Explainability	Reported result
Ghosh & Khandoker [13]	CKD	Clinical characteristics	LR, RF, DT, NB, XGBoost	SHAP (incl. force plots), LIME	Best model selected by accuracy and AUC; figures not reported here
Singamsetty et al. [14]	CKD	CKD clinical data	RF, Extra Trees, LR (L2), ANN; GridSearchCV tuning	SHAP, LIME	99.07% accuracy
Jayarathna & Hewapathirana [15]	CKD progression	Historical patient data	Deep learning (proposed)	SHAP, LIME	Not reported (proposal)
Nguycharoen [16]	CKD in high-risk CVD patients	Retrospective cohort, history and lab data	LR, DT, RF, XGBoost, ANN	Global/local interpretation, counterfactuals, scoped rules, bias inspection	RF sensitivity 88.2%
Wang et al. [17]	Coronary artery disease	Five combined public CAD datasets	AutoGluon ensemble vs. baselines	SHAP	Ensemble outperformed baselines; figures not reported here
Al-Ataby & Attia [18]	Heart disease; heart failure mortality	Public heart failure datasets	LR, RF, GB, stacked ensemble	SHAP, web apps	GB: AUC 96%, accuracy 89% (internal); AUC 94% (external)
Andritsch & Dike [19]	CVD	Framingham Heart Study	LR, SVM, DT, RF, AutoInt, FT-Transformer, Category Embedding	SHAP, LIME	Comparative interpretability analysis; figures not reported here

VI. ANALYSIS

Several patterns emerge from the reviewed work.

Tree ensembles and SHAP are the default combination. Random forests, gradient boosting and stacked ensembles dominate, and SHAP is the most frequent explainer, often paired with LIME [13], [14], [18], [19]. This reflects the tabular nature of clinical data and the practicality of tree-based SHAP, and also suggests that deep learning has so far offered limited advantage on small structured datasets.

Reported accuracy is often very high but poorly contextualised. Near-perfect accuracy on a benchmark such as the one in [14] may reflect easily separable data, small sample sizes or leakage, and does not guarantee performance in routine care. Studies that report external validation, such as [18], give far stronger evidence, yet are the minority.

Metric choice should follow the clinical task. For screening, sensitivity and the false-negative rate matter most, as [16] explicitly recognised. Reporting AUC, calibration and precision-recall performance alongside accuracy is particularly important for imbalanced outcomes.

Explanations are rarely evaluated. Most studies display SHAP or LIME plots and take them as evidence of interpretability. Few measure fidelity, stability across resamples or across model types, or clinician comprehension. Ghassemi and colleagues caution that current post-hoc explanations may not deliver the assurance they appear to, and that they should be validated rather than presumed [22]. The cross-model comparison in [19] and the bias and rule-agreement check in [16] are encouraging steps.

Joint CKD-CVD models are scarce. Despite the clinical rationale for CKM-oriented prediction [4], almost all studies address a single disease. Shared predictors such as renal function markers, diabetes and blood pressure suggest that a joint model could be both more efficient and more informative.

Open challenges. (i) Small, single-centre and often decades-old benchmark datasets limit generalisation. (ii) Missing data and class imbalance are handled inconsistently. (iii) Few studies address fairness across sex, age and ethnicity. (iv) Prospective and clinical-impact evaluations are almost absent. (v) Privacy-preserving training across hospitals, for example federated learning, remains largely unexplored for explainable CKD-CVD models. (vi) Explanations should be translated into clinically actionable and regulator-ready documentation.

Future directions. Promising directions include multi-task cardiorenal models, longitudinal models that predict progression and not just current status, counterfactual and causal explanations that identify modifiable risk factors, standardised benchmarks for explanation quality, federated and privacy-preserving learning, and prospective studies with clinician feedback.

VII. CONCLUSION

Explainable AI has become a standard component of ML-based prediction for both chronic kidney disease and cardiovascular disease. The reviewed studies show that tree-based ensembles combined with SHAP and LIME can deliver accurate and readable risk predictions from routine clinical data, and that richer explanation frameworks with counterfactuals, rules and bias checks are beginning to appear. However, the evidence base remains limited by small datasets, scarce external validation, little formal evaluation of explanation quality and very few joint CKD-CVD models. An intelligent framework that integrates imbalance-aware learning, joint cardiorenal risk modelling, multi-level explanation and rigorous clinical validation could help close these gaps and move explainable prediction from research prototypes toward trustworthy early screening in practice.

REFERENCES

[1] G. A. Roth et al., “Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study,” Journal of the American College of Cardiology, vol. 76, no. 25, pp. 2982–3021, 2020, doi: 10.1016/j.jacc.2020.11.010.

[2] GBD Chronic Kidney Disease Collaboration, “Global, regional, and national burden of chronic kidney disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017,” The Lancet, vol. 395, no. 10225, pp. 709–733, 2020, doi: 10.1016/S0140-6736(20)30045-3.

[3] M. J. Sarnak et al., “kidney disease as a risk factor for development of cardiovascular disease: A statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention,” Circulation, vol. 108, no. 17, pp. 2154–2169, 2003, doi: 10.1161/01.CIR.0000095676.90936.80.

[4] C. E. Ndumele et al., “Cardiovascular-kidney-metabolic health: A presidential advisory from the American Heart Association,” Circulation, vol. 148, no. 20, 2023, doi: 10.1161/CIR.0000000000001184.

[5] C. Rudin, "Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead," Nature Machine Intelligence, vol. 1, no. 5, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[6] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793–4813, 2021, doi: 10.1109/TNNLS.2020.3027314.

[7] Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," IEEE Access, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30 (NeurIPS), 2017, pp. 4765–4774.

[10] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," Harvard Journal of Law & Technology, vol. 31, no. 2, pp. 841–887, 2018.

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[13] S. K. Ghosh and A. H. Khandoker, "Investigation on explainable machine learning models to predict chronic kidney diseases," Scientific Reports, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-54375-4.

[14] S. Singamsetty, S. Ghanta, S. Biswas, and A. Pradhan, "Enhancing machine learning-based forecasting of chronic renal disease with explainable AI," PeerJ Computer Science, vol. 10, p. e2291, 2024, doi: 10.7717/peerj-cs.2291.

[15] M. Jayarathna and I. Hewapathirana, "Explainable deep learning for chronic kidney disease prediction," in Applied Data Science & Artificial Intelligence (ADScAI) Symposium, 2025, doi: 10.31705/ADScAI.2025.14.

[16] N. Nguycharoen, "Explainable machine learning system for predicting chronic kidney disease in high-risk cardiovascular patients," arXiv preprint arXiv:2404.11148, 2024.

[17] J. Wang, Q. Xue, C. W. J. Zhang, K. K. L. Wong, and Z. Liu, "Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework," Frontiers in Cardiovascular Medicine, vol. 11, 2024, doi: 10.3389/fcvm.2024.1360548.

[18] Al-Ataby and H. Attia, "Explainable hybrid models for cardiovascular disease detection and mortality prediction," IAES International Journal of Artificial Intelligence, vol. 15, no. 1, pp. 191–212, 2026, doi: 10.11591/ijai. v15.i1. pp191-212.

[19] J. Andritsch and J. Dike, "XAI-driven explainability for cardiovascular diseases prediction," Journal of Informatics and Web Engineering, 2026.

[20] L. Rubini, P. Soundarapandian, and P. Eswaran, "chronic kidney disease," UCI Machine Learning Repository, 2015.

[21] Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "heart disease," UCI Machine Learning Repository, 1988.

[22] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," The Lancet Digital Health, vol. 3, no. 11, pp. e745–e750, 2021, doi: 10.1016/S2589-7500(21)00208-9.