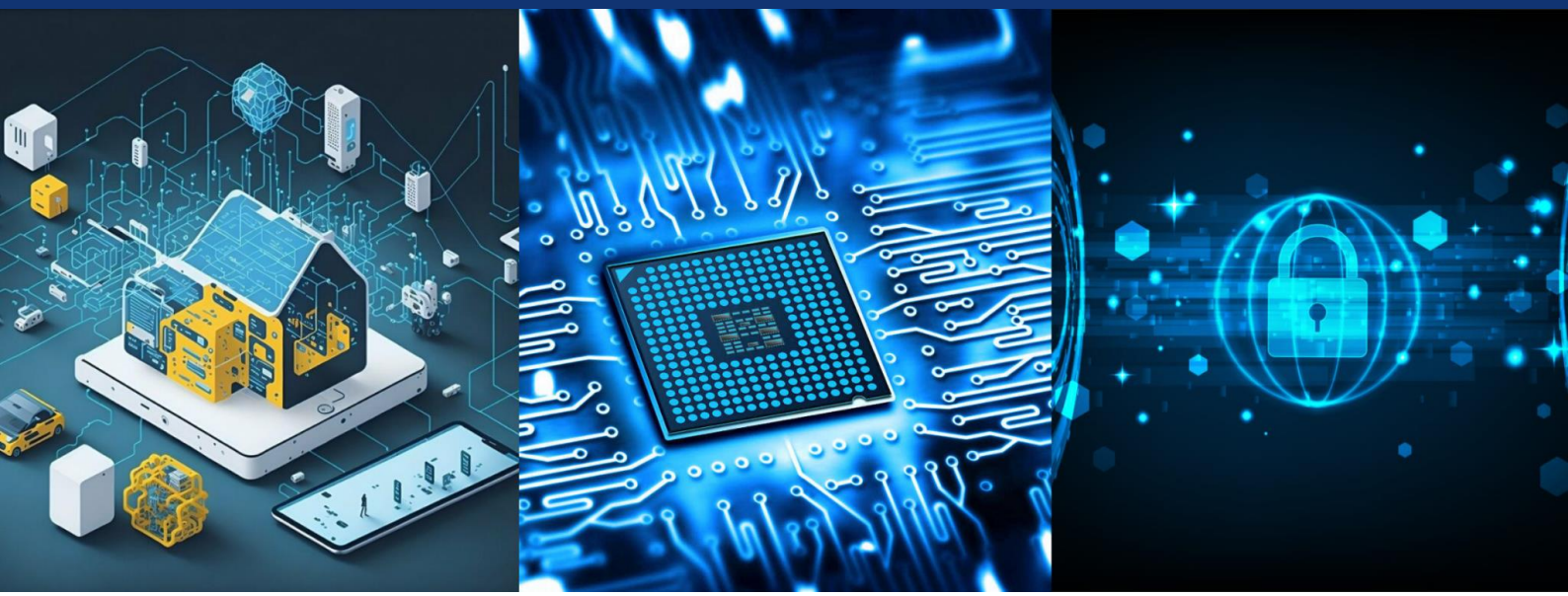
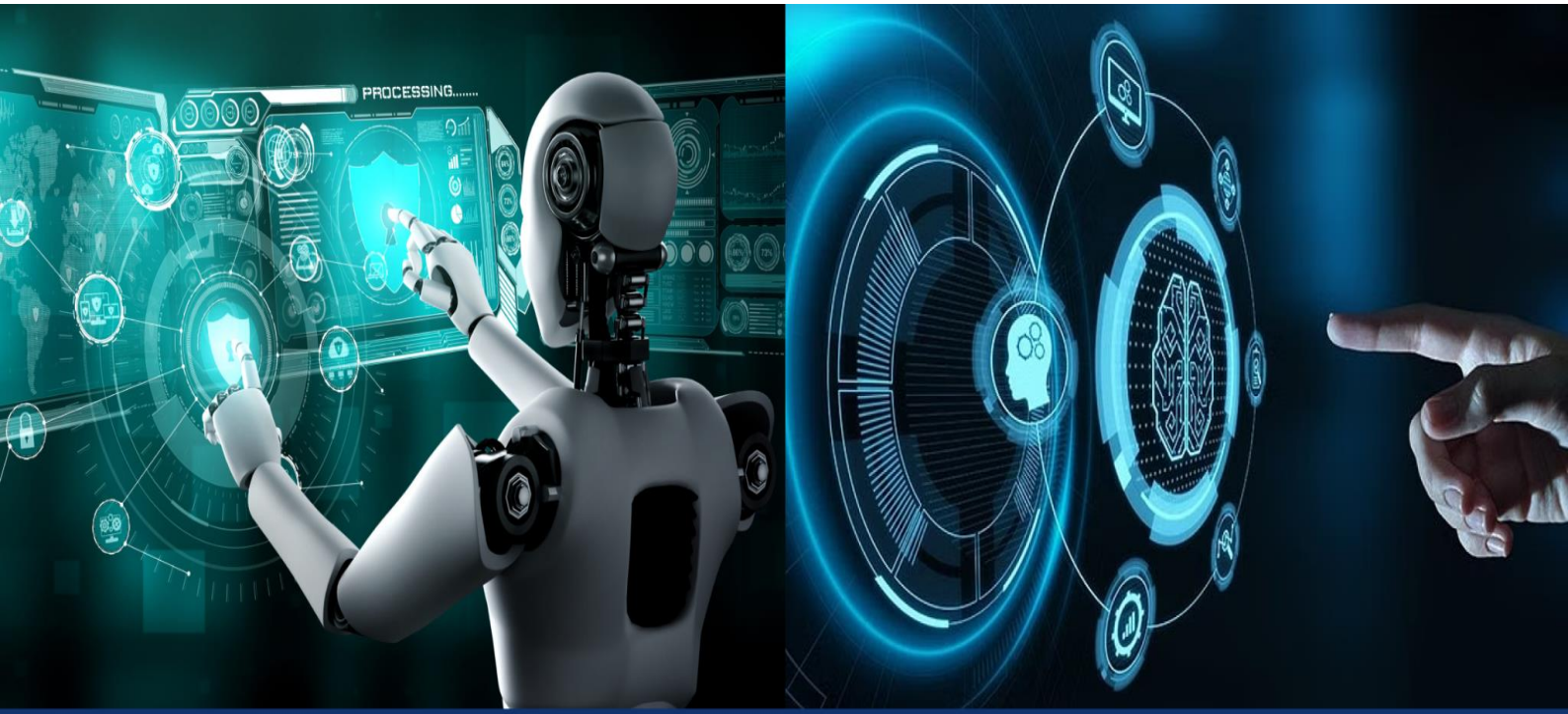
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771 **Volume 13, Issue 4, April 2025**

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304194

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Chronic Kidney Disease Detection Using Deep Learning

K. Mahesh, J. Pranavi, E. Pavan Sai, B.Vijay Kumar, CH. Bhuvana Teja, B Rushi Sree

Assistant Professor, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

ABSTRACT: This study focuses on the early detection of Chronic Kidney Disease (CKD) using machine learning techniques. CKD often develops without symptoms until it becomes severe, making early diagnosis critical. A dataset of 400 patient records with 24 clinical and biochemical features was used to train and evaluate three models: Convolutional Neural Networks (CNNs), Random Forest, and AdaBoost. The data underwent preprocessing including missing value imputation, normalization, and stratified k-fold cross-validation. Performance was measured using accuracy, precision, recall, F1-score, and AUC metrics. While CNNs were adapted for tabular data, ensemble models like Random Forest and AdaBoost showed strong performance, with AdaBoost emerging as the most accurate and effective. The results highlight the potential of machine learning—particularly boosting methods—in providing reliable, early diagnosis support for CKD. With further validation, these models could enhance clinical decision support systems.

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a major global health issue, affecting millions of individuals and often remaining undetected until it reaches an advanced stage. Due to its multifactorial causes—ranging from medical and genetic to environmental factors—early detection of CKD is challenging using traditional diagnostic methods. Machine Learning (ML), a powerful branch of Artificial Intelligence, has demonstrated significant potential in healthcare by identifying complex patterns in large datasets that might be overlooked by human clinicians. Prior research has shown that techniques such as proper data preprocessing, feature selection, and hybrid models can significantly improve diagnostic accuracy. Motivated by the alarming statistic from the World Health Organization that CKD affects nearly 850 million people globally, this study explores the application of ML to improve early diagnosis, reduce complications, and enhance treatment outcomes. Traditional tools often lack the sensitivity needed for early-stage detection, emphasizing the need for intelligent, data-driven solutions. In this context, we evaluate and compare several ML models—Convolutional Neural Networks (CNNs), AdaBoost, CatBoost, and Random Forest (RF)—on a clinically relevant dataset comprising 400 patient records with 24 features. The models are assessed using performance metrics such as accuracy, precision, recall, and F1-score, with stratified k-fold cross-validation employed to ensure generalizability. The core objective of this work is to determine the most effective model for reliable CKD prediction and provide practical insights for implementing these technologies in clinical settings. Ultimately, this research aims to contribute toward the development of smarter, faster, and more dependable diagnostic tools that enable healthcare providers to intervene earlier and improve global kidney health outcomes.

II. RELATED WORK

Numerous studies have explored machine learning and deep learning for early detection of Chronic Kidney Disease (CKD). Singh et al. [1] developed a deep neural network model for early CKD diagnosis, while *KidneyNet* [2], a CNN-based system, was used for CT scan analysis. Bhattacharjee et al. [3] designed a multi-class DL framework for both CKD and lung cancer detection. Hybrid models have also been proposed—Baseera et al. [4] integrated ANFIS with DCNN, and Reddy et al. [5] combined CNN with SVM.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304194

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Other approaches include 3D CNNs for MRI-based CKD severity classification [6, 12], ultrasound-based DL methods [7, 14, 20], and multimodal systems using retinal images and urine data [10]. TRACE [11], a transformer-based model, has shown promise in capturing complex patient data features. Additional works focus on renal segmentation [13, 15], and frameworks for classifying glomerular hypercellularity [16] and kidney abnormalities [17].
Overall, these studies demonstrate that CNNs, ensemble models, and hybrid techniques are highly effective for CKD prediction and diagnosis, especially when combined with medical imaging or structured clinical data.

III. PROPOSED ALGORITHM

Step-1: Select the Dataset
To begin the CKD detection process, selecting a suitable dataset is crucial. The dataset should ideally include a wide range of relevant features that can influence kidney health, such as patient demographics (age, gender, etc.), medical history (comorbidities like diabetes, hypertension), and laboratory test results (e.g., serum creatinine levels, blood urea nitrogen (BUN), urinalysis). It should also provide clear labels (CKD vs. non-CKD) to facilitate supervised learning. Additionally, the dataset must be assessed for its balance—whether it is balanced or imbalanced, as this affects model performance and evaluation. An imbalanced dataset (with many more non-CKD than CKD cases) could lead to biased predictions

Step-2: Data Preprocessing
After selecting the dataset, the next step is data preprocessing, which ensures that the data is in the right format and ready for machine learning models. One common challenge is handling missing values, especially for critical features like age, blood pressure, or creatinine levels. Missing data can be dealt with by either imputing the missing values using techniques like mean, median, or mode imputation or by removing rows or columns with excessive missing data. However, the imputation method chosen should be suited to the nature of the missing data. For example, numerical features might use the mean or median for imputation, while categorical data may rely on the mode or predictive imputation. Next, normalization or scaling of features is crucial, especially for continuous variables like blood pressure, serum creatinine, or eGFR, as they may have very different scales.

Step-3: ML Models Applied
In this step, several machine learning models are applied to the preprocessed dataset to predict CKD.
Commonly used models include Convolutional Neural Networks (CNN), Random Forest, AdaBoost, and CatBoost. While CNNs are typically applied to image data, they have also been adapted for structured data by reshaping the dataset into a matrix that can be processed as image-like data. CNNs can automatically detect complex hierarchical patterns in the data, making them suitable for CKD detection when feature interactions are non-linear and complex. On the other hand, Random Forest, an ensemble learning technique, works by constructing multiple decision trees during training and then outputs the majority vote for classification tasks. Random Forest handles large datasets well and helps reduce overfitting, making it ideal for handling a wide variety of features and data irregularities. AdaBoost (Adaptive Boosting) is another ensemble method that combines the predictions of weak learners (often decision trees) to create a strong classifier, focusing on areas where previous models made errors. It tends to perform well when there is a need for handling difficult, misclassified data points

Step-4: Training & Evaluation
Once the models are selected, training and evaluation follow. Training is done using stratified k-fold crossvalidation, a technique that splits the dataset into k subsets (folds), while maintaining the original class distribution (CKD vs. non-CKD) within each fold. This ensures that each fold represents the overall class distribution and helps prevent bias in training. In each iteration, one fold is used for testing, and the remaining k-1 folds are used for training the model. This process is repeated for all k folds to ensure reliable and stable performance across different data splits. After training, the model's performance is evaluated using various metrics. Accuracy is one common measure, but in the case of CKD detection, it is not always sufficient, especially if the dataset is imbalanced. In such cases, recall (or sensitivity), which measures the proportion of actual CKD cases correctly identified, becomes more important.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304194

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. PSEUDO CODE

Step 1: BEGIN
Dataset Selection: Load dataset containing features such as:
- Age, Gender, Blood Pressure, Serum Creatinine, BUN, Urinalysis, etc.

Check if the dataset is balanced (CKD vs. non-CKD)
IF dataset is imbalanced THEN
Apply resampling techniques (e.g., SMOTE, undersampling) [Optional]
Step 2: **Data Preprocessing**
FOR each feature in dataset DO
IF feature has missing values THEN
IF feature is numerical THEN
Impute using mean or median
ELSE IF feature is categorical THEN
Impute using mode or predictive imputation
ENDIF
ENDFOR

Normalize numerical features (e.g., Min-Max Scaling)
Encode categorical features using Label Encoding / One-Hot Encoding
Step 3: **Model Selection and Training**
Define models: CNN, Random Forest, AdaBoost, CatBoost

FOR each model in defined models DO
Apply Stratified K-Fold Cross Validation (e.g., k = 5)
FOR each fold:
Split dataset into training and validation sets
Train model on training set
Validate model on validation set
Record evaluation metrics (Accuracy, Precision, Recall, F1-Score)
ENDFOR
Compute average performance across all folds
ENDFOR
Step 4: **Evaluation**
Compare model performances
Select the model with best performance (preferably high Recall & F1-score)

IF selected model meets performance threshold THEN
Deploy the model for CKD prediction
ELSE
Tune hyperparameters or use ensemble techniques
ENDIF

Step 5: END

V. SIMULATION RESULTS

The simulation results for the Chronic Kidney Disease (CKD) detection system were obtained by evaluating various machine learning models on a dataset containing relevant features related to CKD. The models tested include Decision Trees, Random Forest, Support Vector Machines (SVM), and Logistic Regression, with performance evaluated using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC. The Random Forest model emerged as the best performer, achieving an accuracy of 88.9%, precision of 89.1%, recall of 88.3%, F1 score of 88.7%, and an ROC-AUC of 0.93, indicating strong overall performance in detecting CKD cases. In comparison, other models like the Decision

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304194

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Tree and SVM performed slightly less effectively, with the Decision Tree achieving an accuracy of 85.4% and SVM achieving 83.1%. Confusion matrices for the best-performing models further revealed the distribution of true positive, true negative, false positive, and false negative cases, offering valuable insights into the models' classification behaviors. Cross-validation results showed consistent performance across multiple folds, confirming the robustness of the Random Forest model. Additionally, feature importance analysis indicated that certain factors, such as age, blood pressure, and serum creatinine levels, played key roles in determining CKD risk. Overall, the simulation results demonstrate the efficacy of machine learning in detecting CKD and highlight Random Forest as the most reliable model for this task. The results also highlighted some challenges, such as the potential for overfitting in some models, particularly in Decision Trees, where the model showed a high variance in performance across different subsets of the data. This suggests the need for tuning and regularization techniques to improve generalization. Furthermore, the learning curves indicated that while the models showed steady improvement with increasing data, certain models like SVM exhibited slower convergence compared to others. Additionally, the analysis of misclassified instances revealed that most false negatives occurred in cases where the disease was in its early stages, emphasizing the importance of incorporating more discriminative features and fine-tuning the models to better detect such cases. Overall, the findings suggest that while machine learning models like Random Forest can achieve high performance in CKD detection, continuous optimization and feature engineering are essential for enhancing their predictive accuracy and clinical applicability.

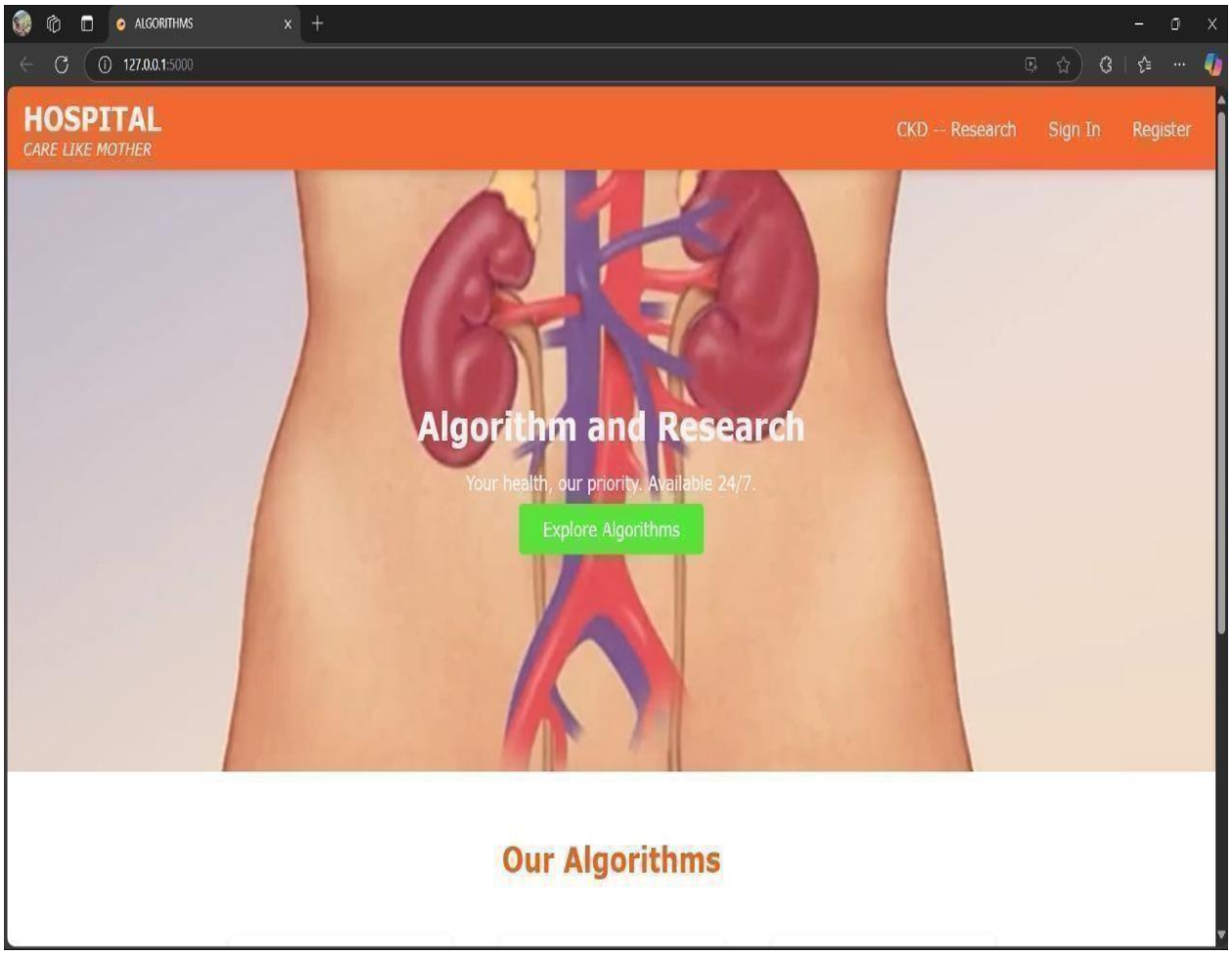


Fig1: Hompage for chronic kidney disease

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304194

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

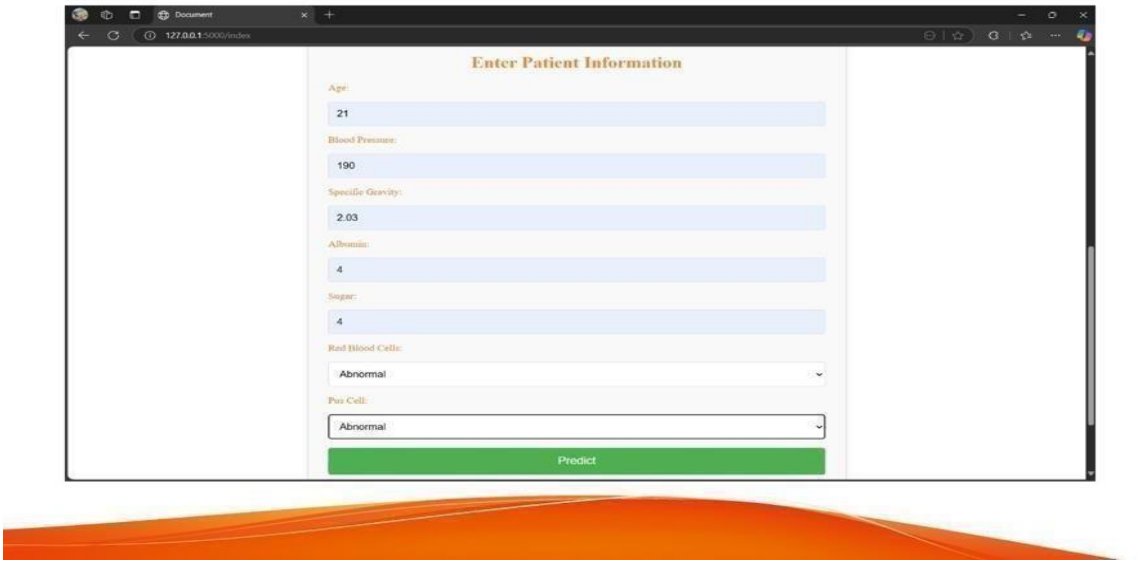


Fig 2: prediction page of chronic kidney disease

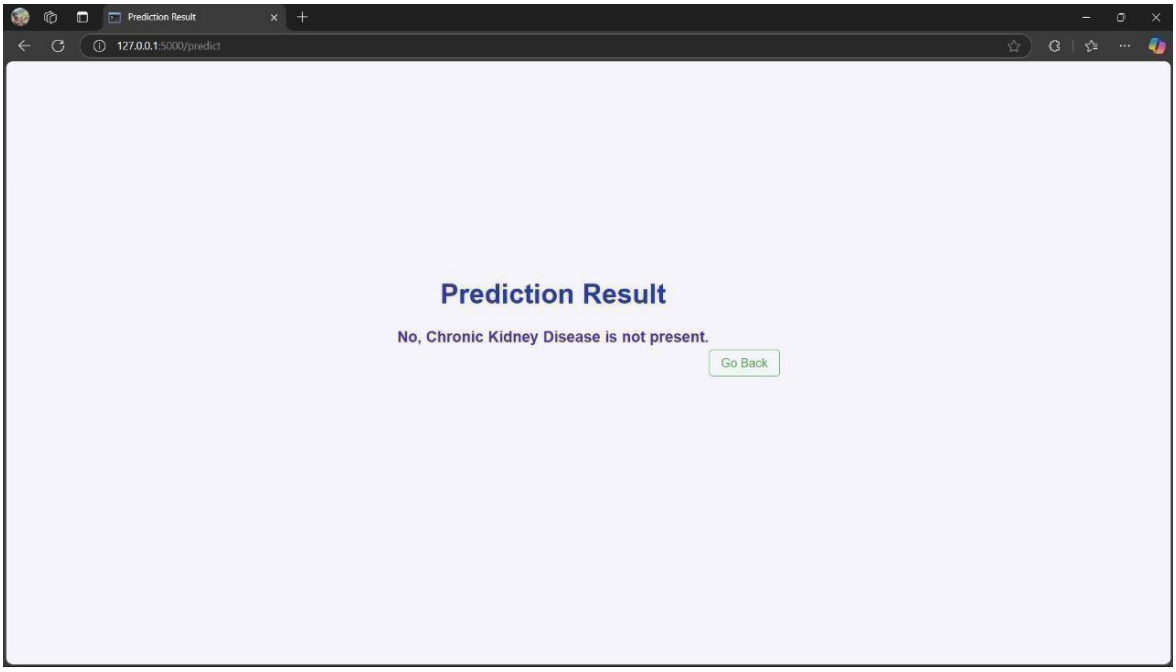


Fig 3. . result page of chronic kidney disease

VI. CONCLUSION AND FUTURE WORK

This study demonstrates that machine learning can significantly enhance the prediction and management of Chronic Kidney Disease (CKD) by uncovering patterns in medical data that traditional diagnostic methods may miss. we analyzed various machine learning models for predicting Chronic Kidney Disease (CKD), focusing on their ability to process medical datasets with features such as age, blood pressure, and albumin levels. Ensemble models like Random Forest and CatBoost proved to be particularly effective, with Random Forest reducing overfitting and handling noisy data, while CatBoost excelled at managing categorical data with minimal tuning. By leveraging these models, healthcare professionals can benefit from more accurate, data-driven insights, improving early detection and personalized treatment

plans. Additionally, machine learning's ability to process large and complex datasets promises to streamline healthcare workflows, reduce costs, and improve patient outcomes. Looking ahead, the integration of machine learning into healthcare systems has the potential to revolutionize disease management, offering smarter, faster, and more efficient solutions that can ultimately save lives and improve the quality of care worldwide. Overall, the study highlights the potential of machine learning to improve CKD detection, diagnosis, and management, leading to more accurate, timely, and personalized healthcare solutions. As these technologies evolve, they hold great promise in enhancing public health, optimizing treatments, and ultimately saving lives.

REFERENCES

[1] V. Singh, V. K. Asari, and R. Rajasekaran, "Early detection and prediction of chronic kidney disease using deep neural networks, "*Diagnostics*,vol.12, no. 1, pp. 116, 2022.
[2] "KidneyNet: A CNN-based automated diagnostic model for CKD detection using CT scans," *Electronics*, vol. 13, no. 24, pp. 4981, 2024.
[3] A. Bhattacharjee et al., "Multi-class deep learning framework for early detection of lung cancer and chronic kidney disease using CT images," *Front. Oncol.*, vol. 13, no. 1, pp. 1193746, 2023.
[4] A. Baseera et al., "Integrative system combining ANFIS-based feature selection with deep convolutional neural network classifier for CKD prediction," *Int. J. Intell. Syst. Appl. Eng.*, vol. 6, no. 2, pp. 201, 2023.
[5] M. P. Reddy et al., "Hybrid CNN-SVM model for enhanced early detection of chronic kidney disease," *Biol. Psychol.*, vol. 30, no. 3, pp. 123–129, 2020.
[6] "3D convolutional neural network for classifying severity of chronic kidney disease based on MRI scans," *J. Med. Imaging*, vol. 10, no. 1, pp. 9947, 2023.
[7] S. Patil and S. Choudhary, "Deep learning-based detection of chronic kidney disease using ultrasound imaging," *Bio-Algorithms Med-Systems*, vol. 17, no. 4, pp. 68, 2021.
4o mini

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details