International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 25 Issue 03, MAR, 2025

A MACHINE LEARNING APPROACH FOR ACCURATE PREDICTION OF CHRONIC KIDNEY DISEASE AND ITS STAGES

Mrs.D. Naga Vardhani[1], Taraka Ranganath Sai Madasu[2], Sai Prakash Chowdary Pemmasani[3], Giri Sai Harha Yaramachu[4], Pradeep Kumar Jandrajupalli [5]

[1]Assistant Professor, IT Department, Vasireddy Venkatadri Institute of Technology, Namburu, Guntur, Andhra Pradesh -522508

[2345]UG Students, IT Department, Vasireddy Venkatadri Institute of Technology, Namburu, Guntur, Andhra Pradesh -522508

Mail: ***vardhani467@gmail.com***

Abstract: *Chronic kidney disease (CKD) is a progressive condition affecting millions worldwide and poses significant challenges to healthcare systems. Early detection and accurate stage prediction are crucial for timely interventions and personalized treatment. This work employs robust ensemble learning models to predict CKD presence and eGFR to specify the stage of CKD. A dataset from the UCI Machine Learning Repository, which contains 400 instances of patients, has been used. The methodology includes essential data preprocessing steps, such as data cleaning and feature selection, to enhance model performance. The models, such as Random Forest, Gradient Boosting (XGBoost), and MLP with XGBoost (stack model), are employed individually for prediction. In addition to these advanced models, the project incorporates the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation for precise Glomerular Filtration Rate (GFR) calculation, ensuring accurate staging of the disease. These models are evaluated using key metrics such as accuracy, sensitivity, specificity, and F1 score. The Random Forest model achieves a 100% accuracy rate, and it is deployed into a user-friendly website that enables users or healthcare professionals to enter patient details and receive CKD predictions along with stage classifications, facilitating early detection and effective disease management.*

Keywords: chronic kidney disease (CKD), Data Preprocessing, Ensemble Learning Models, Glomerular Filtration Rate (GFR), Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) Equation.

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a progressive condition characterized by a gradual loss of kidney function over time. It is a major global health issue that can lead to kidney failure, requiring dialysis or a kidney transplant for survival. CKD is defined as abnormalities in kidney structure or function lasting for more than three months with health implications. It is a significant global health concern affecting millions of individuals worldwide. The kidneys play a crucial role in filtering waste, maintaining electrolyte balance, and regulating blood pressure. However, conditions such as diabetes, hypertension, and glomerulonephritis can impair kidney function, leading to CKD. If left untreated, CKD can progress to End-Stage Renal Disease (ESRD), requiring dialysis or a kidney transplant. Chronic Kidney Disease (CKD) affected approximately 753 million people worldwide in 2016, including 417 million females and 336 million males. The disease led to 1.2 million deaths in 2015, a significant increase from 409,000 deaths in 1990. Among the leading causes of CKD-related deaths, high blood pressure accounted for 550,000 cases, followed by diabetes with 418,000 cases and glomerulonephritis with 238,000 cases.

CKD can result from conditions such as diabetes, high blood pressure, glomerulonephritis, and polycystic kidney disease. A family history of CKD is considered a major risk factor. Diagnosis typically involves blood tests to assess the estimated glomerular filtration rate (eGFR) and urine tests to detect albumin levels. In some cases, additional diagnostic procedures such as ultrasound imaging or a kidney biopsy may be performed to identify the underlying cause of the disease.

CKD is categorized into five stages based on Glomerular Filtration Rate (GFR) with stages 1 and 2 representing mild kidney impairment and stage 5 indicating total kidney failure, requiring dialysis or transplantation. Early detection and accurate stage prediction are crucial for timely interventions and personalized treatment. Machine Learning techniques can help in predicting CKD with proper training. So, for that, this work employs a dataset from the UCI Machine Learning Repository, which contains 400 instances of patients' data collected from the hospital for nearly 2 months. The models, such as Random Forest, Gradient Boosting (XGBoost), and MLP with XGBoost (stack model), are employed individually for prediction. In addition to these advanced models, the project incorporates the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation for precise Glomerular Filtration Rate (GFR) calculation, ensuring accurate staging of the disease.

II. RELATED WORK

Rady and Anwar [1] conducted a comparative study on predicting kidney disease stages using Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Radial Basis Function (RBF) algorithms. Using a dataset comprising 400 cases with 24 features. The findings indicated that PNN achieved the highest overall classification accuracy of 96.7%, outperforming the other models evaluated in the study.

S. Tekale et al. [2] conducted a study titled "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm", utilizing a dataset comprising 400 instances and 14 selected features. They experimented with a

Decision Tree and Support Vector Machine (SVM) after reducing the original 25 features to 14 through preprocessing. Their results indicated that SVM performed better, achieving an accuracy of 96.75%, making it the more effective model in their study.

K.M. Almustafa [3] developed a classification system for kidney disease using a dataset comprising 400 cases with 24 features. Multiple machine learning classifiers were evaluated, with J48 and Decision Tree (DT) emerging as the top-performing models, achieving an accuracy of 99%. After applying feature selection, the performance of these classifiers further improved, along with Naïve Bayes and K-Nearest Neighbors (KNN), demonstrating the effectiveness of feature selection in enhancing model accuracy.

M.A. Islam et al. [4] conducted a study on early CKD detection using machine learning, utilizing a dataset of 400 cases with 24 attributes (13 categorical and 11 numerical). Following data preprocessing, they applied Principal Component Analysis (PCA) to identify the most significant features for CKD prediction. Among the evaluated models, XGBoost demonstrated superior performance, achieving an accuracy of 98.33% with the original dataset, which further improved to 99.16% after PCA was implemented. Other classifiers also attained 98.33% accuracy before PCA was applied.

Poonia et al. [5] explored various machine learning algorithms for kidney disease prediction, including K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Naïve Bayes (NB), and Logistic Regression. They also applied Recursive Feature Elimination (RFE) and the Chi-Square test as feature selection techniques. Using a dataset of 400 patients, each with 24 attributes, they developed and analyzed predictive models. Their findings revealed that a Logistic Regression model, optimized with features selected using the Chi-Square technique, achieved the highest accuracy of 98.75%.

Dibaba Adeba Debal [6] conducted a study on both binary and multi-classification for stage prediction using a Random Forest (RF), Support Vector Machine (SVM), and Decision Tree (DT). Using a dataset of 1718 instances with 19 features where 12 are numeric and 7 are nominal. The results from the experiments indicated that RF based on recursive feature elimination with cross-validation has better performance than SVM and DT with the highest accuracy of 99.8% for binary class. XGBoost has 82.56% accuracy for five classes.

III. Proposed Work

A. *Dataset Collection*

The data set used for this work is employed from the UCI Machine Learning Repository. The name of the dataset is “chronic kidney disease,” which has been collected from the hospital for nearly 2 months. The dataset consists of 400 instances of patient data with 24 features and one target class with a mix of 11 numerical and 14 nominal variables; 250 instances are CKD, and 150 are NON-CKD. The dataset consists of several missing values in several columns, affecting the Machine Learning Models' performance. Columns like rbc, pc, pcc, ba, htn, dm, cad, appet, pe, ane, and class are categorical. They need to be converted into numerical values. So, for that, this work needs to preprocess the data. A description of the Dataset is provided in TABLE 1.

B. *Data Preprocessing*

Data preprocessing is a crucial step in any machine learning pipeline, ensuring that the dataset is clean, consistent, and suitable for model training. In this work, preprocessing was performed to handle missing values, remove anomalies, encode categorical variables, and normalize numerical features to improve model accuracy and generalization.

1) Data Cleaning:

Data cleaning was performed to eliminate inconsistencies and prepare the dataset for analysis. This included:

a) Handling Missing Values: The dataset contained missing values due to improper data recording or unavailable patient information. Missing values were identified and replaced using the median for numerical attributes and the mode for categorical attributes

b) Removing Anomalies: Data anomalies such as erroneous characters (‘?’, ‘\t?’) were replaced with NaN and subsequently handled appropriately. Extra spaces in string values were removed to maintain consistency

c) Standardizing Column Names: Column names were stripped of unnecessary spaces to ensure uniformity during data preprocessing.

2) Encoding Categorical Variables: Machine learning models require numerical input; hence, categorical variables were transformed using label encoding. Binary categorical attributes (e.g., presence or absence of a condition) were mapped to 0 (No) and 1 (Yes), while multi-class attributes were converted into numerical equivalents. The following mappings were used*:*

- Medical Conditions: Normal = 1, Abnormal = 0
- Presence-based Attributes (e.g., Bacteria, Pedal Edema): Present = 1, Not Present = 0
- Appetite: Good = 1, Poor = 0
- Hypertension: Yes = 1, No = 0
- Target Variable**:** CKD = 1, Not CKD = 0

3) Handling Outliers and Data Normalization

Outliers were detected using statistical methods such as interquartile range (IQR) and were treated accordingly. Since numerical attributes varied significantly in scale, feature normalization was performed using Standardization (Z-score normalization) to improve model stability. The standardization formula used was:

$$X' = \frac{X-\mu}{\sigma}$$

Where X is the original value, μ is the mean, and σ is the standard deviation.

Table- 1: Dataset Description

Column Name	Datatype	Missing values	Range
Age (age)	numerical	9	2-90
Blood Pressure (bp)	numerical	12	50-180
Specific Gravity (sg)	nominal	47	1.005,1.010,1.015,1.020,1.025
Albumin (al)	nominal	46	0,1,2,3,4,5
Sugar (su)	nominal	49	0,1,2,3,4,5
Red Blood Cells (rbc)	nominal	152	normal, abnormal
Pus Cell (pc)	nominal	65	normal, abnormal
Pus Cell clumps (pcc)	nominal	4	Present, not present
Bacteria(ba)	nominal	4	Present, not present
Blood Glucose Random(bgr)	numerical	44	22 – 490
Blood Urea (bu)	numerical	19	1.5 - 391
Serum Creatinine (sc)	numerical	17	0.4 – 76
Sodium (sod)	numerical	87	4.5 -163
Potassium (pot)	numerical	88	2.5 - 47
Hemoglobin (hemo)	numerical	52	3.1 - 17.8
Packed Cell Volume(pcv)	numerical	70	9 – 54
White Blood Cell Count (wc)	numerical	105	2200 – 26400
Red Blood Cell Count (rc)	numerical	130	2.1 – 8
Hypertension (htn)	nominal	2	yes, no
Diabetes Mellitus(dm)	nominal	2	yes, no
Coronary Artery Disease (cad)	nominal	2	yes, no
Appetite (appet)	nominal	1	Good, poor
Pedal Edema (pe)	nominal	1	yes, no
Anemia (ane)	nominal	1	yes, no
Class	nominal	0	CKD, NONCKD

C. Feature Selection

Feature selection is the process of selecting the most relevant and significant features from a dataset to improve model performance, reduce overfitting, and enhance interpretability. It helps in eliminating redundant or irrelevant features that may negatively impact a model's accuracy and computational efficiency. This work has used Recursive Feature Elimination (RFE). It is a feature selection technique that recursively removes the least important features while building a model to identify the most relevant subset of features.

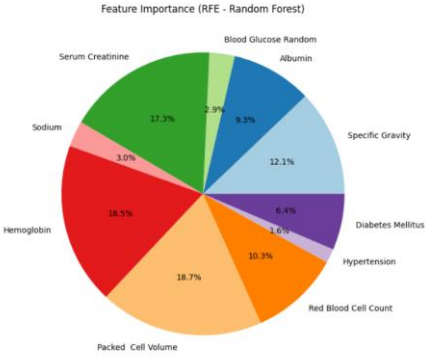


Fig- 1: Top 10 Features selected using RFE

D. *Machine Learning Models*

Ensemble Learning is a machine learning technique that combines multiple models, known as base learners or weak learners, to improve predictive accuracy, reduce variance, and enhance model generalization. By aggregating the outputs of multiple models, ensemble learning produces more robust and reliable predictions compared to individual models. This approach is widely used in classification, regression, and anomaly detection tasks, leveraging diverse algorithms to optimize performance and mitigate errors.
Types of Ensemble Learning Methods:

- *Bagging (Bootstrap Aggregating):* This reduces variance and prevents overfitting by training multiple instances of the same model on different subsets of the data. Example: Random Forest, which combines multiple decision trees trained on different bootstrapped samples.
- *Boosting:* Reduces bias by training models sequentially, where each new model corrects the error of the previous ones. Example: XGBoost (*Extreme Gradient Boosting)*
- *Stacking (Stacked Generalization*): Combines multiple base models to create a stronger predictive model. It works by training several base learners and then using a meta-model as the final decision layer. Example: Combining MLP and

XGBoost, with another model acting as the final decision layer.

1) *Random Forest*

Random Forest is a powerful ensemble learning algorithm that constructs multiple decision trees and combines their outputs to enhance accuracy and reduce overfitting. It operates using Bootstrap Aggregation (Bagging), where each tree is trained on a random subset of data, and feature randomness, selecting a subset of features at each split to improve generalization. By averaging predictions in regression tasks or using majority voting in classification, it mitigates overfitting compared to individual decision trees.

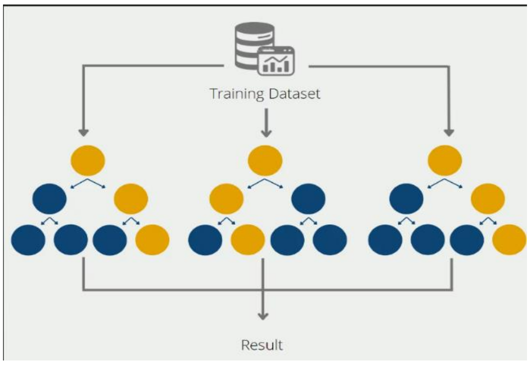


Fig- 2: Random Forest Architecture

2) *XGBoost*

XGBoost (Extreme Gradient Boosting) is a powerful, optimized implementation of the Gradient Boosting algorithm designed for high efficiency, speed, and accuracy. It builds decision trees sequentially, where each tree corrects the errors of the previous ones, reducing bias and improving predictions. XGBoost incorporates L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting and supports parallel processing for faster execution.

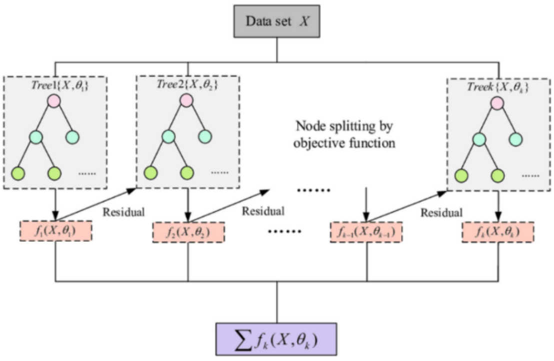


Fig- 3: XGBoost Architecture

3) *Multi-Layer Preceptron with XGBoost(stack model)*

Stacking is an ensemble learning technique that combines multiple machine learning models to improve predictive performance. Unlike bagging and boosting, which focus on combining similar models, stacking leverages the diversity of different algorithms by training a meta-model to aggregate their outputs. In stacking, base learners (such as decision trees, neural networks, or gradient boosting models) make predictions, and these predictions are then used as input features for a higher-level meta-model, which learns to optimally combine them. This approach helps capture different patterns in the data, reduce bias and variance, and enhance overall accuracy, making stacking a powerful method for complex classification and regression tasks.

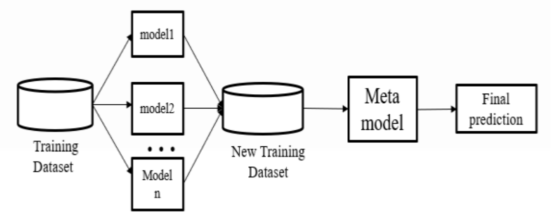


Fig- 4: Stack Architecture

Multi-layer perceptron (MLP) is a class of artificial neural networks (ANNs) composed of multiple layers of neurons, including an input layer, one or more hidden layers, and an output layer. Each neuron in a layer is fully connected to neurons in the next layer, with weighted connections that are adjusted using backpropagation and gradient descent to minimize error. MLP uses activation functions like ReLU, Sigmoid, or Tanh to introduce non-linearity, making it effective for complex pattern recognition and classification tasks.

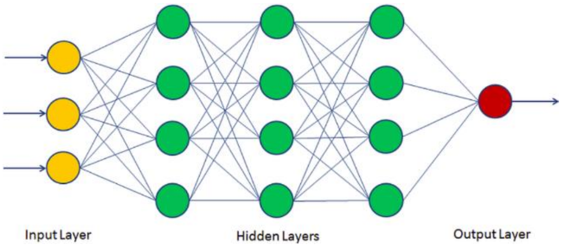


Fig- 5: MLP Architecture

Stacking Multi-Layer Perceptron (MLP) with XGBoost creates a powerful ensemble model that combines the strengths of both algorithms. In this approach, MLP and XGBoost are trained separately on the dataset, and their predictions are then used as input features for a meta-model, which typically is another machine learning algorithm (e.g., Logistic Regression). MLP captures complex, non-linear relationships in the data, while XGBoost efficiently handles structured features and reduces overfitting through boosting. This stacked model improves predictive performance by leveraging the complementary strengths of deep learning and gradient boosting.

E. Chronic Kidney Disease (CKD) Stages

The classification of chronic kidney disease (CKD) stages according to GFR (Glomerular Filtration Rate) follows the guidelines set by the Kidney Disease: Improving Global Outcomes (KDIGO) organization. It is also consistent with standards from the National Kidney Foundation (NKF) and the Kidney Disease Outcomes Quality Initiative (KDOQI). Chronic Kidney Disease (CKD) is classified into five stages based on the Glomerular Filtration Rate (GFR), which measures kidney function. Stage 1 (GFR ≥ 90 mL/min/1.73m²) and Stage 2 (GFR 60–89) indicate mild kidney damage with normal or slightly reduced function. Stage 3 (GFR 30–59) is moderate CKD, where symptoms may start appearing. Stage 4 (GFR 15–29) is severe CKD, requiring medical management to slow its progression. Stage 5 (GFR < 15) is kidney failure, often necessitating dialysis or a kidney transplant.

The most common formula used to calculate GFR is the CKD-EPI equation:

$$GFR = 141 \times \min(SCr/\kappa, 1)^{\alpha} \times \max(SCr/\kappa, 1)^{-1.209} \times 0.993^{Age} \times 1.018^{(\text{if Female})} \times 1.159^{(\text{if African American})}$$

Fig- 6: CKD-EPI equation.

Table- 2: CKD Stages Description

CKD Stages	GFR (mL/min/1.73m²)	Kidney Function
Stage 1	>=90	Normal
Stage 2	60 - 89	Mild decrease
Stage 3a	45 - 59	Moderate
Stage 3b	30 - 44	Moderate-Severe
Stage 4	15 - 29	Severe
Stage 5	< 15	Failure (ESRD)

F. Models Evaluation

All the models are evaluated using key metrics such as accuracy, sensitivity, specificity, and F1 score. Model evaluation is the process of assessing the performance of a trained machine learning model using various metrics and techniques to determine its accuracy, reliability, and generalization ability on unseen data. It involves comparing predicted outputs with actual values to ensure the model meets desired performance standards before deployment.

TP, TN, FP, and FN are key components of the confusion matrix, which helps evaluate the performance of a model.

- True Positive (TP): The model correctly predicts a positive instance (e.g., detecting CKD in a patient who actually has CKD).

- True Negative (TN): The model correctly predicts a negative instance (e.g., predicting no CKD in a person who truly does not have CKD).

- False Positive (FP): The model incorrectly predicts a positive instance when it is actually negative (e.g., diagnosing CKD in a healthy person – a false alarm).

- False Negative (FN): The model incorrectly predicts a negative instance when it is actually positive (e.g., failing to detect CKD in a patient who has the disease – a missed diagnosis).

Key Metrics for Model Evaluation

1) Accuracy: Measures the percentage of correctly predicted instances out of the total instances. Suitable for balanced datasets.

$$Accuracy = (TP + TN)/\ (TP + TN + FP + FN)$$

2) Precision: Indicates how many of the predicted positive cases were actually positive. Useful for minimizing false positives.

$$\text{Precision} = (TP)/\ (TP + FP)$$

3) Recall (Sensitivity): Measures the model's ability to detect actual positive cases. Important when false negatives are costly.

$$\text{Recall} = (TP)/\ (TP + FN)$$

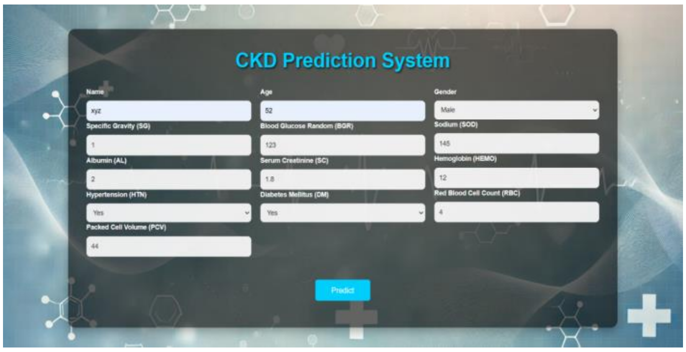
4) F1-Score: The harmonic mean of precision and recall, balancing false positives and false negatives.

$$F1\text{-Score} = 2(\text{Precision} * \text{Recall})/\ (\text{Precision} + \text{Recall})$$

5) Confusion Matrix: A table that shows the number of TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives), helping visualize classification performance.

G. User-Friendly Web Interface

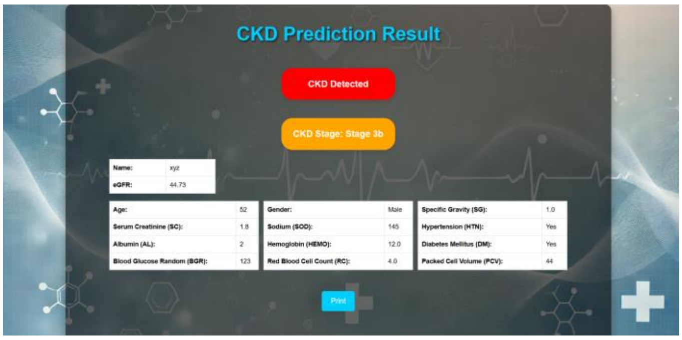
A user-friendly web interface was developed, allowing users to input patient data for CKD detection. Since the Random Forest model achieved the highest accuracy, it was selected as the final model. If the model detects CKD, the CKD-EPI equation is used to calculate the estimated glomerular filtration rate (eGFR). Based on the eGFR value, the CKD stage is determined, and the final output is displayed to the user. The front end of the application was designed using HTML, CSS, and JavaScript, ensuring an intuitive and responsive user experience. These technologies allow for an interactive and visually appealing interface where users can easily input patient details. On the back end, Flask, a lightweight and powerful Python web framework, was utilized to handle data processing, model inference, and communication between the user interface and the machine learning model. Flask efficiently processes input data, passes it to the trained model for prediction, and returns the results to the user in real time.



CKD Prediction System

Name: xyz
Age: 52
Gender: Male
Specific Gravity (SG): 1
Blood Glucose Random (BGR): 123
Sodium (SOD): 145
Albumin (AL): 2
Serum Creatinine (SC): 1.8
Hemoglobin (HEMO): 12
Hypertension (HTN): Yes
Diabetes Mellitus (DM): Yes
Red Blood Cell Count (RBC): 4
Packed Cell Volume (PCV): 44

Fig- 7: User Input Form



CKD Prediction Result

CKD Detected

CKD Stage: Stage 3b

Name:	xyz
eGFR:	44.73

Age:	52	Gender:	Male	Specific Gravity (SG):	1.0
Serum Creatinine (SC):	1.8	Sodium (SOD):	145	Hypertension (HTN):	Yes
Albumin (AL):	2	Hemoglobin (HEMO):	12.0	Diabetes Mellitus (DM):	Yes
Blood Glucose Random (BGR):	123	Red Blood Cell Count (RC):	4.0	Packed Cell Volume (PCV):	44

Fig- 8: Prediction Result Page

IV. Result And Discussion

All models are trained on 80% of the complete dataset, and testing is done on the remaining 20% of the dataset. Random Forest, MLP, and MLP + XGBoost achieved 100% accuracy, and XGBoost achieved 98.75% accuracy.

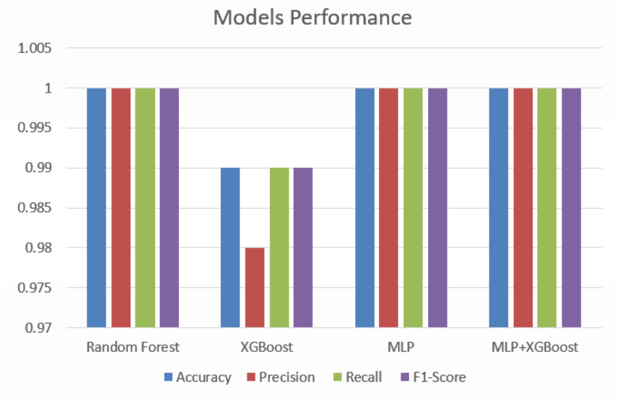
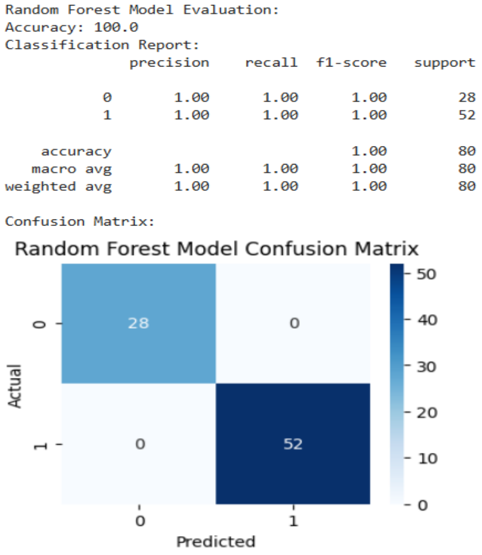
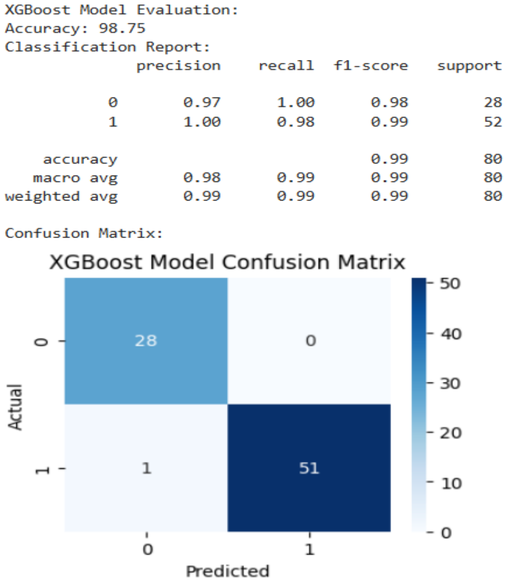


Fig- 9: models' performance



```
Random Forest Model Evaluation:
Accuracy: 100.0
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        28
           1       1.00      1.00      1.00        52

    accuracy                           1.00        80
   macro avg       1.00      1.00      1.00        80
weighted avg       1.00      1.00      1.00        80

Confusion Matrix:
```

Fig- 10: Random Forest Evaluation Result

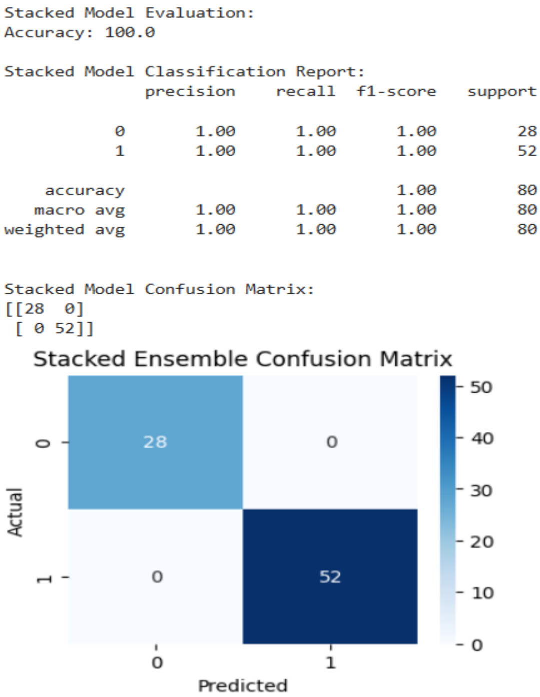


```
XGBoost Model Evaluation:
Accuracy: 98.75
Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.98        28
           1       1.00      0.98      0.99        52

    accuracy                           0.99        80
   macro avg       0.98      0.99      0.99        80
weighted avg       0.99      0.99      0.99        80

Confusion Matrix:
```

Fig- 11: XGBoost Evaluation Result

Table- 3: COMPARISON OF WORKS WITH SAME DATASET

Authors	Models	Accuracy
Rady and Anwar [1]	PNN	96.7%
S. Tekale et al. [2]	SVM	96.75%
K.M. Almustafa [3]	J48, DT	99%
M.A. Islam et al. [4]	XGBoost with PCA	99.16%
Poonia et al. [5]	LR	98.75%
our work	RF, stack model	100%



```
Stacked Model Evaluation:
Accuracy: 100.0

Stacked Model Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        28
           1       1.00      1.00      1.00        52

    accuracy                           1.00        80
   macro avg       1.00      1.00      1.00        80
weighted avg       1.00      1.00      1.00        80

Stacked Model Confusion Matrix:
[[28  0]
 [ 0 52]]
```

Fig- 12: Stack Model Evaluation

V. Conclusion

Early detection is essential for both healthcare professionals and patients to prevent and delay the progression of chronic kidney disease, reducing the risk of kidney failure. This work, "Chronic Kidney Disease Prediction and Stage Classification," successfully developed a comprehensive machine learning framework for the early detection and stage classification of CKD. By leveraging advanced ensemble learning models as Random Forest, which achieved 100% accuracy, XGBoost 98.75%, and MLP with XGBoost 100%, since “Random Forest model” achieved a higher accuracy, it is deployed into a user-friendly website that enables users or healthcare professionals to enter patient details and receive CKD predictions along with stage classifications, facilitating early detection and effective disease management.

VI. Future Scope

Integration with AI-driven Healthcare Platforms: Combining the CKD prediction model with other AI tools, such as imaging analysis (e.g., kidney scans) or symptom analysis, could create a more comprehensive diagnostic system that supports healthcare professionals in decision-making.

Additionally, expanding the dataset by incorporating more patient records would enhance model accuracy, robustness, and generalizability, leading to improved predictive performance and better patient outcomes.

VII. Reference

[1] El-Houssainy A. Rady, Ayman S. Anwar. Prediction of kidney disease stages using data mining algorithms 2019. https://doi.org/10.1016/j.imu.2019.100178

[2] Tekale S, Shingavi P, Wandhekar S, Chatorikar A. Prediction of chronic kidney disease using machine learning algorithm. Disease. 2018;7(10):92–6.

[3] Prediction of chronic kidney disease using different classification algorithms. Inform Med Unlocked, 24 (2021), Article 100631, 10.1016/j.imu.2021.100631

[4] Islam MdA, Majumder MdZ, Hussein MdA. Chronic kidney disease prediction based on machine learning algorithms. J Pathol Inform 2023; 14:100189. https://doi.org/10.1016/j.jpi.2023.100189

[5] Debal DA, Sitote TM. Chronic kidney disease prediction using machine learning techniques. J Big Data 2022;9(1). https://doi.org/10.1186/s40537-022-00657-5

[6] R.C. Poonia, M.K. Gupta, I. Abunadi, *et al.* Intelligent diagnostic prediction and classification models for detection of kidney disease. Healthcare, 10 (2022), p. 371, 10.3390/healthcare10020371

[7] Dr.Venkata Kishore Kumar Rejeti, "Distributed Denial of Service Attack Prevention from Traffic Flow for Network Performance Enhancement", Proceedings of the Second International Conference on Smart Electronics and Communication (ICOSEC) in IEEE. DVD Part Number: CFP21V90-DVD; ISBN: 978-1-6654-3367-9, 2021.