

Chronic Kidney Disease Prediction Based On Machine Learning Algorithms

Hizam Ahmed Al-Saeedi

Problem Statement:

In today's era everyone is trying to be conscious about health although due to workload and busy schedule one gives attention to the health when it shows any symptoms of some kind. But CKD is a disease which doesn't show symptoms at all or in some cases it doesn't show any disease specific symptoms it is hard to predict, detect and prevent such a disease and this could lead to permanent health damage, but machine learning can be a hope in this problem it is best in prediction and analysis. By using data of CKD patients with 14 attributes and 400 records we are going to use various machine learning techniques like Decision Tree, SVM, etc. To build a model with maximum accuracy of predicting whether CKD or not and if yes then its Severity (Tekale et al., 2018).

This paper analyses the suitability of several machine learning algorithms for CKD prediction in consideration of health-related parameters like age, blood pressure, and blood glucose levels. It suggests that decision trees, possibly combined with ensemble techniques like Random Forests, might be suitable for application in clinical environments because they provide insights which are useful for medical practitioners (Kethineni et al., 2025).

In this study, both binary and multi classification for stage prediction have been carried out. The prediction models used include Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT). Analysis of variance and recursive feature elimination using cross validation have been applied for feature selection. Evaluation of the models was done using tenfold cross-validation. The results from the experiments indicated that RF based on recursive feature elimination with cross validation has better performance than SVM and DT (Debal & Sitote, 2022).

This Paper proposes an intelligent classification and prediction model. It uses modified Deep Belief Network (DBN) as classification algorithm to predict the kidney related diseases and the Softmax as activation function and the Categorical Cross-entropy as a loss function (Elkholy et al., 2021).

Research Objectives:

This study aims to develop an efficient and scalable machine learning (ML) framework for the early prediction of Chronic Kidney Disease (CKD). The key objectives include:

Developing a Systematic ML Approach for CKD Classification

Implement various ML techniques to classify CKD effectively.

Ensure the model is scalable and adaptable to real-world healthcare applications.

Optimize computational efficiency to allow deployment in resource-limited settings.

Building ML Models for CKD Prediction Using Patient Health Data

Train predictive models using a combination of clinical, demographic, and biochemical data.

Apply supervised learning algorithms, such as classification models, for disease detection.

Address common data challenges, including missing values, noise, and class imbalance in CKD datasets.

Comparing Different Classification Models

Evaluate traditional ML models such as Logistic Regression, Decision Trees, and Random Forest. Explore advanced models, including XGBoost, LightGBM, and deep learning techniques (ANNs, CNNs, LSTMs).

Assess model performance using various evaluation metrics:

Accuracy – Measures overall model correctness.

Precision & Recall – Evaluates trade-offs between false positives and false negatives.

F1-score & AUC-ROC – Ensures a balanced assessment of model performance.

Identifying the Most Important Clinical Features for CKD Diagnosis

Use feature selection techniques like Recursive Feature Elimination (RFE), Mutual Information, and SHAP analysis.

Rank key biomarkers affecting CKD progression, such as:

Creatinine levels – A primary indicator of kidney function.

Hemoglobin levels – Low hemoglobin may indicate declining kidney function.

Albumin-to-Creatinine Ratio (ACR) – An essential early marker for CKD.

Enhancing Model Interpretability for Healthcare Adoption

Implement explainable AI (XAI) techniques like SHAP and LIME to improve transparency in decision-making. Develop a clinician-friendly interface to facilitate model integration into hospitals and healthcare settings. Validate the model's real-world performance using independent and multi-hospital datasets.

Literature Review

Chronic Kidney Disease (CKD) is a progressive condition that affects kidney function, leading to severe complications if not diagnosed early. Machine Learning (ML) algorithms have become increasingly popular in CKD prediction due to their ability to detect patterns in patient data, automate diagnosis, and support clinical decision-making. However, several challenges still need to be addressed, including data quality, computational complexity, and interpretability in medical settings.

Machine Learning Models for CKD Prediction

Traditional CKD diagnosis relies on laboratory tests, which can be expensive and time-consuming. To improve early detection, ML models such as Decision Trees (DT), Random Forests (RF), and Support Vector Machines (SVM) have been used to classify CKD based on patient data. Recent studies have also explored integrating telemedicine and cloud-based ML models to enable faster, remote diagnosis, improving healthcare accessibility. However, practical challenges remain, including handling complex datasets and optimizing computational resources for real-world implementation.

Comparison of Classification Algorithms

Selecting the best ML model requires careful comparison of different algorithms: Traditional models like Logistic Regression are simple and interpretable but may struggle with complex patterns. Advanced models like XGBoost, LightGBM, and Deep Learning (ANNs, CNNs, LSTMs) achieve higher accuracy but demand more computational power. Model performance is assessed using key metrics: Accuracy – Measures overall prediction correctness. Precision & Recall – Balances false positive and false negative rates. F1-score & AUC-ROC – Provides a holistic evaluation of model effectiveness.

Identifying Key Clinical Features for CKD Diagnosis

Feature selection is crucial to identifying the most influential biomarkers in CKD diagnosis. Techniques like SHAP (Shapley Additive Explanations) and Recursive Feature Elimination (RFE) help determine which clinical features contribute most to model predictions. Key biomarkers include:

Creatinine Levels – A primary indicator of kidney function deterioration.

Hemoglobin Levels – Low levels may indicate worsening kidney function.

Albumin-to-Creatinine Ratio (ACR) – A key early marker for CKD progression.

Improving Model Interpretability for Clinical Adoption

For ML models to be trusted and widely adopted in clinical settings, they must be transparent and explainable. Techniques such as SHAP and LIME help interpret model decisions, making predictions more understandable for doctors and healthcare professionals. Additionally, creating user-friendly interfaces for ML tools can increase acceptance in hospitals and clinics, making AI-driven diagnostics more practical and accessible.

References:

- Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, 9(1). <https://doi.org/10.1186/s40537-022-00657-5>
- Elkholy, S. M. M., Rezk, A., & Saleh, A. A. E. F. (2021). Early Prediction of Chronic Kidney Disease Using Deep Belief Network. *IEEE Access*, 9, 135542–135549. <https://doi.org/10.1109/ACCESS.2021.3114306>
- Kethineni, Likitha., Nithinchandra, Kumar, N., & Sk, S. Sultana. (2025). Chronic Kidney Disease Prediction Based On Machine Learning Algorithms. *ITM Web of Conferences*, 74, 01004. <https://doi.org/10.1051/itmconf/20257401004>
- Tekale, S., Shingavi, P., & Wandhekar, S. (2018). Prediction of Chronic Kidney Disease Using Machine Learning Algorithm. *IJARCCCE*, 7(10), 92–96. <https://doi.org/10.17148/IJARCCCE.2018.71021>