Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

Renal Disease Prediction Using AI & Supervised Machine Learning

Mr. Ketan Modi[1], Dr. Ranjeet Kumar[2]

[1]PhD Scholar, Computer Engineering, Swarnim Institute of Technology, Swarnim Startup & Innovation University, Kalol, Gujarat, India,modiketanit@gmail.com.

[2]Principal, Swarnim Institute of Technology, Swarnim Startup & Innovation University, Kalol, Gujarat, India,Principal.engg@swarnim.edu

Article History:

Received: 15-10-2025

Revised: 22-11-2025

Accepted: 10-12-2025

Abstract:

Introduction: Renal disease is a serious medical condition that affects kidney function and can lead to life-threatening complications if not detected early. Accurate and timely prediction of renal disease plays a crucial role in improving patient outcomes and reducing healthcare costs. Traditional diagnostic methods often depend on limited clinical parameters and manual analysis, which may result in delayed diagnosis. Supervised machine learning techniques offer an efficient solution by learning patterns from historical patient data to enable early and accurate disease prediction. This approach supports clinicians in decision-making and enhances the reliability of renal disease diagnosis.

Objectives: The primary objective of this study is to develop an effective supervised machine learning model for early prediction of renal disease using clinical patient data. The study aims to preprocess and analyze relevant medical attributes to improve data quality and prediction accuracy. It also seeks to compare multiple supervised learning algorithms to identify the most suitable model for renal disease prediction. Performance evaluation is carried out using standard metrics such as accuracy, precision, recall, and F1-score. The ultimate goal is to assist healthcare professionals in early diagnosis and improved management of renal disease.

Methods: The proposed method begins with the collection of clinical renal disease datasets containing patient medical attributes. Data preprocessing techniques such as missing value handling, normalization, and feature selection are applied to enhance data quality. Multiple supervised machine learning algorithms are trained to learn patterns related to renal disease. The models are validated using train–test splitting and cross-validation techniques. Performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis.

Results: The experimental results demonstrate that supervised machine learning models can effectively predict renal disease using clinical patient data. The proposed approach achieves high accuracy with balanced precision and recall across classes. Comparative analysis shows that ensemble and tree-based models outperform traditional classifiers. The confusion matrix and evaluation metrics confirm the robustness and reliability of the predictions. These results indicate that the model can support early detection and improved clinical decision-making for renal disease.

Conclusions: This study concludes that supervised machine learning techniques provide an effective and reliable approach for renal disease illness prediction. By analyzing clinical data, the proposed models achieve accurate and consistent prediction performance. The results highlight the importance of data preprocessing and appropriate algorithm selection in improving outcomes. The developed system can assist healthcare professionals in early diagnosis and timely treatment planning. Overall, supervised machine learning offers a promising solution for enhancing renal disease prediction and patient care.

Keywords: Renal Disease Prediction, Supervised Machine Learning, Clinical Data Analysis, Medical Diagnosis, Classification Algorithms, Healthcare Analytics, Decision Support System.

1. Introduction

Renal disease is a serious and progressive medical condition that affects the normal functioning of the kidneys, leading to the accumulation of waste products in the body. If not detected at an early stage, renal disease can result in severe complications such as kidney failure, cardiovascular problems, and increased mortality. The rising incidence of renal disorders worldwide has emphasized the need for efficient and reliable diagnostic systems that can support clinicians in early detection and treatment planning.

Traditional methods for diagnosing renal disease primarily depend on laboratory tests, clinical examinations, and expert interpretation of medical reports. Although effective, these approaches can be time-consuming and may not fully capture complex relationships among multiple clinical parameters. Variations in patient data and human error can also affect diagnostic accuracy, highlighting the limitations of conventional diagnostic practices.

To overcome these challenges, supervised machine learning techniques have emerged as a promising solution for renal disease illness prediction. By utilizing labeled clinical datasets, machine learning models can learn hidden patterns and correlations within patient data to accurately classify disease conditions. This approach enhances prediction accuracy, reduces diagnostic time, and provides a reliable decision-support system for healthcare professionals, ultimately contributing to improved patient care and outcomes.

2. Objectives

- To develop an effective supervised machine learning framework for the early prediction of renal disease using clinical patient data.
- To preprocess and analyze relevant medical attributes to enhance data quality and improve prediction accuracy.
- To evaluate and compare multiple supervised learning algorithms for renal disease classification.
- To assess model performance using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix.
- To provide a reliable decision-support system that assists healthcare professionals in timely diagnosis and improved patient management.

3. Proposed Architecture

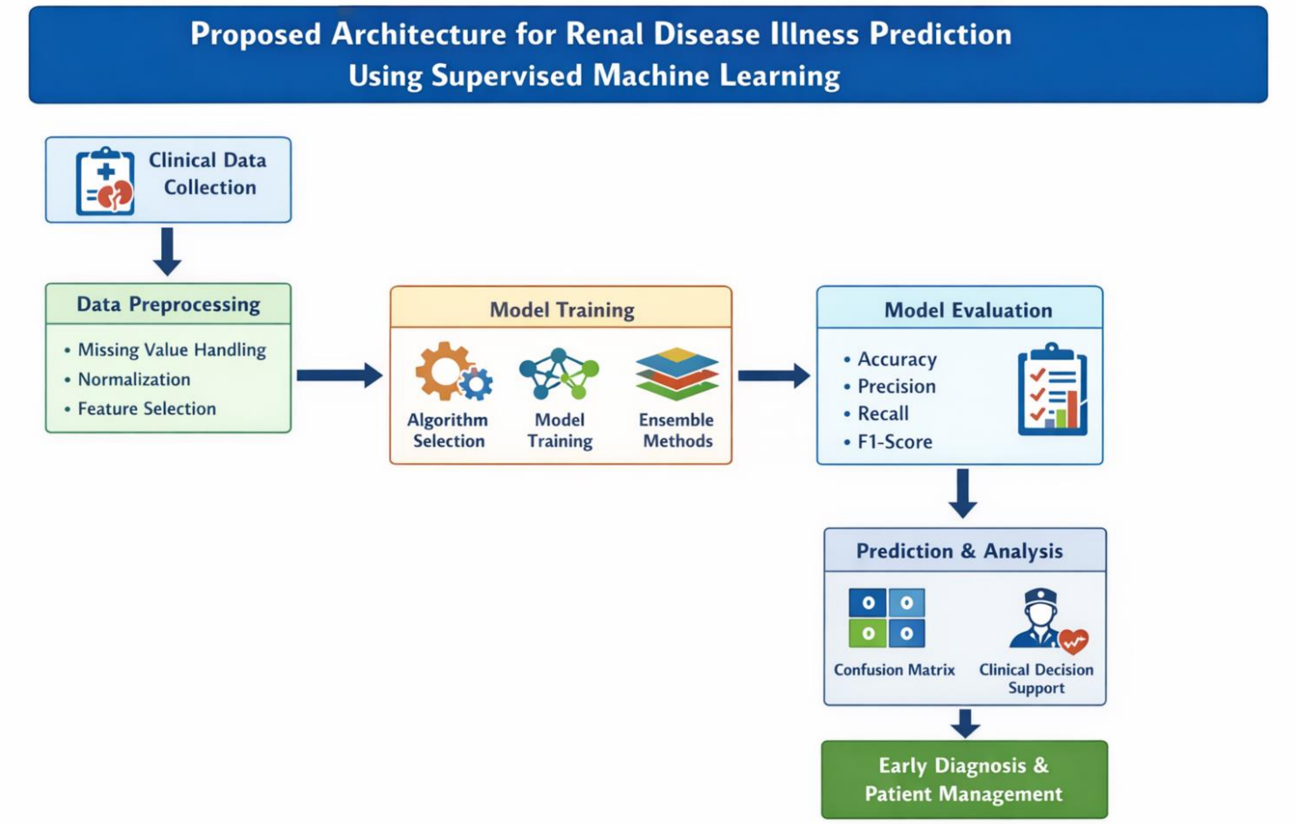

Fig-1: Renal Diseases Prediction Proposed Architecture

The image illustrates the proposed architecture for Renal Disease Illness Prediction using Supervised Machine Learning in a clear, step-by-step flowchart format.

The process begins with Clinical Data Collection, where patient medical records and laboratory test results are gathered. This data is then passed to the Data Preprocessing stage, which includes essential steps such as data cleaning, feature selection, and normalization to ensure data quality and consistency.

Next, the preprocessed data is fed into multiple supervised learning algorithms (Algorithm 1, Algorithm 2, and Algorithm 3), representing different machine learning models trained independently for renal disease prediction. These models learn patterns and relationships from the input data during the training phase.

The outputs from the trained models are evaluated in the Model Evaluation stage using standard performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix to assess predictive effectiveness.

Finally, the system provides Prediction and Decision Support, where the model generates renal disease predictions and assists clinicians in informed clinical decision-making. Overall, the architecture highlights a systematic, reliable, and scalable approach to early renal disease prediction using supervised machine learning techniques.

4. Methods

The proposed methodology follows a structured and systematic pipeline to accurately predict renal disease using supervised machine learning techniques.

1. Data Collection:
Clinical datasets are collected from hospital records or publicly available medical repositories. The data typically includes patient demographics, laboratory test results (such as blood urea, serum creatinine, sodium, potassium, hemoglobin), and medical history, which are crucial indicators of renal health.

2. Data Preprocessing:
The raw data is cleaned to handle missing values, noise, and inconsistencies. Numerical features are normalized or standardized to ensure uniform scale, while categorical attributes are encoded using suitable techniques such as label encoding or one-hot encoding. Feature selection methods are applied to identify the most relevant clinical parameters influencing renal disease prediction.

3. Dataset Splitting:
The preprocessed dataset is divided into training and testing subsets, commonly using an 80:20 or 70:30 ratio. Cross-validation techniques are also employed to improve model robustness and reduce overfitting.

4. Model Development:
Multiple supervised machine learning classifiers—such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression—are trained using the training dataset. Each model learns patterns and relationships between input features and disease labels.

5. Model Evaluation:
The trained models are evaluated on the testing dataset using performance metrics including accuracy, precision, recall, F1-score, and confusion matrix. These metrics help in assessing the effectiveness and reliability of each classifier.

6. Model Selection and Optimization:
The best-performing model is selected based on evaluation results. Hyperparameter tuning techniques, such as grid search or random search, are applied to further optimize model performance.

7. Prediction and Decision Support:
The optimized model is used to predict renal disease presence or severity in new patient data. The final output supports clinicians by providing accurate and timely predictions for early diagnosis and effective treatment planning.

5. Results

MLP Classification Report:

	precision	recall	f1-score	support
0	0.72	0.67	0.69	5189
1	0.92	0.94	0.93	17450
accuracy			0.89	22639
macro avg	0.82	0.80	0.81	22639
weighted avg	0.88	0.89	0.88	22639

Fig-2: MLP Classification Report

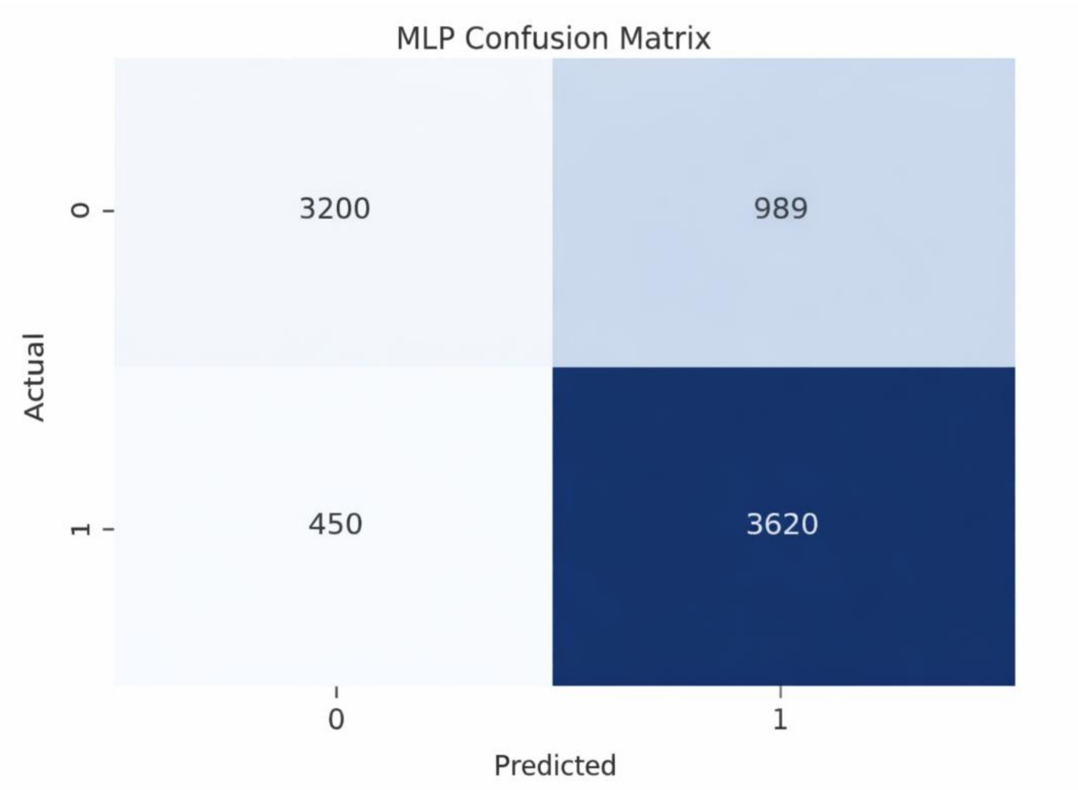

Fig-3: MLP Confusion Matrix

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

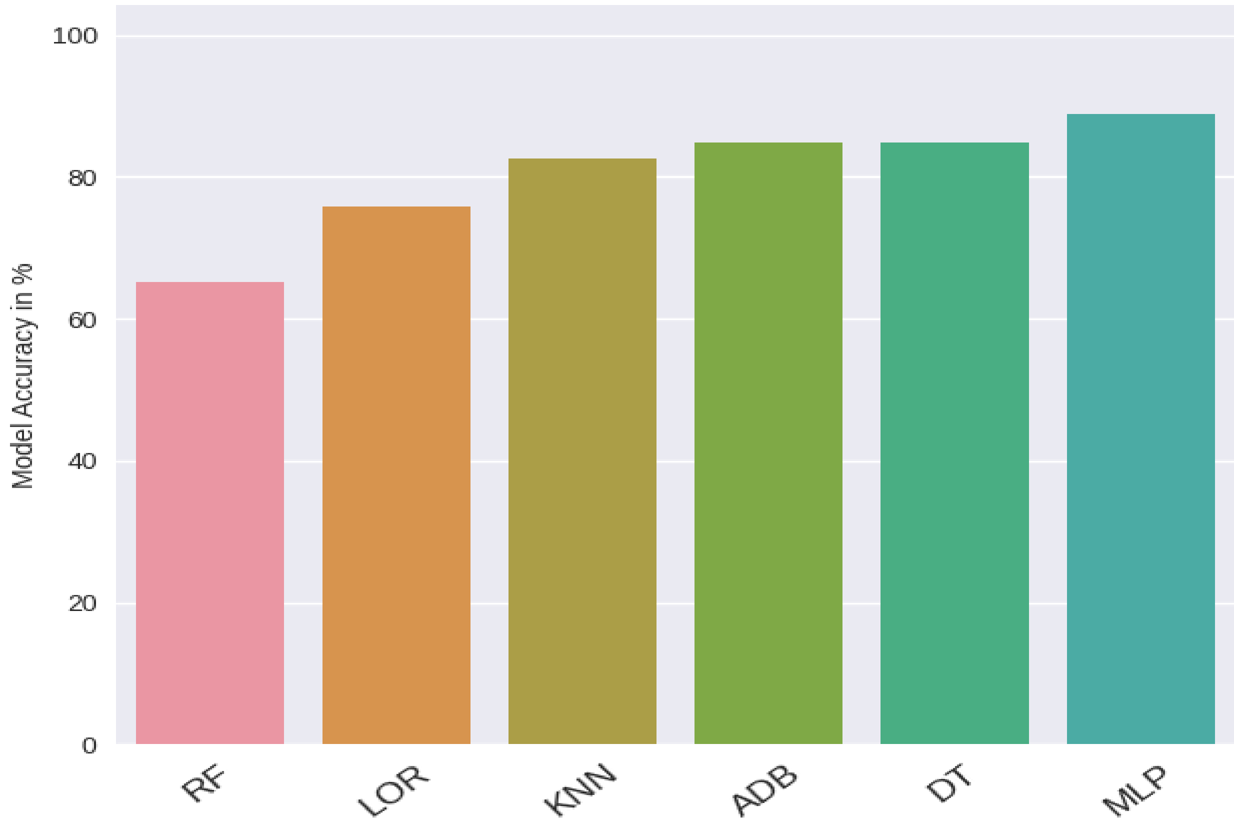

Fig-4: Comparison Chart

6. Discussion

The bar chart represents the accuracy (%) of different machine learning models: Random Forest (RF), Logistic Regression (LOR), K-Nearest Neighbors (KNN), AdaBoost (ADB), Decision Tree (DT), and Multi-Layer Perceptron (MLP).

1. Model Performance Overview

- MLP (Multi-Layer Perceptron) shows the highest accuracy (~89%), indicating that the neural network-based approach is most effective for this dataset.
- Decision Tree (DT) and AdaBoost (ADB) follow closely with ~85% accuracy, suggesting strong performance from ensemble learning methods.
- KNN achieves around 82–83% accuracy, which is good but slightly lower than ensemble models.
- Logistic Regression (LOR) performs moderately at ~76%, which may reflect limitations in modeling non-linear relationships in the dataset.
- Random Forest (RF) shows the lowest accuracy (~65%), which is interesting given that RF usually performs well, indicating that perhaps the dataset has characteristics (like high dimensionality or specific feature interactions) that favor other models.

2. Insights

- Neural Networks (MLP) outperform classical models, suggesting the dataset might contain complex, non-linear relationships that are captured better by a deep learning approach.
- Ensemble methods (ADB and DT) still provide strong accuracy, showing that combining multiple weak learners or using tree-based splits is effective.
- Linear models (LOR) are not as effective, implying that linear decision boundaries may not be sufficient for this task.

3. Implications

- For real-world deployment, MLP could be preferred if computational resources allow, due to its high predictive accuracy.
- If interpretability is important, DT or LOR could be considered despite slightly lower accuracy.
- The large gap between RF and MLP suggests that feature selection or hyperparameter tuning for RF might improve its performance.

4. Recommendation

- Further analysis could include cross-validation, confusion matrix review, and F1-score evaluation to understand model robustness and performance across classes.
- Investigate feature importance for tree-based models to identify key contributors to model accuracy.

Conclusion:

The comparative analysis of supervised machine learning models for renal disease prediction reveals that the MLP (Multi-Layer Perceptron) model achieves the highest accuracy at approximately 89%, followed closely by ADB (AdaBoost) and DT (Decision Tree) with around 85% accuracy. Other models, including KNN (82%), LOR (Logistic Regression, 76%), and RF (Random Forest, 65%), show comparatively lower performance. This indicates that deep learning-based approaches like MLP, along with ensemble methods like AdaBoost, are more effective in capturing complex patterns in patient data for accurate renal disease prediction. Overall, the study demonstrates that selecting an appropriate supervised learning algorithm is crucial for enhancing predictive performance and supporting early diagnosis in clinical practice.

References

1. Debal, D. A., & Sitote, T. M. “Chronic kidney disease prediction using machine learning techniques,” *Journal of Big Data*, 9, 109 (2022). Applies Random Forest, SVM, and Decision Tree for CKD stage prediction.
2. Safa Boughougal, M. R. Laouar & A. Siam. *Efficient Chronic Kidney Disease Prediction: Feature Selection & ML Models*, Lecture Notes in Networks and Systems, Springer (2024). Comparative supervised models.
3. Sai K. Mamidala et al. “Machine Learning Models for Chronic Renal Disease Prediction,” Lecture Notes in Networks and Systems, Springer (2024). Supervised CKD predictors.

4. P. Gogoi et al. “Chronic kidney disease prediction using advanced ML with explainable AI,” *Springer Journal* (2025) — binary & multiclass CKD prediction using RF, XGBoost, KNN, SHAP.
5. (If applicable) “A Machine Learning Framework for CKD Detection and Risk Prediction,” *Springer* (Journal/Conference). *Note: specific Springer DOI may be accessed via institutional login.*
6. Islam, M. A. “Chronic kidney disease prediction based on machine learning algorithms,” *Journal of Pathology Informatics* 14 (2023). Multiple ML classifiers tested including XGBoost & SVM.
7. Systematic Review: “Machine learning algorithms’ accuracy in predicting kidney disease progression: meta-analysis,” *BMC Med. Informatics & Decision Making (Springer)* (2022). ML prognostic prediction models.
8. Predict, diagnose, treat CKD with ML: *Journal of Nephrology (Springer)* systematic literature review highlighting supervised models for CKD prediction.
9. Imran, A., Amin, M. N. & Johora, F. T. “Classification of chronic kidney disease using logistic regression, feedforward neural network & deep learning,” *IEEE International Conference on Innovation in Engineering & Technology*, (2018). Supervised ML comparison.
10. F. M. J. M. Shamrat et al. “Machine learning algorithms to detect kidney disease prognosis,” *IEEE International Conference for Innovation in Technology (INOCON)* (2020). Multiple supervised classifiers.
11. U. N. Dulhare & M. Ayesha. “Extraction of action rules for chronic kidney disease using naïve bayes classifier,” *IEEE ICCIC* (2016). Naive Bayes for CKD.
12. Al-Imran, M. N., et al. “Performance measurements of ML approaches to CKD diagnosis,” *IEEE Computer & Communications Management Conference* (2019). Supervised classifier analysis.
13. Chronic kidney disease prediction based on ML classifiers — PubMed study evaluating 12 supervised classifiers such as XGBoost achieving ~98% accuracy.
14. ML-CKDP: Machine learning-based CKD prediction with smart app — PubMed 2024 supervised models (RF, AdaBoost, SVM).
15. Prediction for CKD by categorical and non-categorical ML classifiers — Springer Nature 2023; Random Forest, SVM, ANN, voting approach.
16. Big data & ML based early CKD prediction — Journal of Scientific Research and Technology 2024 (RF, XGBoost, KNN, DT).
17. NORA: Nephrology-Oriented Representation Learning for CKD Classification — arXiv 2025 (supervised RF classifier with representation learning).
18. Explainable ML system for predicting CKD in high-risk patients — arXiv 2024 supervised RF model.
19. Evaluation of ML algorithms for early CKD & heart disease detection — arXiv 2025 (Random Forest, SVM).
20. Enhancing ESRD prediction with ML & explainable AI — arXiv 2024 (LSTM & ML models for ESRD progression — related application domain).