"Comparative Prediction Analysis using ML Algorithms for CKD and Diabetes "

Assistant Professor Naveen Kumar B,DoS in CSE **,**navinaveen_b@yahoo.co.in

Varsha Bhajantri, MTechScholar, UBDTCE Davangere, smartvsb123@gmail.com

Abstract-Chronic Kidney Disease and Diabetes are major chronic illnesses requiring early detection to avoid complications. This project introduces a web-based prediction system using machine learning. For CKD, three feature selection methods Anova,Mutual Information,RFE with random forest were combined to identify impotent key clinical features, which improved model efficiency and accuracy. CatBoost 0.98, while for diabetes, XGboost model reached 0.82 accuracy. The final models were deployed in a user-friendly web platform.

Keywords-XGBoost,CKD,Risk Prediction, CatBoost,Machine Learning, , Predictive Analytics.

I. INTRODUCTION

Diabetes and CKD is most serious health challenges in present century.Diabetes affect more then 400 millions of peoples(WHO).Every year new cases significantly increasing. Similarly CKD affect around 10% of total global population.Both ckd and diabetes is economic and social burden.Research in the year 2020 total 7.8millions people are living with CKD. Approximately 210,000 new cases of kidney failure are diagnosed annually in India.Every year around 1.4 million individuals in India.

Kidneys act as filters that remove excess water, salts, and by-products of metabolism such as urea, creatine, and uric acid, along with certain drugs and toxins. When kidney function declines, these substances accumulate in the body, eventually leading to serious complications if left untreated.CKD has emerged as a critical health problem, affecting a large portion of the population. Early detection is therefore essential, and physicians rely on tests such as estimated glomerular filtration rate (eGFR), urine analysis, and blood pressure monitoring to evaluate kidney status.The severity of CKD is categorized into stages using GFR values. The earliest stages may show little to no symptoms despite underlying damage, while middle stages often present with tiredness, swelling, and urinary changes due to waste build-up. In advanced stages, kidney function is severely impaired, symptoms worsen, and patients may require dialysis or a transplant. The final stage, also called end-stage renal disease, represents complete kidney failure, where survival depends on medical replacement therapies.

Stages of CKD:Stages of CKD is classified based on eGFR. Lower the GFR worse the kidney function.Stage 1:Here Kidney are still functioning normally,but there may be structural damage.Stage 2:Here Kidney function is slightly decreased,but there may still be no symptoms.Stage 3:In this stage waste start building up in the blood Symptoms may include fatigue,swelling,back pain, and urination changes.Stage 4:Severely reduced kidney functions.Decline in kidney function with significant symptoms such as swelling,fatigue,nausea,preparing for dialysis or transplant of kidney.Stage 5:It is the end stage only dialysis or kidney transplant is end stage of renal disorder.

2023 study by ICMR reported that in India over 10.1 crore people suffering diabetes.Symptoms of diabetes is increase carving ,appetite frequent urination ,increased thirst. Diabetes leads to heart failure and chronic kidney disease blindness and stroke.It damages the kidneys filtering units.Prolong high blood sugar damages kidney filtering capacity leading to diabetic kidney diseases.DKD is serious obstacle of diabetes mellitus that affect kidney. Project target is to develop a single platform CKD and Diabetes is combined and Catboost classifier and XGBoost classifier are used in the backed to predict the result accurately.A light weight web framework Flask is used to develop and API ,main goal is to improve the clinical outcome and reduces the disease burden.Dual-disease prediction system is developed to assess ckd and diabetes risk.

II. RELATED WORK

[1].Comparative Study of ML Methods for Early Diabetes Prediction

Paper results Neural Networks on the PIDD dataset. The result shows using Neural Network classifier achieved the highest accuracy 78% and 76% of accuracy by Random Forest. In paper it is noted that using ensemble method and proper feature feature selection method it is possible to get better accuracy. Primary focus is only on accuracy.Only accuracy not always correct when dataset is imbalanced.To improve the study other metric like precision, recall, AUC verified.

Citation: Alzboon, M. S., Al-Batah, M., Alqaraleh, M., Abuashour, A., & Bader, A. F. (2025).

[2].Transparent & Accurate Prediction with Explainable AI

In this study,Prediction without balance data it show overfitting,hence ML models were combined with XAI tools, with SMOTE used to balance data. The model achieved around 92.5 % accuracy and an AUC of 0.975. Key predictive features included BMI, age, physical activity, income, and general health. The authors emphasized that transparency is as important

as accuracy.But data set used is survey based data set it may intriduce in accuracy.
Citation: Khokhar, P. B., Pentangelo, V., Palomba, F., & Gravino, C. (2025). Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable AI.
[3].Local Dataset Study in North Kashmir
Using clinical data collected locally in the Bandipora district,(RF, MLP, SVM, Decision Tree, Logistic Regression) used. Random Forest gave the best performance with ~98 % accuracy, followed by Gradient Boosting (~97 %), MLP (~90.99 %), etc. Data balancing and feature selection were used to combat class imbalance and overfitting.
Citation: Bhat, [First Name], et al. (2022). Prevalence and Early Prediction of Diabetes Using Machine Learning in North Kashmir: A Case Study.
[4]Ensemble & Feature-Selection Hybrid Models
Project study focused on combining ensemble learning with feature selection (e.g. Boruta) and optimization techniques. On the PIDD dataset, their hybrid model achieved around 98 % accuracy, showing that combining feature selection + ensemble classification results than standalone models.
Citation: Raja, et al. (2023). Prediction of Diabetes Disease using an Ensemble of Machine Learning Multi-Classifier Models. BMC Bioinformatics.
[5]. Fine-Tuned CatBoost with Nature-Inspired Optimization (Haque et al., 2025)
Haque and colleagues (2025) explored different machine learning algorithms for CKD detection, giving special focus to a CatBoost model refined with parameter tuning. To improve feature selection and reduce the impact of outliers, they applied optimization strategies based on simulated annealing and cuckoo search. Among the compared models (Random Forest, Logistic Regression, and MLP), the optimized CatBoost achieved the best results, recording an accuracy close to 98.75% and an AUC of nearly 0.999. Their analysis highlighted the diagnostic importance of attributes such as urine,albumin concentration, hemoglobin level, and diabetic status.
[6]. Deep Learning for Renal Replacement Therapy Prediction (Leung et al., 2024)
Retrospective cohort analysis to predict the timing of renal replacement therapy (RRT) among CKD patients. Using deep learning, they demonstrated that data-driven predictions could provide timely identification of hight risk cases beyond the capacity of traditional methods. This work points to the possibility of using AI systems to support nephrologists in planning and optimizing treatment before kidney failure reaches a critical stage.
[7]. Systematic Review of AI/ML in CKD Progression (PubMed, 2024)
A systematic review published in 2024 examined existing machine learning and AI approaches for predicting CKD progression and transition to ESRD or dialysis. The review included studies that employed logistic regression, SVMs, random forest, neural networks, and deep learning methods. While most approaches showed encouraging results, the review emphasized that differences in dataset sizes, follow-up duration, model design, and reporting metrics make direct comparison difficult. The authors stressed the need for standardized protocols and validation across diverse populations to strengthen clinical applicability.

III METHOD

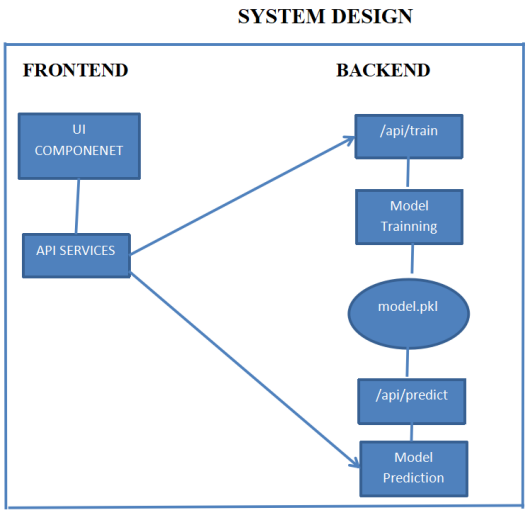


Figure:System Architecture for the Machine Learning Model Deployment

1.Frontend Layer:The frontend represents the user interface of the system, where users interact with the application. It consists of two key components.
UI Component: This module provides input fields and forms through which users can enter data or upload datasets. It also displays the prediction results in an interpretable format.
API Service: The frontend does not directly interact with the machine learning model. Instead, it communicates with the backend through API calls, typically implemented using HTTP methods such as POST and GET.
2. Backend Layer:The backend is responsible for executing the core machine learning tasks, including model training and prediction. It exposes two primary API endpoints. /api/train: This endpoint enables the training of the machine learning model. When training data is provided, the backend processes the dataset, applies preprocessing techniques, and trains the chosen algorithm. The resulting trained model is stored in a serialized file format, commonly model.pkl, which ensures that the model can be reused without retraining.
/api/predict: This endpoint is designed for prediction. It loads the previously trained model (model.pkl) and uses it to generate predictions based on new user input data. The output is then returned to the frontend in a structured format, such as JSON.
3. Model Serialization (model.pkl):The serialized model acts as a bridge between training and prediction phases. Once the model is trained, it is stored as a .pkl file using Python's serialization libraries. This allows

efficient loading of the model for future prediction tasks without repeating the training process.

4. Communication Workflow:The interaction between the frontend and backend follows a structured workflow. Users submit input data through the UI, which triggers API requests to the backend. In the case of training, the /api/train end point is invoked, and for predictions, the /api/predict endpoint is used. The backend responds with results, which are then displayed to the user through the frontend interface.This layered architecture not only ensures a clear separation of user interface and machine learning logic but also facilitates reusability, maintainability, and scalability of the system.

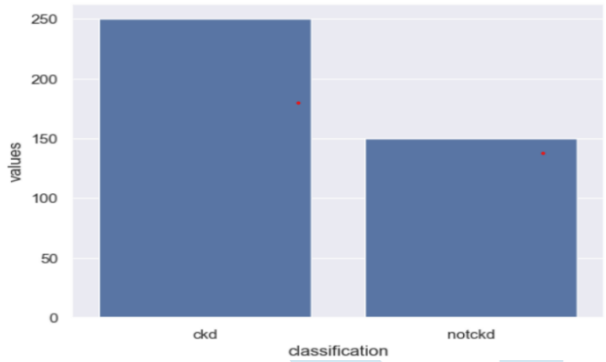
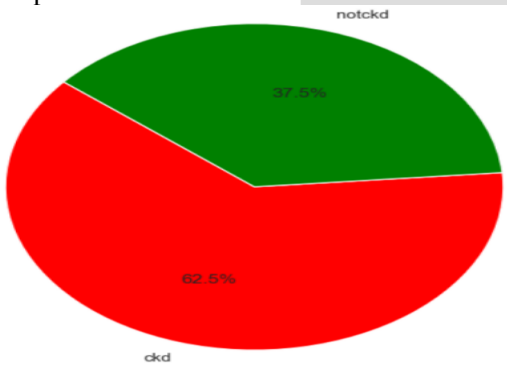


Figure:ckd dataset

Among 400 patient records 62.5% patients are having chronic kidney disease that is 250 and 37.5% patients are free from chronic kidney disease that is 150 records .The dataset consisting of 11 numerical attributes and 14 categorical attributes.Every features plays important role in disease predictions descriptions of attributes.



Percentage_of_ckd_notckd

Age:Kidney function naturally declines with age. Research shows that CKD prevalence significantly increases in populations over 60 years old. However, age is not always a determining factor, as genetic conditions like Polycystic Kidney Disease (PKD) or Alport Syndrome can cause kidney dysfunction even in younger individuals.

Blood Pressure (bp):The kidneys contain tiny filtering units called glomeruli. High blood pressure can damage these filters, reducing their ability to function effectively. Conversely, damaged kidneys may also cause secondary hypertension by retaining excess sodium and water, while releasing hormones that further elevate blood pressure.

Specific Gravity (SG):A low SG value (below 1.005) indicates poor concentration ability, which often reflects impaired kidney function.

Albumin (AL):Healthy kidneys prevent albumin (a type of protein) from passing into urine. When the glomeruli are damaged, albumin leaks into the urine (a condition known as proteinuria), signaling reduced kidney filtration efficiency.

Sugar in Urine (su):The presence of sugar in urine is abnormal. Values range from 0 (normal) to 5 (high). Elevated sugar in urine often results from diabetes and can damage kidney filters, reducing the glomerular filtration rate (GFR).

RBC: Low red blood cell counts are often associated with anemia in CKD patients. Blood in urine (hematuria) may also indicate kidney damage.

WBC: Reduced WBC counts weaken the immune system, making patients more susceptible to infections.

Pus Cells (PC):Higher values may be associated with progressive CKD or recurrent infections.

Pus Cell Clumps (PCC):Clumped pus cells are a marker of severe infection, strongly linked with advanced UTIs that can worsen kidney damage.

Bacteria (BA):Bacteria in urine, detected under a microscope, is another sign of UTI. Since CKD patients often have weaker immunity, UTIs are more frequent and require medical attention.

Blood Glucose Random (BGR):This test measures blood sugar at any random time of day. Consistently high values suggest diabetes, which is a leading cause of CKD over time.

Blood Urea (BU):Elevated blood urea indicates that kidneys are failing to remove waste effectively. In severe cases, red blood cells may also appear in urine, turning it pink, red, or brown—an alarming sign of kidney damage.

Serum Creatinine (SC):Creatinine is a waste product from muscle metabolism. High serum creatinine levels are a key biomarker of kidney dysfunction, as healthy kidneys normally filter it out.

Sodium (SOD):Sodium is an essential electrolyte that maintains blood pressure, fluid balance, and nerve function. Abnormal sodium levels can reflect kidney dysfunction.

Potassium (POT):High potassium levels (hyperkalemia) are one of the most dangerous complications of CKD, as they can cause life-threatening heart rhythm disturbances.

Hemoglobin(hemo),Packed Cell Volume(pcv):Lower pcv level indicates low level of hemoglobin.It is strongly corelated with with chronic kidney disease

Hypertension(htn):Hypertension reduces the kidney filtering ability,damages the blood vessels in the kidney and ckd it self cause secondary hyper tension.Diabetes

Mellitus(dm):Diabetes mellitus it is metabolic disorder with high blood suger level

Diabetes Mellitus (DM):A metabolic disorder characterized by chronically high blood sugar. Long-term uncontrolled diabetes is the most common cause of CKD worldwide.

Coronary Artery Disease (CAD):Patients with CKD are at higher risk of CAD due to narrowed or blocked coronary arteries.
Appetite (Appet):Loss of appetite is a common symptom in later stages of CKD. It often results from the buildup of toxins in the blood (uremia).
Pedal Edema (PE):Swelling in the ankles, feet, or legs due to fluid retention is a hallmark symptom of kidney dysfunction.
Anemia (Ane):CKD patients often develop anemia due to impaired production of erythropoietin (a hormone made by the kidneys).
Classification (Target Attribute):The target attribute in the dataset indicates whether a person has CKD or not. ML models use the above clinical and biochemical features to predict the classification outcome.
Dataset Information:

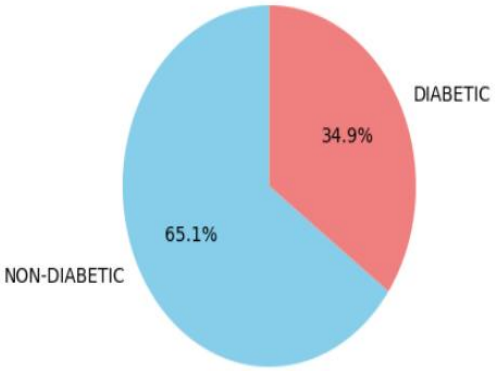


Figure:Percentage of diabetes patients in PIDD
It totally consisting of 9 attributes and 768 records .65.1% of the data in a PIDD indicate that 500 records are Non-Diabetes Patients record.34.9% of the data in a PIDD indicate that 268 records are Diabetes Patients record.

5.1.2 Pre-processing

One of the most important steps in preprocessing is handling missing values.The total number of missing values for each attribute was analyzed and represented in a graphical format for better understanding. From the visualization, it was observed that several attributes had a high or moderate number of missing values. These include:Red Blood Cells,White Blood Cell Count (WBC),Sodium,Potassium (POT), Packed Cell Volume (PCV),Pus Cells (PC)Hemoglobin (Hemo),Sugar in Urine,Albumin ,Specific Gravity ,Blood Glucose Random (BGR),Red Blood Cell Count (RBCC)In total, 12 attributes were identified with high to moderate levels of missing data, which required appropriate imputation or handling techniques.On the other hand, certain attributes were found to have very few missing values, such as:Hypertension (HTN),Diabetes Mellitus,Coronary Artery Disease and Appetite (Appet), Anemia,Age,Blood Pressure Since the proportion of missing values in these attributes was low, simple imputation methods (such as filling with mode/mean/median) or dropping those rows had minimal impact on overall data integrity.
Handling Missing Entries Two different imputation techniques were applied depending on the type of attribute:Numerical Features:Handled using the MICE (Multiple Imputation by Chained Equations) technique.MICE was selected because it models each incomplete feature as a function of the other features in the dataset.It captures complex relationships across variables and iteratively updates estimates, which improves accuracy compared to simple imputation methods (e.g., mean or median).This method is especially effective in clinical datasets where attributes are interrelated (e.g., hemoglobin, packed cell volume, and red blood cells).
Categorical Features: .Mode imputation is computationally simple, preserves categorical distribution, and prevents data distortion in variable.The choice of MICE for numerical features and Mode imputation for categorical features ensured that the dataset retained its underlying statistical relationships, thereby reducing bias and improving the performance.

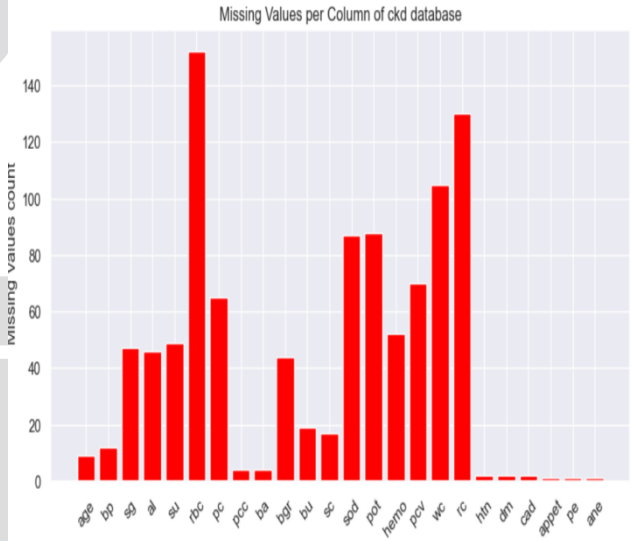


Figure: Missing values count CKD

Outlier Detection

Outliers are unusual records that differ strongly from the rest of the data, often showing values that are much higher or lower than expected. In the CKD dataset, such points were checked using the Interquartile Range approach, which is a reliable way of measuring how spread out the central data is.
When applied to the CKD dataset, the features with the largest number of flagged outliers were:Urine sugar level (SU),Serum creatinine,Blood urea (BU),Blood pressure (BP),Random blood glucose (BGR),Sodium (SOD)However, in medical datasets, extreme values do not always mean “errors.” For example:A very high serum creatinine level may be completely valid in a patient with advanced kidney damage.Similarly, very low sodium or extremely high blood urea levels could reflect true medical conditions.Only unrealistic values caused by human error or data entry mistakes should be considered for removal.Therefore, while the IQR method is useful for spotting unusual data points, domain knowledge is essential to decide whether a point is truly an outlier or a genuine medical reading.

Handling Missing Values in Diabetes Dataset:In the diabetes dataset, missing or invalid entries were

carefully analyzed before applying machine learning models.Pregnancies: The value 0 in the number of pregnancies attribute is considered valid, since some women may never become pregnant. Therefore, these entries were not treated as missing values.Attributes without Missing Values: Features such as Age and Diabetes Pedigree Function contained no missing entries, so no preprocessing was required for these variables.Attributes with Missing Values: Several attributes contained missing or invalid entries that were represented as zeros, which are not physiologically possible. These included:Glucose,Blood Pressure,Skin Thickness,Insulin,BMI For example, a glucose level of zero is not realistic in a living individual, and thus such cases were treated as missing values.

High-Missing Attributes: Among these, Insulin and Skin Thickness had the highest proportion of missing entries, requiring careful handling during preprocessing.

Imputation Strategy: To resolve the missing values, imputation techniques were applied. Depending on the attribute type and distribution, suitable statistical methods used to replace missing entries while preserving the overall characteristics of the dataset.

Figure :Missing Value count Diabetes

Outlier Detection in Diabetes Dataset:

Data points that fall outside the expected clinical range and can influence the quality and accuracy of analysis. To identify such values in the diabetes dataset, visualization techniques, particularly boxplots, were applied.The analysis revealed that:High insulin values and high glucose values were frequently flagged as outliers.However, these values are clinically possible in patients with severe diabetes and therefore should not be removed.Only outliers resulting from human errors, measurement mistakes, or data entry issues were considered for correction or removal.The decision-making process ensured that extreme but valid clinical cases were preserved, while unreliable records were handled to improve the overall reliability of the dataset.

Encoding Categorical Data:

Certain attributes in the CKD and diabetes datasets are categorical in nature (e.g., Hypertension: yes/no, Red Blood Cells: normal/abnormal, Appetite: good/poor, Pedal Edema: yes/no). Categorical values converted to numerical form using encoding techniques.Binary categories (e.g., yes/no, good/poor) were encoded as 0 and 1.Attributes such as Diabetes Mellitus (DM) were already available in numerical form, so no further encoding was required.This transformation ensured that categorical variables could be processed directly by the model without loss of information.

Feature Scaling

ML models are sensitive to the scale of continuous variables. Features with large value ranges may dominate other features during training.

To address this, standardization was applied:

$z=(x-\mu)/\sigma$

where: x = original value μ = mean of the feature σ = standard deviation of the feature

3. Features Selection

Feature selection was performed to improve computational efficiency, reduce noise, and minimize the risk of overfitting. By identifying the most relevant predictors, the model was able to focus only on medically significant factors that contribute meaningfully to CKD prediction.

Multiple methods were applied to evaluate feature importance.ANOVA (Analysis of Variance): Assessed the statistical significance of individual features with respect to the target variable.Mutual Information (MI): Measured the dependency between features and the output label, capturing both linear and non-linear relationships

Random Forest Classifier (RFC) Feature Importance: Utilized the learning method to rank predictors based on their contribution to decision-making in tree-based models.

Combining the results of these techniques, the following top 7 features were consistently identified as the most influential for prediction:Packed Cell Volume (PCV),Specific Gravity (SG),Albumin (AL),Diabetes Mellitus (DM),Hypertension (HTN),Serum Creatinine (SC),Hemoglobin (Hemo).These selected features form the optimal subset for training different machine learning algorithms, ensuring that the models achieve high predictive accuracy while remaining medically interpretable.

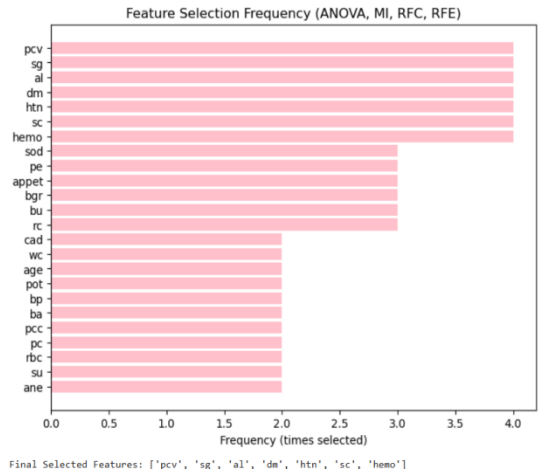


Figure:CKD Feature selection

This plot shows how well six different machine learning models performed on the dataset using all

available features compared to using only a selected subset of important features.

Feature Selection for Diabetes Dataset

Diabetes dataset is relatively small in size; therefore, the complete dataset was utilized directly for model development and web deployment. However, to improve interpretability and highlight the most significant predictors, Random Forest Feature Importance was applied to rank the attributes.

The analysis identified the following features as the most influential for diabetes prediction:

Glucose – A primary indicator of abnormal blood sugar levels and a strong clinical marker for diabetes.BMI (Body Mass Index) – Reflects body weight relative to height; obesity is a well-known risk factor.Age – Older individuals have a higher likelihood of developing diabetes due to metabolic changes.Diabetes Pedigree Function (DPF) – Captures genetic predisposition and family history of diabetes.These selected features align with both clinical evidence and statistical importance, ensuring that the predictive models remain both accurate and medically meaningful.

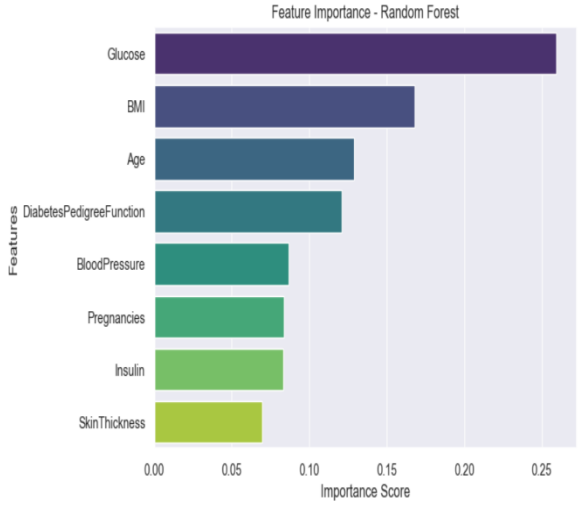


Figure:Diabetes Feature selection

Model Construction :

Logistic Regression (LR):Served as the baseline. It is statistically interpretable and helps identify risk factors (e.g., high serum creatinine or low hemoglobin strongly linked with CKD). However, performance was moderate compared to ensemble methods.Used as a baseline, giving insight into how features like glucose, BMI, and age affect diabetes risk.Random Forest (RF):An ensemble of decision trees that reduced overfitting and provided high accuracy. Random Forest also produced feature importance rankings, which validated the medical significance of predictors such as creatinine level, specific gravity, hemoglobin, packed cell volume, and albumin.Gradient Boosting (GB):Sequentially built decision trees to minimize errors from previous models. Effective in capturing non-linear relationships between clinical features. Support Vector Classifier (SVC):Plotted decision boundaries in multidimensional space. It performed well for cases where CKD risk factors showed complex, non-linear separation. K-Nearest Neighbors (KNN):Classified patients based on similarity to closest neighbors. Performance was acceptable after standardization, but it was more sensitive to noise compared to ensemble methods. XGBoost (Extreme Gradient Boosting):

Optimized for speed and accuracy. Provided higher predictive power compared to traditional boosting methods. CatBoost Classifier:Especially effective because it handles categorical attributes (like hypertension, diabetes, and appetite) without extensive preprocessing. CatBoost achieved excellent accuracy while also reducing bias.

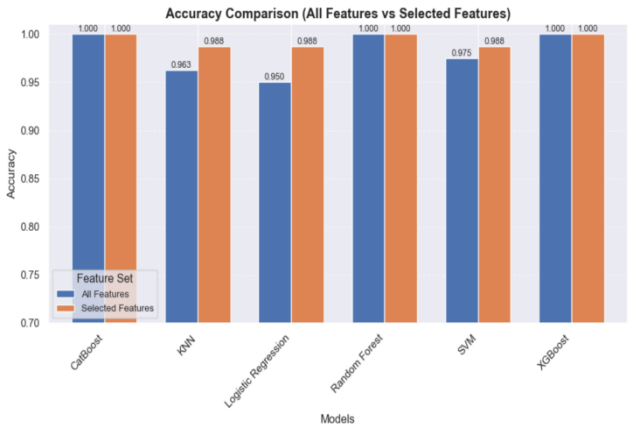
5.1.5. Backend and Frontend Implementation

The final selected model was deployed using Flask as the backend framework. After training, the machine learning model was serialized and saved in .pkl format using libraries such as Joblib or Pickle. Flask acted as a lightweight web server and exposed the model through RESTful APIs.The API was designed to:Accept patient details (clinical attributes like serum creatinine, hemoglobin, blood pressure, glucose, etc.) as input in JSON format.Preprocess the input data to match the format used during training (handling scaling, encoding, or missing values where necessary).Run the prediction using the loaded .pkl model.Return the prediction output as a response, indicating whether the patient is likely to have Chronic Kidney Disease (CKD) or not.On the frontend, a React.js interface was integrated with the Flask API. Users can enter health parameters in a simple form, and once submitted, the input is sent to the backend. The Flask API processes the request, performs the analysis, and sends back the prediction result, which is visible to user.

This deployment ensures that the system can:Operate as a decision-support tool for healthcare professionals.Allow easy interaction for patients through a web interface.Provide fast, accurate, and consistent predictions without retraining the model each time.

IV Performance Evaluation

The models were validated using train-test split and cross-validation.Accuracy comparison with and without feature selection for ckd.

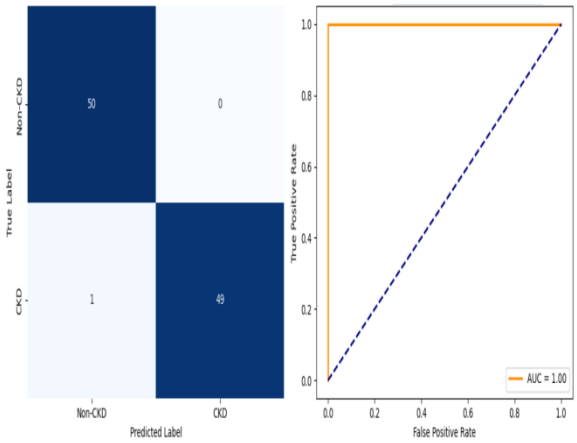


Fetature comparison plot

Maximum accuracy for CKD is obtained by Random forest and Catboost and XGboost.Individual Catboost model achived maximum accuracy

Classification Report:

	precision	recall	f1-score	support
0	0.96	1.00	0.98	50
1	1.00	0.96	0.98	50
accuracy			0.98	100
macro avg	0.98	0.98	0.98	100
weighted avg	0.98	0.98	0.98	100



1. Confusion Matrix :The confusion matrix compares the predicted labels against the actual labels for CKD and Non-CKD.showing that it is extremely reliable in identifying CKD patients while avoiding false alarms.
2. ROC Curve (Receiver Operating Characteristic):In the figure, the orange line almost touches the top-left corner, which represents an ideal classifier.The AUC (Area Under the Curve) value is 1.00, which means the model has perfect discrimination ability between CKD and Non-CKD patients.

Model Performance Evaluation for Diabetes Prediction

Classification Report:

	precision	recall	f1-score	support
0	0.81	0.79	0.80	100
1	0.79	0.81	0.80	100
accuracy			0.80	200
macro avg	0.80	0.80	0.80	200
weighted avg	0.80	0.80	0.80	200

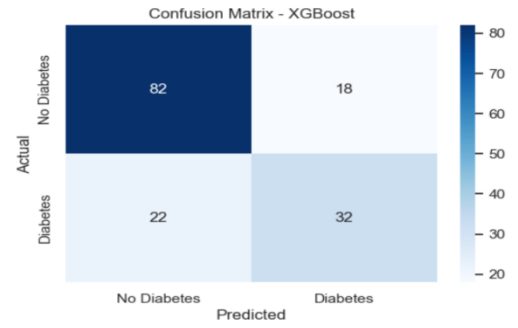


Figure:Confusion matrix for diabetes

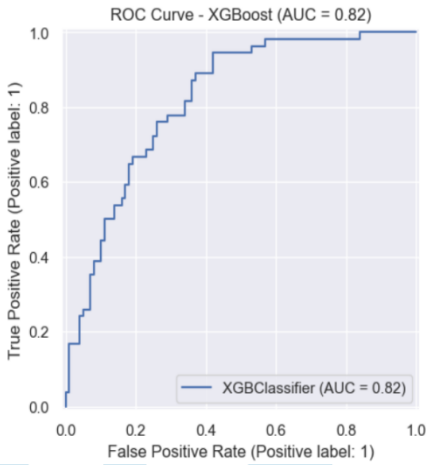


Figure :ROC-Curve matrix for diabetes

Confusion Matrix :The confusion matrix displays comparison between actual and predicted outcomes.Maximum accuracy for diabetes dataset predicted by XGboost model is AUC is 082.

From the results:TP: This indicates that the model performed well overall but made a moderate number of misclassifications in both diabetic and non-diabetic groups.

2. ROC Curve :It evaluates the model's ability to distinguish between diabetic and non-diabetic patients.The curve lies significantly above the diagonal reference line.The AUC value is 0.82, which indicates a good level of discrimination between the two class.

V. CONCLUSION

This project developed a web-based system for predicting both CKD and Diabetes using ML. For CKD, Different feature selection methods were combined to identify ten important attributes, which improved the efficiency of the models. RF achieved the highest accuracy of 0.98, while CatBoost reached 0.98, and for Diabetes, the XGboost model gave 0.82 accuracy. The system was deployed in a user-friendly web platform, making it accessible for practical use.Unlike many existing works that concentrate on a single disease or apply only one method, this project combines two critical diseases in one platform and introduces a combining three feature selection approach for CKD,making it more comprehensive, accurate, and closer to real-world healthcare needs.Overall,This project demonstrates that machine learning integrated with web application can provide valuable tools for better healthcare outcomes.

REFERENCES

[1].Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. (2022). KDIGO 2022 Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease. Kidney International, 102(5S), S1–S127. https://doi.org/10.1016/j.kint.2022.07.001

[2].de Boer, I. H., Khunti, K., Sadusky, T., Tuttle, K. R., Neuen, B. L., & de Zeeuw, D. (2022). Diabetes management in chronic kidney disease: A consensus report by the American Diabetes Association (ADA) and Kidney Disease: Improving Global Outcomes

(KDIGO). Diabetes Care, 45(12), 3075–3090 https://doi.org/10.2337/dci22-0027
[3].Khunti, K., Perkovic, V., & de Boer, I. H. (2021). Chronic kidney disease in diabetes: Guidelines from KDIGO. American Family Physician, 103(11), 698–700.
[4].Mottl, A. K., Tuttle, K. R., Rovin, B. H., & others. (2024). KDOQI commentary on the KDIGO 2022 clinical practice guideline for diabetes management in CKD. American Journal of Kidney Diseases, 83(1), 1–15. https://doi.org/10.1053/j.ajkd.2023.08.018
[5].Tobe, S. W., McFarlane, P., & Senior, P. A. (2025). Chronic kidney disease in diabetes: A clinical practice guideline. Canadian Journal of Diabetes, 49(1), 1–18. https://doi.org/10.1016/j.jcjd.2024.11.002