International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Chronic Kidney Disease Prediction Using ML

Mr.Aashish Janardhan Mallah, Mr.Abhishek Jagdish Mishra, Mr.Pritish Ashok Shetty, Mr.Kapil Deendayal Mewati, Mr.Sandesh Patil

Department of Data Engineering, Universal college of Engineering, University of Mumbai, Maharashtra, India

Email; abhimishra5987@gmail.com , pritishshetty32@gmail.com

Abstract

Chronic Kidney Disease (CKD) is a serious global health problem that often remains undetected until advanced stages, leading to complications such as hypertension, anemia, and kidney failure. Early identification of CKD is important for improving treatment outcomes and reducing health risks. This study proposes a machine learning–based approach for early CKD prediction using clinical health data. The system performs data preprocessing and feature selection before applying two ensemble learning models: Random Forest and AdaBoost. These models are evaluated using standard performance metrics to determine their effectiveness in classifying CKD cases. The model is integrated into a Streamlit-based web application that enables real-time CKD risk assessment. The proposed system offers an accurate, interpretable, and scalable solution that can support early diagnosis and assist healthcare professionals in clinical decision-making.

Keywords: Chronic Kidney Disease (CKD), Machine Learning, Random Forest, AdaBoost, Healthcare Analytics.

1.Introduction

Chronic Kidney Disease (CKD) is a major global health concern characterized by the gradual loss of kidney function over time. The disease often progresses silently, with minimal symptoms in its early stages, which makes timely diagnosis difficult and increases the risk of severe complications. Recent studies, including the Global Burden of Disease report, highlight a significant rise in CKD cases worldwide, affecting nearly 850 million people, particularly in developing regions with limited healthcare resources. To address this challenge, this study proposes a machine learning–based system for early CKD risk prediction using clinical and demographic data such as blood pressure, blood glucose levels, serum creatinine, and hemoglobin. The developed predictive model is integrated into a web-based application to provide an accessible and cost-effective tool that assists healthcare professionals and individuals in early disease detection and preventive healthcare.

1.1 Background

Chronic Kidney Disease (CKD) is a progressive medical condition in which the kidneys gradually lose their ability to filter waste and excess fluids from the blood. If not detected early, CKD can lead to severe complications such as hypertension, anemia, cardiovascular disease, and kidney failure. Early diagnosis plays a critical role in preventing disease progression and improving patient outcomes.

In recent years, machine learning techniques have been widely applied in the healthcare domain to assist in disease diagnosis and prediction. These techniques can analyze large clinical datasets and identify patterns that may not be easily detectable through traditional medical analysis. Several studies have explored the use of machine learning algorithms to improve CKD prediction accuracy and assist healthcare professionals in early diagnosis [2].

1.2 Problem Definition

Traditional methods of diagnosing chronic kidney disease often rely on manual evaluation of laboratory results and clinical observations, which can be time-consuming and

prone to human error. As medical datasets continue to grow, there is a need for automated systems that can efficiently analyze patient data and support clinical decision-making.

Previous studies have applied various machine learning algorithms such as decision trees, support vector machines, neural networks, and gradient boosting techniques for CKD prediction. While these models have demonstrated promising results, many studies focus primarily on algorithm comparison without providing an accessible platform for real-time prediction and analysis [3][4].

algorithms and its deployment through a web-based interface.

The system focuses on analyzing patient clinical attributes to determine the likelihood of CKD. The machine learning models are trained using historical medical datasets and then used to predict disease risk based on new patient input.

The project also includes the development of a user-friendly Streamlit web application where users can input health parameters and receive instant prediction results.

Several studies have explored the use of machine learning techniques for predicting chronic diseases using clinical datasets

Author	Method Used	Accuracy	Research Gap

1.3 Objectives

The main objectives of this project are:

1. To develop a machine learning model for predicting Chronic Kidney Disease using clinical health data.

2. Literature Review

2. To preprocess and analyze CKD datasets for effective model training.
3. To implement ensemble learning algorithms such as Random Forest and AdaBoost.
4. To evaluate model performance using standard classification metrics.
5. To develop a Streamlit-based web application that allows users to obtain real-time CKD prediction results.

1.4 Scope of the Project

The scope of this project includes the development of a CKD prediction system using machine learning

ML-CKDP: Machine Learning-based Chronic Kidney Disease Prediction with Smart Web Application	Multiple machine learning algorithms for CKD prediction integrated with a web-based system	High prediction accuracy reported using ensemble methods	The system evaluates multiple models but does not specifically focus on comparing Random Forest and AdaBoost performance for CKD prediction.
Hafeez Ilyas et al. (2021)	Decision Tree and Random Forest algorithms for CKD detection	Random Forest achieved accuracy close to 99%	The study focused mainly on traditional classifiers and did not implement a deployable prediction interface for practical usage.
M. A. Islam et al. (2023)	XGBoost algorithm with Principal Component Analysis (PCA) feature selection	Around 99% classification accuracy	The research concentrated on feature selection techniques but did not implement a real-time application for patient prediction.
Habiba Urooj et al. (2023)	Multiple machine learning algorithms with feature selection methods	Accuracy reported above 97%	The work compared several models but did not focus on ensemble comparison between Random Forest and AdaBoost.
T. Sruthi et al. (2023)	Machine learning algorithms applied to CKD dataset for classification	Accuracy above 95%	The study focuses on algorithm comparison but does not provide a system that allows real-time prediction through a user interface.

3. Proposed System

3.1 System Overview

The proposed system is a machinelearning-based Chronic Kidney Disease prediction system designed to analyze patient clinical data and determine the likelihood of CKD. The system processes medical attributes such as blood pressure, serum creatinine level, blood glucose level, hemoglobin concentration, and other relevant parameters. These attributes are used as input features for the machine learning models.

The system uses two ensemble learning algorithms:

- Random Forest
- AdaBoost

Both models are trained on a CKD dataset to classify patients as CKD or non-CKD. After training, the best performing model is deployed into a Streamlit web application.

The web application allows users to input patient health information and obtain real-time CKD prediction results.

3.2 Dataset Description

The dataset used in this study is the Chronic Kidney Disease (CKD) dataset obtained from the UCI Machine Learning Repository. This dataset contains 400 patient records with 24 clinical parameters related to kidney health. The dataset is widely used in medical data analysis and machine learning research for CKD prediction tasks.

The dataset includes various demographic, blood test, and urine test attributes that are commonly used by healthcare professionals to diagnose kidney-related disorders. Examples of these parameters include age, blood pressure, specific gravity, albumin, blood glucose level, serum creatinine, hemoglobin level, white blood cell count, and red blood cell count.

These attributes are used as predictive features for identifying whether a patient is likely to suffer from Chronic Kidney Disease. The dataset contains both numerical features (such as blood pressure and creatinine levels) and categorical features (such as presence or absence of certain medical condition.

3.3 Data Preprocessing

Data preprocessing is an essential step in machine learning model development. It ensures that the dataset is clean, consistent, and suitable for model training.

The preprocessing steps performed in this project include:

- Handling missing values in the dataset
- Converting categorical data into numerical format
- Normalizing numerical features
- Splitting the dataset into training and testing sets

These steps improve the quality of the dataset and enhance model performance.

3.4 Machine Learning Algorithms Used

Random Forest

Random Forest is an ensemble learning algorithm that builds multiple decision trees and combines their outputs to generate a final prediction.

AdaBoost (Adaptive Boosting)

AdaBoost is another ensemble learning technique that improves classification accuracy by combining multiple weak learners.

During training, AdaBoost assigns weights to each training sample. Misclassified samples receive higher weights so that the algorithm focuses more on difficult cases in subsequent iterations.

In this project, AdaBoost is implemented alongside Random Forest to evaluate its performance in CKD prediction.

4. Result and Analysis

The performance of the proposed Chronic Kidney Disease prediction system was evaluated using two ensemble machine learning algorithms: Random Forest and AdaBoost. The models were trained using the processed CKD dataset, and their performance was evaluated using standard classification metrics including Accuracy, Precision, Recall, F1 Score, and Confusion Matrix.

4.1 Random Forest Results

The Random Forest classifier produced the following evaluation results:

Metric	Value
Accuracy	1.00
Precision	1.00
Recall	1.00

F1 Score	1.00

The confusion matrix obtained from the Random Forest model is shown below:

Actual / Predicted	Not CKD	CKD
Not CKD	30	0
CKD	0	50

The model achieved **training accuracy of 1.0 and test accuracy of 1.0**, and all samples in the test set were classified correctly according to the confusion matrix.

4.2 AdaBoost Results

The AdaBoost classifier produced the following evaluation results:

Metric	Value
Accuracy	0.9875
Precision	1.00
Recall	0.98
F1 Score	0.9899

The confusion matrix obtained from the AdaBoost model is shown below:

Actual / Predicted	Not CKD	CKD
Not CKD	30	0
CKD	1	49

The AdaBoost model achieved **training accuracy of 1.0 and test accuracy of 0.9875**. Based on the confusion matrix, one CKD case was misclassified while the remaining samples were classified correctly.

4.3 Streamlit interface

A Streamlit-based web application that allows users to obtain CKD prediction results.

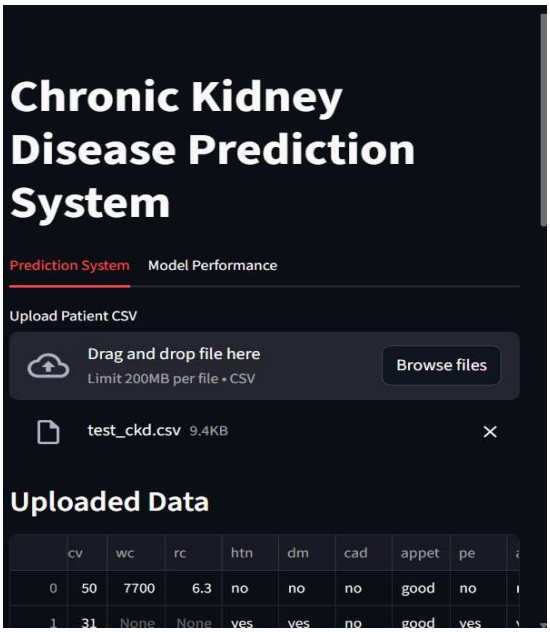

Fig. 4.1 interface

Fig. 4.2 output

4.4 Discussion

The experimental results show that both Random Forest and AdaBoost classifiers were able to classify the CKD dataset with high accuracy. The Random Forest model correctly classified all samples in the test dataset according to the obtained confusion matrix.

The AdaBoost model also performed similarly, with only one misclassified CKD instance based on the confusion matrix.

5. Conclusion

The experimental results showed that the Random Forest model achieved 100% accuracy on the test dataset, while the AdaBoost model achieved 98.75% accuracy.

These results indicate the classification performance of the two models on the dataset used in this study. The trained models were integrated into a Streamlit-based web application, allowing users to upload patient data and obtain CKD prediction results.

5.1 Future Enhancements

Future improvements may include:

- Integration with hospital electronic health records
- Development of a mobile application
- Implementation of additional ensemble techniques
- Inclusion of larger healthcare datasets

5.2 Final Remarks

Machine learning has significant potential in transforming healthcare diagnostics. The proposed CKD prediction system demonstrates how artificial intelligence can assist in early disease detection

REFERENCES

[1] ML-CKDP: Machine Learning-based Chronic Kidney Disease Prediction with Smart Web Application.

[2] Chronic Kidney Disease Diagnosis Using Decision Tree Algorithms,

Hafeez Ilyas et al., BMC Nephrology, vol. 22, 2021.

[3] M. A. Islam et al.,

"Chronic Kidney Disease Prediction Using XGBoost with PCA Feature Selection," 2023.

[4] Habiba Urooj et al.,

"Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms with Feature Selection Techniques," 2023.

[5] T. Sruthi et al.,

"Detection of Kidney Disease Using Machine Learning," International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2023.

