Smart Diagnosis: Hybrid Ensemble Learning For Early Detection of Chronic Kidney Disease Using Clinical Data

Sana Batool[1], Shahnawaz Sheikh[2], Muhammad Ovais Akhter[3]*, Shakir Karim Buksh[4], Abdullah Ayub Khan[5], Muhammad Zohaib Khan[6]

Abstract: **Chronic Kidney Disease (CKD) is a progressive and life-threatening condition that affects millions of people worldwide and is often diagnosed at advanced stages due to the absence of noticeable early symptoms. Early and accurate detection is essential to delay disease progression, reduce complications, and improve patient outcomes. This paper proposes a smart hybrid machine learning framework for the early prediction of CKD using clinical data. The proposed approach integrates Agglomerative Hierarchical Clustering (AHC), Synthetic Minority Over-sampling Technique (SMOTE), Stacking Ensemble Classification (SEC), and Logistic Regression (LR) to address class imbalance and enhance prediction performance. Several machine learning classifiers, including K-Nearest Neighbors (K-NN), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Quadratic Discriminant Analysis (QDA), and Decision Tree (DT), are incorporated into the hybrid framework and comparatively evaluated. Experimental results demonstrate that the proposed hybrid models consistently outperform individual classifiers, with the AHC–SEC-LR–K-NN model achieving the highest classification accuracy of 97.59%, followed by AHC–SEC-LR–MLP (96.39%) and AHC–SEC-LR–SVM (95.90%). The findings indicate that the proposed framework provides a reliable and effective decision-support tool for the early diagnosis of CKD and has significant potential for integration into intelligent healthcare and telemedicine systems to support timely clinical decision-making.**

Keywords: **chronic kidney disease prediction, ensemble machine learning, logistic regression, stacking ensemble classifier, agglomerative hierarchical clustering, support vector machine, decision tree.**

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a silent yet devastating global health challenge that progressively affects the lives of millions of individuals worldwide. It is characterized by a gradual and irreversible decline in kidney function, affecting more than 10% of the global population. The major concern associated with CKD is its asymptomatic nature during the early stages, which prevents timely diagnosis and reduces the possibility of effective intervention. Furthermore, CKD is strongly associated with an increased risk of cardiovascular disease (CVD), which contributes significantly to higher morbidity and mortality rates among affected individuals [1, 2].

[1,2] Riphah International University, Faisalabad
[3,4,5] Bahria University Karachi Campus
[6] Shaheed Mohtarma Benazir Bhutto Institute of Trauma, Karachi
Country: Pakistan
Email: *ovaisakhter.bukc@bahria.edu.pk

As the disease progresses, it may lead to End-Stage Renal Disease (ESRD), a critical condition requiring renal replacement therapies (RRT), including dialysis or kidney transplantation. This progression not only places a significant burden on healthcare systems but also negatively impacts patients' quality of life and survival. Therefore, the latent progression of CKD highlights the importance of early diagnosis, continuous monitoring, and accurate prediction of disease progression.

Considering these challenges, early detection and prediction of CKD progression have become increasingly important [3]. Although conventional diagnostic approaches are clinically valuable, they may not always identify high-risk individuals before significant kidney damage occurs. Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), and data-driven analytics have introduced promising opportunities for improving CKD diagnosis, management, and treatment. These intelligent approaches enable healthcare professionals to analyze complex clinical datasets, identify hidden patterns, and develop predictive models for accurate risk assessment. ML algorithms can process large-scale medical data and detect complex relationships among different clinical variables, supporting early identification of high-risk patients. Consequently, AI-based approaches can facilitate personalized treatment strategies, improve patient monitoring, reduce late-stage interventions, and enhance overall healthcare outcomes [4, 5].

In recent years, ML techniques have demonstrated significant potential in improving early CKD detection and risk prediction. These algorithms are capable of handling large healthcare datasets and identifying non-linear relationships that may remain unnoticed through conventional statistical approaches. Such capabilities enable accurate classification of disease stages and early identification of patients at risk of rapid kidney deterioration. Since CKD is associated with a chronic and progressive decline in kidney function, predictive ML-based approaches provide an effective mechanism for addressing this global healthcare challenge.

Several factors, including diabetes, hypertension, and hereditary disorders, contribute to kidney dysfunction. The kidneys play a vital role in removing waste products, toxins, and excess fluids from the blood; however, the gradual progression of CKD often remains undetected during its initial stages. According to the WHO, CKD affects millions of people worldwide, with many patients remaining unaware of their condition until irreversible damage occurs. Therefore, early detection using advanced ML techniques is essential to slow disease progression, reduce healthcare costs, and improve patient survival and quality of life [6, 7]. The application of ML in healthcare has significantly enhanced the accuracy, efficiency, and reliability of disease diagnosis,

particularly for CKD prediction. Unlike traditional diagnostic methods that often require extensive clinical assessment, ML-based systems can evaluate risk factors using limited inputs, including blood test results, patient history, and demographic information. These capabilities make ML an effective tool for reducing diagnostic delays and supporting personalized healthcare strategies [8].

The main goal of this paper is to develop an intelligent solution for early CKD prediction by applying advanced machine learning techniques. The study aims to improve CKD detection accuracy and diagnostic reliability by evaluating and comparing different Ensemble Machine Learning (EML) methods. Furthermore, the performance of ensemble learning approaches, particularly LR and SEC, is optimized to achieve more accurate CKD prediction. The proposed approach also focuses on developing a decision-support system that can assist healthcare professionals, including physicians and paramedical staff, in identifying CKD at an early stage and providing timely personalized treatment strategies. Additionally, the study investigates classification and clustering approaches by comparing supervised and unsupervised learning techniques for CKD identification.

This paper proposes an advanced framework for early CKD detection using EML algorithms, particularly focusing on the optimization of SEC and LR models. SEC is an effective ensemble learning approach that improves model generalization by combining predictions from multiple base classifiers through a meta-classifier. In contrast, Logistic Regression is a widely used statistical method for binary classification that estimates the probability of class membership. To comprehensively evaluate different learning strategies, both supervised and unsupervised approaches are incorporated. AHC is applied as an unsupervised learning technique that creates hierarchical clusters by repeatedly combining similar data points. Meanwhile, supervised learning algorithms, including K-NN, MLP, SVM, and QDA, are employed for CKD classification. K-NN classifies samples based on the majority vote of neighboring instances, MLP identifies complex non-linear patterns through interconnected neural layers, SVM determines an optimal decision boundary between classes, and QDA performs classification based on probabilistic distribution assumptions.

These models are evaluated through classification and clustering approaches to effectively differentiate CKD-positive and CKD-negative cases. The proposed framework aims to improve prediction performance, reduce false-positive and false-negative rates, and provide a reliable solution for early CKD diagnosis. By optimizing ensemble-based models, this research highlights the potential of ML-driven approaches in reducing diagnostic challenges and supporting cost-effective healthcare strategies. The remainder of this paper is organized as follows: Section 2 presents a review of existing CKD prediction methods. Section 3 describes the proposed methodology, including the dataset and experimental framework. Section 4 discusses the experimental results and findings, while Section 5 concludes the paper and presents future research directions.

II. LITERATURE REVIEW

Many papers have tested the use of ML technologies to predict CKD, using different datasets and methodological frameworks. Altogether, this research literature has demonstrated the prospects of ML models to improve diagnostic accuracy to a substantial extent, assist in studying early signs of disease, and offer some useful information on risk classification and clinical decision making. Boukenze et al. [9] reported the performance of some of the ML classifiers: SVM, C4.5, bayesian network (BN), and K-NN on WEKA platform in CKD prediction. According to their findings, MLP and C4.5 performed the best, and C4.5 was much higher than other models in terms of ROC analysis. Their work showed promise of ML in CKD diagnosis; however, it also showed shortcomings of including only one dataset and lacking clinical validation, and thus their research implies that further testing in a wide range of populations is necessary. Jena et al. [10] conducted a similar study that highlighted the global burden of CKD and the significance of early diagnosis in reducing the progression and other complications of the disease. They have found that ML is an effective means of processing complicated patient data, especially because CKD is asymptomatic in the early stages. Their review provided an overview of existing methods of diagnostics and the changing role of ML in the healthcare sector, as well as defining the most critical gaps in the research that should be filled through additional research.

The system that was suggested by Mohammed and Beshah [11] is a self-learning knowledge-based system (KBS) used to diagnose and manage early CKD. By applying a limited set of data and a Decision Tree algorithm, they managed to come up with a prototype that can provide users with personalized medical recommendations. It demonstrated a high level of accuracy (91) of the system, which demonstrates the potential of smart tools in improving patient engagement and promoting early CKD detection. Xie et. al. [12] examined how combining renal biomarkers such as the urine albumin to creatinine ratio and cystatin C-based estimated glomerular filtration rate (eGFR) could complement the traditional clinical parameters to forecast heart failure (HF). Their results illustrated that cystatin C-based eGFR most significantly increased the accuracy of reclassification, which is significant as kidney dysfunction contributes to the development of HF and should support the integration of nephrology and cardiology. Equally, Priyanka et al. [13] applied Naive Bayes algorithm in predicting CKD and compared it with some of the ML methods such as ANN, SVM, K-NN, and DT. NB is the most accurate in the prediction, with a rate of 94.6, and it also indicated that probabilistic classifiers are highly applicable in the diagnosis of CKD using structured clinical data. Xiao et al. [14] compared several ML algorithms including Logistic Regression, Elastic Net, Lasso and Ridge Regression, SVM, Random Forest, XGBoost, K-NN, and neural networks to create a model to predict CKD progression. They categorized the severity of the disease using clinical data of 551 patients with proteinuria and 18 features as mild, moderate, and severe outcomes. The best outcome was obtained with LR having AUC of 0.873, sensitivity of 0.83 and specificity of 0.82, justifying its usefulness in clinical risk classification of CKD.

Considering that ML is highly adaptive, it is especially adapted to chronic illnesses, such as CKD, as data regarding patients is dynamic and longitudinal. Amirgaliyev et al. [15] used SVM to classify CKD patients with the help of clinical features and showed impressive results on the test dataset, a 93 percent accuracy, sensitivity, and specificity. Their results supported the possibility of SVM to improve the diagnostic accuracy and disease monitoring over time. Finally, Tsai et al. [16] created a high-performing CKD risk prediction model in Thailand with the help of the RF

algorithm, which had the ability to predict with 92.1% accuracy. To improve model transparency and clinical interpretability, they added SHapley Additive exPlanations (SHAP) that allowed them to learn dual-attribute risk factors, which contributed to a more explainable AI-powered diagnostic tool. Moreno-Sánchez et al. [17] also introduced the explainable artificial intelligence (XAI) paradigm in the context of the initial diagnosis of CKD; in their case, the authors revealed the predictive power of hemoglobin levels, specific gravity, and hypertension. Besides being extremely accurate, this model could provide a cost-effective way of diagnosis and was particularly effective in those environments where resources were limited. Overall, the results indicate that interpretable AI is necessary to diagnose CKD by balancing the complexity of the algorithms and making well-informed and ethical healthcare decisions. More recent studies have proposed hybrid and ensemble methods to improve prediction strength, as is also shown by Ghosh et al. et al. [18]. In this case, the hybrid model that used Naive Bayes, Decision Tree, and Random Forest were better than single classifiers as they achieved an accuracy of 94.99%. Besides, Zhu et al. [19]. Advanced ML use by applying XGBoost to forecast CVD in individuals with CKD. This demonstrated the importance of such features as age, hypertension and eGFR and reached AUC of 0.893. Collectively, the findings indicate that ML has an even greater promising potential in enhancing the early CKD and comorbidity detection and supporting clinical decision-making, especially when it comes to ensemble and explainable models (as shown in Table 1).

Table 1: Benchmarking Accuracy of Hybrid Models in Historical CKD Research

Author(s)	Method/ Model	Dataset	Performance/ Results	Key Findings / Contributions
Boukenze et al. [9]	MLP, SVM, C4.5, K-NN using WEKA, BN	UCI CKD Dataset	MLP & C4.5 best; C4.5 top via ROC	ML's usefulness was demonstrated, but its limits as a single dataset were brought to light.
Jena et al. [10]	Review of ML in CKD diagnosis	Literature Review	-	Highlighted the significance of data-driven diagnostics and identified the relevance of ML in early CKD detection.
Mohammed & Beshah et al. [11]	Knowledge-Based System (KBS) with Decision Tree	Small proprietary dataset	Accuracy: 91%	Created a prototype that allowed for interactive patient involvement and early CKD diagnosis.
Xie et al. [12]	Random Forest + renal biomarkers + SHAP	Thai clinical dataset	Accuracy: 92.1%	Risk prediction is greatly improved by renal biomarkers, and interpretability was enhanced by SHAP.
Priyanka et al. [13]	Naïve Bayes, K-NN, SVM, DT, ANN	-	Naïve Bayes: 94.6%	When it came to organized clinical data, Naïve Bayes performed better than the others.
Xiao et al. [14]	LR, Elastic Net, Lasso, Ridge, SVM, RF, XGBoost, NN, K-NN	551 patients (18 features)	LR AUC: 0.873; Sens/Spec: 0.83/0.82	The most reliable method for predicting the course of CKD was logistic regression.
Amirgaliyev et al. [15]	SVM with clinical features	Clinical testing dataset	Accuracy: 93%	SVM classified early CKD with great accuracy and generalizability.
Tsai et al. [16]	Random Forest with SHAP	Thailand CKD data	Accuracy: 92.1%	Identified important risk variables using SHAP; demonstrated excellent performance in a real-world scenario.
Moreno-Sánchez et al. [17]	Explainable AI (XAI) model	-	High Accuracy	Highlighted models that could be understood; hemoglobin, specific gravity, and hypertension were shown to be crucial.
Ghosh et al. [18]	Hybrid Model: Naïve Bayes + DT + RF (stacked ensemble)	UCI CKD Dataset	Accuracy: 94.99%	The suggested hybrid model improved prediction accuracy and robustness while outperforming base models.
Zhu et al. [19]	XGBoost for CVD prediction in CKD patients + SHAP	8,894 real clinical records (China)	AUC: 0.893; Accuracy: 0.806	ML for predicting comorbidity; scalability was shown; important features (e.g., age, HTN, and eGFR) were discovered.

Using EML and logistic regression (LR) approaches, predictive modeling is used in the biomedical industry to create forecasting or classification algorithms that aid in computer-assisted clinical decision-making. These models employ a collection of input parameters referred to as features or attributes and concentrate on a certain result (or target variable). Accurate predictions or classifications are made possible by ML algorithms that examine these features and reveal hidden patterns that correlate with the result. By examining patient data gathered over time, supervised classification models can be trained in telemonitoring to differentiate between stable health situations and those that are susceptible to deterioration. Daily tele-monitoring data are used as features in this context, and the goal variable for prediction is the exacerbation status (CKD-Positive or CKD-Negative) for each day.

III. PROPOSED METHODOLOGY

The study employs contemporary ML in a systematic and orderly way to predict CKD. The detailed Work Breakdown Structure (WBS) of the CKD shown in Figure 1 gives an example of the proposed methodology:

1. The dataset presented in this study was offered by Google Drive. It also contains patient information and the relevant clinical variables related to CKD.
2. Some preprocessing steps were undertaken to ensure optimal model performance. Data cleaning was the initial process that entailed elimination or filling of gaps or inconsistency values.
3. The analysis of importance of features was then carried out through the Random Forest algorithm to identify and keep the most important characteristics that have the most significant effect on the prediction of CKD. Lastly, feature distributions were placed through the normalization process which enhanced numerical stability and cross-variable comparability.
4. The dataset was further refined using the AHC unsupervised learning method. This improved classification measures such as specificity, F1-score, precision, recall, sensitivity and the overall model accuracy by identifying organic groups in the data.
5. To simplify model training and evaluation, the processed dataset was divided into two subsets. To enable the model to identify background trends, 75 percent of the data was set aside as a training set. The remaining quarter was the test data to test the model on the untested data to ensure that the model is applicable to other unrelated data.
6. SMOTE was only applied to the training set to deal with the issue of class imbalance that is usually observed in medical datasets. In this method, fictitious samples are developed of the minority class so that the model is not prejudiced; it favors the majority class.
7. Some of the ML models employed to develop and evaluate credible prediction models of CKD include SEC, MLP, LR, AHC, QDA, K-NN, SVM and DT. This diversified set of algorithms was carefully selected to make the model more reliable, enhance the accuracy of diagnosing the dataset, and find complex patterns.

The design of the CKD system is presented in Figure 1 that also contains details about the task breakdown structure of the research.

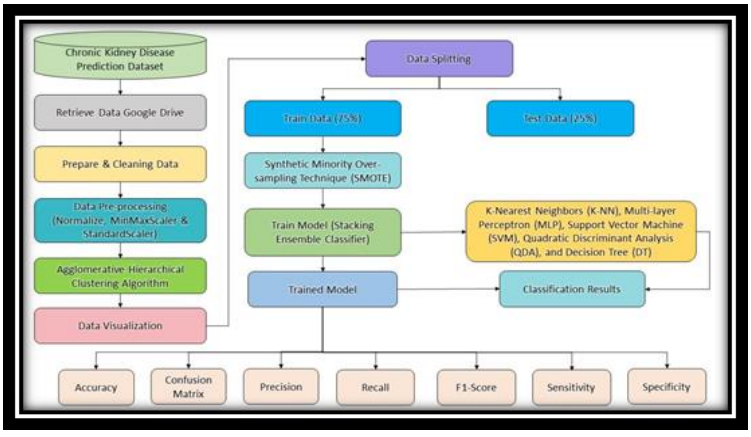


Figure 1: Proposed method for chronic kidney disease

Comprehensive model validation was done using the processed dataset that resulted in 75 percent of the dataset being allocated to training and 25 percent to testing. In order to ensure the quality and consistency of the data, the raw data was initially patched out of the database, and properly polished with much standardization and preprocessing, as appears in the flowchart that accompanies it. The two crucial aspects of the workflow data preparation and classification, in turn, influence the overall performance of a model considerably. This is a purposeful and systematic approach to methodology that increases the strength of the model as well as its generalizability to a range of diverse clinical situations in addition to the accuracy of the predicted.

1. Preprocessing

The dataset undergoes a data cleaning stage where the data is analyzed thoroughly using clustering methods to identify meaningful patterns. To reveal hidden structures in the data, Agglomerative Hierarchical Clustering (AHC) was selected to cluster similar data points in the data based on their inherent attributes. This unstructured learning plan, described further on, is critical towards organizing and simplifying the data. To obtain accurate and insightful predicting modeling, the fined output of this step is further enhanced and ready to duty the next classification stage.

A. Data collection

The Chronic Kidney Disease Prediction (CKDP) data was available on Kaggle. The 1,659 patient records on the popular site of open-access ML data sets depict 25 distinct features that denote the various clinical and behavioral variables in relation to CKD. Having 1,524 cases as CKD-Positive and 135 cases as CKD-Negative, the target variable classifies people into two categories CKD-Positive and CKD-Negative [20]. This rich and diverse data gives an excellent start to the development of advanced predictive models and performing comprehensive healthcare analytics. It captures significant variables like blood urea nitrogen levels, serum creatinine, glomerular filtration rate (GFR), and lifestyle habits, including smoking and NSAID use. The Table 2 offers a detailed description of the key nature of the data set, which can be useful to the researchers in their quest to enhance early diagnosis and treatment of CKD (as exhibited in Table 2).

Table 2: Displays the original dataset utilized for predicting CKD diagnosis

Parameters of the Dataset	Characteristics of CKDP
PatientID	Unique identifier for each patient
Age	Age of the patient (in years)
Gender	Gender of the patient (e.g., Male, Female)
BMI	Body Mass Index (kg/m²)
Smoking	Smoking status (e.g., Yes, No)
AlcoholConsumption	Alcohol consumption status (e.g., Yes, No)
GFR	Glomerular Filtration Rate (mL/min/1.73 m²)
SerumCreatinine	Serum creatinine level (mg/dL)
BUNLevels	Blood Urea Nitrogen level (mg/dL)
ACR	Albumin-to-Creatinine Ratio (mg/g)
ProteinInUrine	Presence of protein in urine (e.g., Normal, Trace, High)

HbA1c	Glycated Hemoglobin percentage (used for diabetes monitoring)
FastingBloodSugar	Fasting blood glucose level (mg/dL)
SystolicBP	Systolic blood pressure (mmHg)
DiastolicBP	Diastolic blood pressure (mmHg)
PreviousAcuteKidneyInjury	History of acute kidney injury (e.g., Yes, No)
FamilyHistoryKidneyDisease	Family history of kidney disease (e.g., Yes, No)
NSAIDsUse	Use of Non-Steroidal Anti-Inflammatory Drugs (e.g., Yes, No)
SerumElectrolytesSodium	Sodium level in serum (mEq/L)
SerumElectrolytesPotassium	Potassium level in serum (mEq/L)
SerumElectrolytesCalcium	Calcium level in serum (mg/dL)
SerumElectrolytesPhosphorus	Phosphorus level in serum (mg/dL)
ACEInhibitors	Use of ACE inhibitors (e.g., Yes, No)
Diuretics	Use of diuretics (e.g., Yes, No)
Diagnosis	Final CKD diagnosis (e.g., CKD, Non-CKD)

Figure 2 shows a graphic comparison of the prevalence of (CKD) in men and women. The x-axis is represented by the males as 0 and the females as 1. The results indicate that the prevalence of (CKD) was much higher among women, indicating that it is possible that there is a gender difference in the occurrence of the condition. It is a major consideration to make on the predictive modeling and targeted healthcare interventions. Besides, the distribution of CKD patients according to diagnosis category is presented in one more field of Figure 3, where 0 indicates CKD-Negative cases, and 1 indicates CKD-Positive cases. The vast difference between the bar heights reflects the much higher percentage of patients with CKD, and this fact indicates the dire necessity of the early diagnosis methods. Furthermore, Figure 4 displays the age distribution of patients, and it gives an understanding of the age groups that have been affected the most by CKD. All these visual indicators help in selecting the most relevant features that can be used in the training of a model with less relevant features being removed in the next stage of feature optimization.

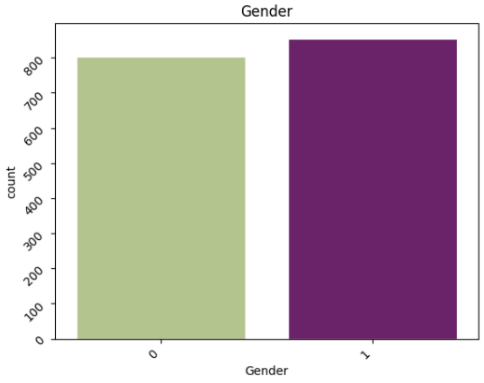


Figure 2: Quantity of facts male and female CKD patients

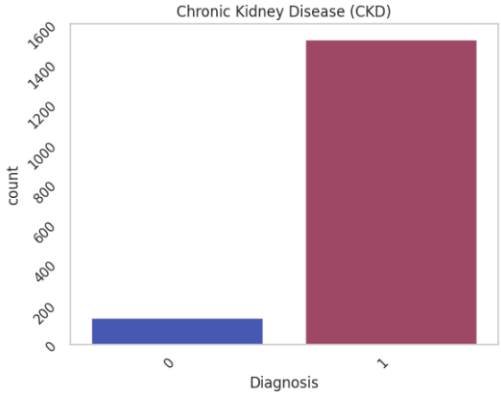


Figure 3: Quantity of CKD patients

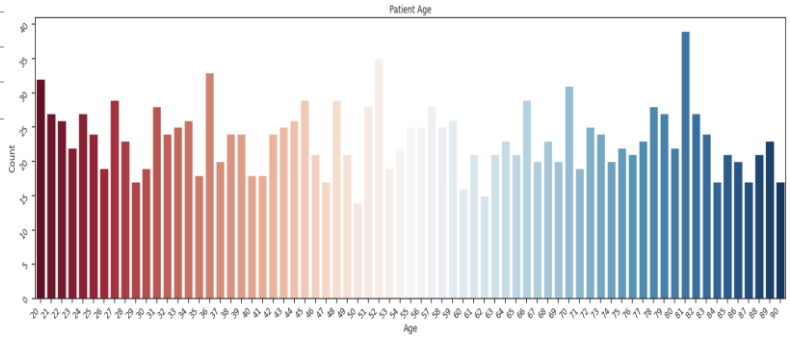


Figure 4: Number of CKD patients age wise.

B. Feature importance

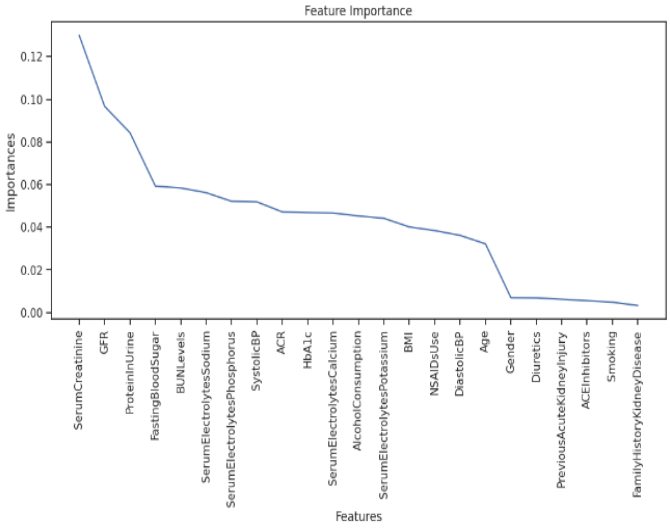
The input variables receive scores according to the importance of the features that reveal the degree to which they influence the predictions of a model. The higher the scores the more relevant. In this study, the best features were selected with the help of Recursive Feature Elimination (RFE) as an iterative algorithm that ranks and eliminates features based on their influence. This enhanced RFE approach is more than the traditional methods because multiple algorithms are combined with XGBoost, SVM, and Random Forest to combine mathematical evaluation with strategic variable analysis and predictive performance. This multi-algorithm method enhances the efficiency of the model, and it ensures the accurate selection of features [21, 22].

Significance of the features in the CKD model, along with significant scores, is further disaggregated in Table 3 that brings out the contribution made by each feature to the model accuracy. This tabular analysis will help to understand better the clinical and lifestyle factors which produce the strongest tendency of the prediction results. Moreover, Figure 3 illustrates the same scores visually, which can easily help one to interpret the meaning of each attribute. Table 3, combined with Figure 5, represents a detailed feature analysis of importance, which offers valuable data to make a model more efficient and guide future research on CKD prediction (as shown in Table 3).

Table 3: Importance dataset feature for CKDP

S. no.	Feature	Importance Score Feature
1	SerumCreatinine	0.130010
2	GFR	0.096711
3	ProteinInUrine	0.084314
4	FastingBloodSugar	0.059279

5	BUNLevels	0.058417
6	SerumElectrolytesSodium	0.056214
7	SerumElectrolytesPhosphorus	0.052169
8	SystolicBP	0.051890
9	ACR	0.047182
10	HbA1c	0.046866
11	SerumElectrolytesCalcium	0.046684
12	AlcoholConsumption	0.045293
13	SerumElectrolytesPotassium	0.044193
14	BMI	0.040151
15	NSAIDsUse	0.038443
16	DiastolicBP	0.036193
17	Age	0.032204
18	Gender	0.006966
19	Diuretics	0.006885
20	PreviousAcuteKidneyInjury	0.006229
21	ACEInhibitors	0.005559
22	Smoking	0.004847



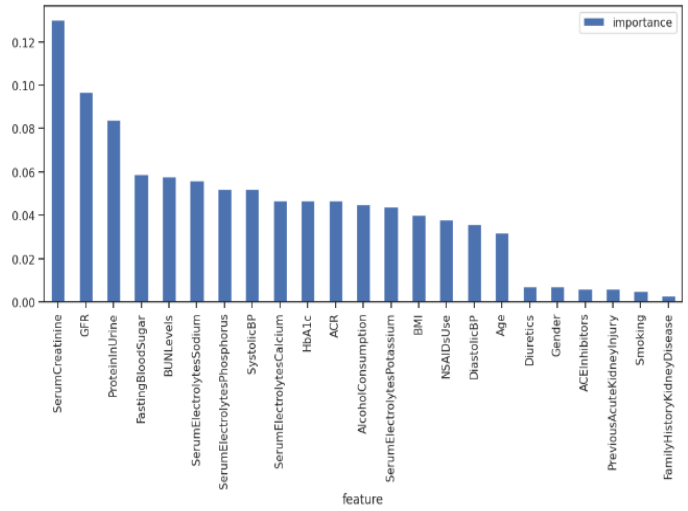


Figure 5: Feature importance dataset of (CKDP)

Spotting key steps like data normalization, preprocessing, and simulation, as well as induction requirements, this paper looks at the rudimentary methods of CKD prognostication. It further examines after-processing methods, test criteria, variables of complexity and the overall effectiveness of the prediction system. In order to ensure consistency and model readiness, the process begins with the acquisition of raw data which is then enhanced by standardization and normalization. The clean and arranged data are summarized in Table 4. Also, Figure 6 provides a graphical representation of complex relationships in the data without the application of Agglomerative Hierarchical Clustering (AHC). In order to distinguish pattern of features easily, data point X would be plotted in this image as a black circle, and point Y would be a light pink circle (shown in Table 4).

Table 4: Shows the dataset for predicting CKD

25-Dimension
[[-3.58728672 3.57858931 -0.81003711 ... -0.54317401 -0.57295377 2.93354214]
[-3.56911771 1.07981974 -0.81003711 ... -0.54317401 -0.57295377 3.12691778]
[-3.56124411 3.38064585 2.79950211 ... -0.54317401 3.5802017 2.27110426]...
[0.6940669 -0.49813297 -0.81003711 ... -0.54317401 0.18269251 -1.55190203]
[0.67670253 -0.43725846 -0.81003711 ... -0.54317401 -0.57295377 -0.85955276]
[0.70739589 -1.04267391 -0.15205668 ... -0.54317401 0.18412209 -0.85500884]]

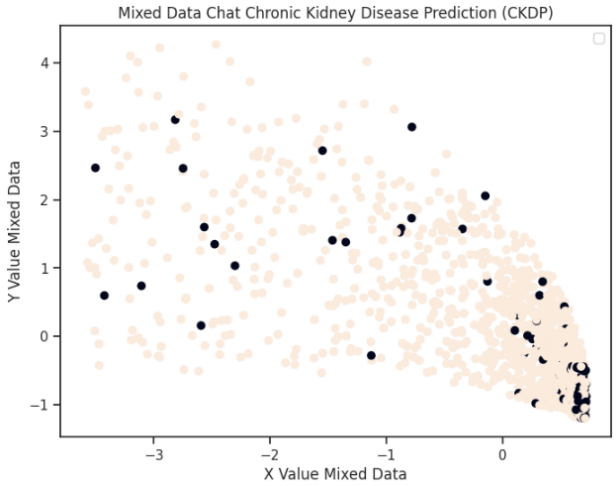


Figure 6: Mixed data chat of (CKDP)

C. Agglomerative hierarchical clustering method

Clustering is a significant unsupervised learning which means that related data points are grouped together without the use of predefined labels. The Agglomerative Hierarchical Clustering (AHC), also called AGNES (Agglomerative Nesting), is a bottom-up method that initially takes each point of data as a cluster. It then repeatedly combines the nearest two clusters according to a metric of similarity and the process goes on until only one, all-inclusive cluster is created. The resultant data is a dendrogram a three-like diagram, which visually displays the hierarchical grouping of information and the sequence in which data clusters together. Users may identify valuable groupings by the dendrogram by slicing them at a given height based on certain criteria. An identifier is then used to identify each resultant cluster to be used in subsequent analysis or interpretation. This hierarchical method provides dynamically and flexible partitioning of the clusters where the user can determine the level of granularity in the cluster by determining the level at which to separate the dendrogram [23, 24, and 25].

- To quantify how similar each data point is, calculate the proximity matrix with some appropriate distance measure (e.g. the Manhattan distance or the Euclidean distance).
- Begin construction of an organizational cluster tree (dendrogram) by using a linking function on the distance matrix.
- The connection criterion is applied to merge the two nearest data points or clusters into a new cluster.
- Repeat steps two and three several times, successively clustering the data points together until all the data has been clustered into one root.
- Once the last group of all the observations has been created, they terminate the operation.

The method of preprocessing the information by organizing it, analyzing the information at a high level is Agglomerative Hierarchical Clustering (AHC), and a unified method of data collection is applied to generate mixed-data representation. This will enhance interpretability, reduce repetition and reveal important patterns in the data. Clustering divides the dataset into two distinct clusters and assigns a probability value to every data point, this value indicating the likelihood of a data point being part of a particular cluster. The final product of the process is a matrix that correlates the samples and clustered samples. All the points in this work represent centroid, the shorter the distance, the higher the connection. AHC algorithm is applied to a dataset that consists of 25 characteristics and binary-class. It can be applied to both cluster detecting and the software-related problems prediction on the Prediction (CKDP) dataset that contains 25 critical variables. The AHC model splits the data visually between benign and malignant, and the results of the graphical representation are understandable. Figures 7 and 8 present the agglomerative clustering results of CKDP once the raw and unstructured data had been transformed into a form that was easy to understand and interpret (as in Table 5).

Table 5: Agglomerative Hierarchical Clustering Two Clusters Pre-processed CKDP Dataset

25-Dimension
[[-3.58728672 3.57858931 -0.81003711 ... -0.54317401 0.57295377 2.93354214 1]
[-3.56911771 1.07981974 -0.81003711 ... -0.54317401 0.57295377 3.12691778 1]
[-3.56124411 3.38064585 2.79950211 ... -0.54317401 3.58020175 2.27110426 1] ...
[0.6940669 -0.49813297 -0.81003711 ... -0.54317401 0.18269251 -1.55190203 0]
[0.67670253 -0.43725846 -0.81003711 ... -0.54317401 0.57295377 0.85955276 0]
[0.70739589 -1.04267391 -0.15205668 ... -0.54317401 0.18412209 -0.85500884 1]]

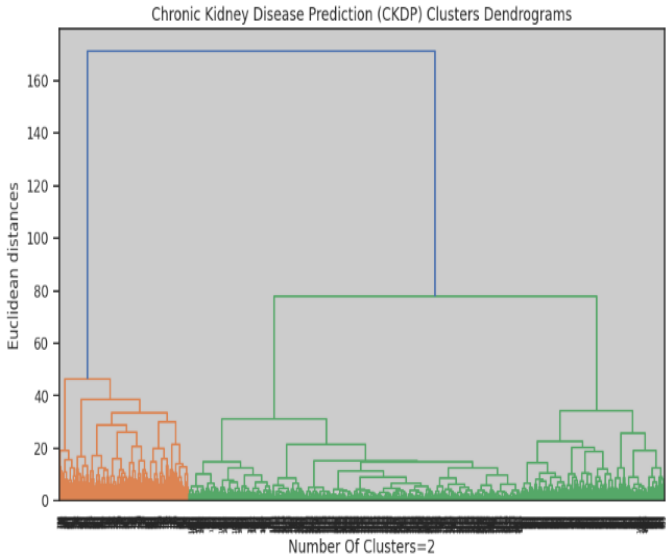


Figure 7: Agglomerative hierarchical clustering dendrograms (CKDP)

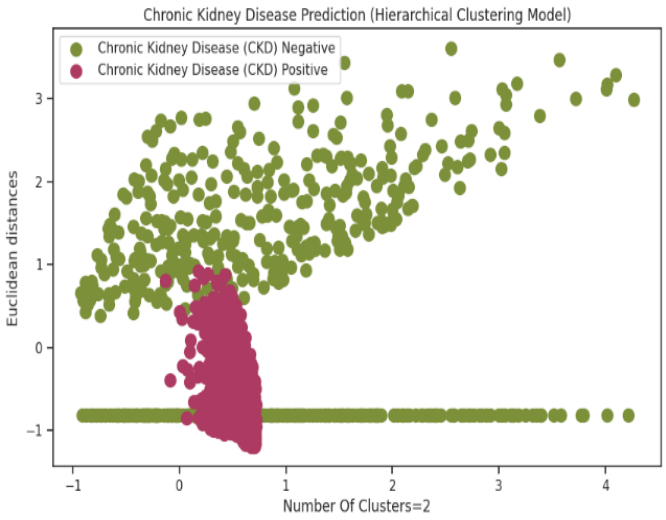


Figure 8: Agglomerative Hierarchical Clustering Two Clusters (CKDP)

Such metrics as F-measure, accuracy and recall are taken to evaluate the extent to which the clustering feature detection was applied effectively using various clustering methods. These measurements offer a detailed analysis of the capacity of the system to identify and categorize the associated elements of data. A worry score is also calculated by calculating the deviation of the system against the expected norms in order to categorize the outcomes into legitimate, doubtful and anomalous results. This approach will be able to check proper data grouping and interpretation since it will quantify the clustering accuracy and enhance the ability of the system to detect patterns and anomalies according to pre-influenced benchmarks.

2. CLASSIFICATION

In order to find the best accurate classification model for the provided dataset, the classification process—which is powered by training data—focuses on classifying observed data into specified categories. It uses labeled samples to learn decision limits as a type of supervised learning. Several classification algorithms are used and experimentally evaluated in order to choose the best model, enabling a thorough comparison of their performances. This assessment aids in identifying the classifier that produces the best accuracy and dependability for the features of the CKDP dataset.

A. Smote

To redistribute the balance between classes, SMOTE, which is an over-sampling method, will produce artificial representatives of the minority group. As opposed to replicating the existing data, it interpolates new cases between the existing minority samples, and this reduces overfitting and enhances the efficiency of the classifiers on imbalanced datasets.

B. Ensemble machine learning (EML)

EML involves using several models to predict the results that are more accurate and reliable. Ensemble learning has techniques such as begging, improving and stacking. This is used when there is uncertainty in the data used to minimize bias and variation to improve generalization.

C. Stacking ensemble classifier (SEC)

A complex system of ensemble learning referred to as the Stacking Ensemble Classifier (SEC) involves many individual classifiers that are merged to enhance prediction accuracy in general. This technique is to train a set of various models (SVM, K-NN, MLP, and Decision Tree) independently on the same data set. These models are then inputted into the last model which is referred to as the meta-learner which in most cases is a logistic regression classifier. The meta-learner effectively uses the merits of each model by examining what the basic models have produced to come up with the final prediction. In complex tasks such as CKDP, such layered architecture enhances the model robustness, reduces overfitting, and it leads to improved performance [48].

D. Logistic regression (LR)

LR is a statistical classification technique that represents the probability of event occurrence (e.g. the presence of CKD in a person) as dependent on a single or multiple predictor variables. Its applicability and readability make it a common tool in medical prediction issues and it approximates probability with the use of logit function [26].

E. K-Nearest Neighbors (K-NN) algorithm

The K-NN is a popular supervised learning algorithm that can be used effectively both in regression and in the classification problems. It makes predictions based on certain close neighbors and the values and labels of the closest neighbors of the data point to predict the outcome. K-NN, being a non-parametric method that does not require any assumptions about the actual distribution of the information, is a highly flexible method that can be used to work with both numerical and categorically based data. Although it is not as sensitive, or sophisticated as more complex models, its simplicity, ease of operation, and reasonable effectiveness on a wide range of datasets has made it an appealing option [27, 28].

A popular evaluation metric to assess classification problems, binary and multiclass problems, is the confusion matrix. It is an important measure to assess the predictive performance in CKD model by comparing the expected and actual classifications. The confusion matrix represents an important component of the performance evaluation in medical diagnostic because it provides detailed information about how accurate the model is, the recall, the precision, and the overall efficiency of the model through the presentation of the true positives, the false positive, the true negatives and the false negative in a structured way [36,37].

- The proportion of correctly diagnosed cases of illness in all actual cases of disease is measured by a value known as the True Positive Rate (TPR) or sometimes called the Detection Rate (DR). It shows the model's sensitivity and accuracy in detecting positive cases. An ideal but rare case in practical applications such as CKDP is that the TPR value is 1, which represents perfect detection, implying that all actual cases of the disease have been diagnosed correctly. This metric (the mathematical definition of which is given below) is necessary to evaluate the model's ability to detect people with the condition:

$$\text{True Positive Rate (TPR)} = \frac{TP}{(TP + FN)} \qquad (1)$$

- The percentage of healthy or normal cases that are incorrectly labelled as diseased by the prediction model is called the False Positive Rate, or FPR. It highlights the prominence of false alarm which is of importance in

medical diagnosis where it is important to reduce unnecessary anxiety. A low FPR implies more model precision while a high FPR can lead to incorrect diagnosis and unnecessary treatment. Maintaining a low FPR in the context of CKDP ensures dependability and confidence in the results of the system. The formula that is used to find out the FPR is as follows:

$$\textbf{False Positive Rate (FPR)} = \frac{\textbf{FP}}{(\textbf{FP} + \textbf{TN})} \qquad \textbf{(2)}$$

- The percentage of actual cases of CKD that were missed and mistaken as normal by the CKDP model is called the False Negative Rate. This statistic is important since false negatives have serious health implications of delaying diagnosis and treatment. A high FNR has negative effect on 'clinical credibility' on models by implying their inability to identify real cases of a disease. Therefore, ensuring early and accurate detection means reducing the FNR. To calculate the FNR the following formula is used:

$$\textbf{False Negative Rate (FNR)} = \frac{\textbf{FN}}{(\textbf{FN} + \textbf{TP})} \qquad \textbf{(3)}$$

The confusion matrix in this study is as follows [[A B] [C D]],

- Where A is the number of properly predicted negative cases, indicating the model is able to correctly identify occurrences which are not diseases.
- The number of false positives, i.e., the number of people in the population that were incorrectly classified as positive and therefore could have been responsible for false alarms, is denoted with the letter B.
- C stands for false negatives, which are the cases where missed detections were a consequence of true positive cases but were forecasted as negative.
- D is the number of accurately predicted positive cases and shows how well the model can detect actual cases of the disease.

With the aim to make the best possible predictions to the actual class labels, the model can predict labels for both known and unknown samples of data after it has been trained. Figure 9 shows the use of a confusion matrix to evaluate the performance of the CKDP algorithm. The result of the prediction performance of this model can be clearly observed from this matrix, which indicates that the model is able to accurately classify the cases with and without CKD throughout the data set in the dataset [38, 39, and 40].

The K-NN model's suitability for this task was proven through the confusion matrix that was crucial in verifying the predicted labels in detection as well as in prediction phase. The results of using K-NN method on a synthetic dataset is shown in Figure 10, which shows how well it classifies data. The Receiver Operating Characteristic (ROC) curve which gives a visual of the tradeoff between true positive and false positive rates was used to analyses the performance of the model in detail. We were able to improve the overall accuracy and efficiency of our estimation methods and give important insights into the prediction patterns by putting the ROC analysis in our evaluation [41, 42].

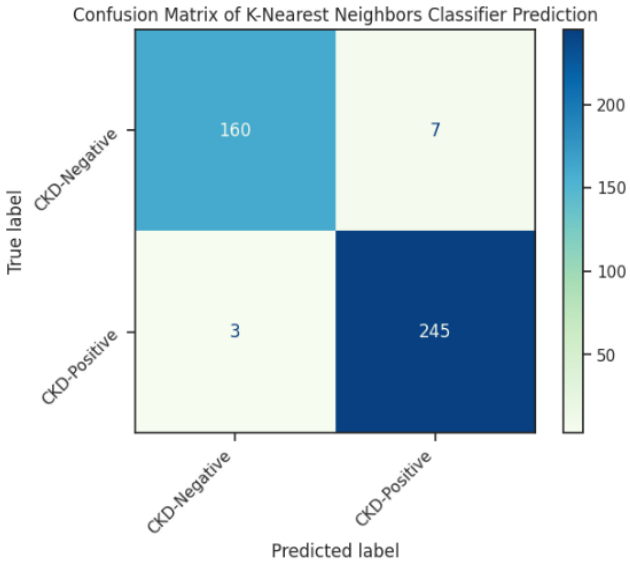


Figure 9: Confusion Matrix of K-NN

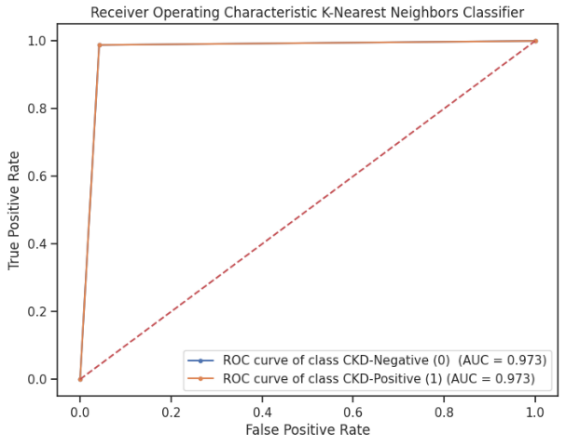


Figure 10: Illustrates the ROC Curve for K-NN

F. Multi-Layer Perceptron (MLP) Algorithm

MLP Algorithm. An improved type of supervised learning system built from multiple connected perceptron's is called Multi-Layer Perceptron (MLP). It consists of an output layer which gives predictions, one or more hidden layers which analyze data, and an input layer which receives data. The model can find complex, non-linear relationships in the data due to the hidden layers, which serve as the primary computational engine. MLPs can be used to accurately simulate any continuous function by adjusting the number of hidden layers and activation functions, such as linear, sigmoid or non-linear functions. due to their adaptability, MLPs are often used in jobs that involve strong classification and prediction ability and are suitable for solving problems that involve data that is non-linearly separable or interdependent [29, 30]. The Multi-Layer Perceptron (MLP) method can learn complex patterns and relations due to this process which is successful in mapping the data into a high-dimensional environment. With the aim of making its predictions as similar as possible to results, the model can reliably predict labels for both known and unknown data samples after it has been trained. Figure 11 shows graphically the performance of the MLP model as a confusion matrix, in which the accuracy of classification and the ability to differentiate the true and false predictions are emphasized [43, 44].

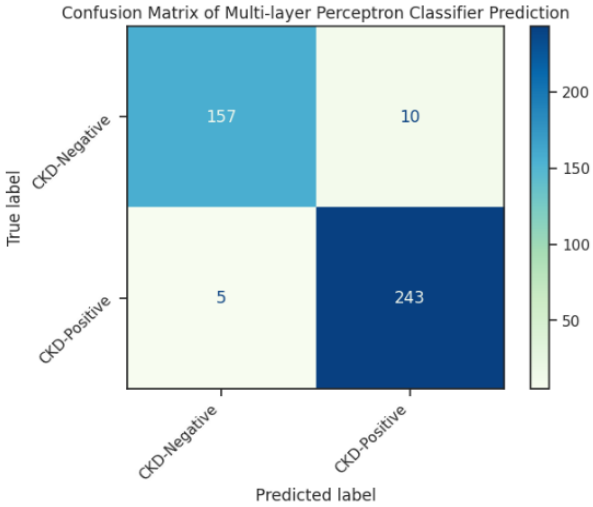


Figure 11: Confusion Matrix of MLP Algorithm

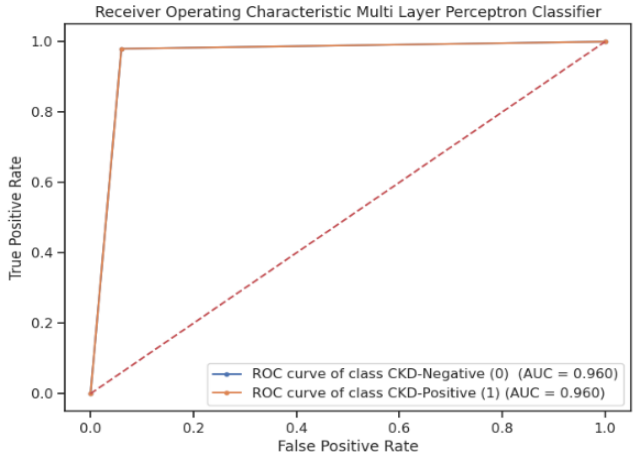


Figure 12: Illustrates the ROC Curve for MLP

The confusion matrix was key to checking the expected labels for the detection as well as the prediction assuming that Multi-Layer Perceptron (MLP) is a good fit for this case. The classification performance of the MLP method is evident from the visualization of the results obtained after having applied it to a synthetic dataset in Figure 12. The Receiver Operating Characteristic (ROC) curve that allows a more comprehensive understanding of the false positive and true positive rates was then used to further assess the efficacy of the model. We were able to improve the precision and resilience of our method of making estimations by incorporating ROC analysis which gave us deeper insights into patterns of prediction.

G. Support vector machine (SVM) algorithm

Strong ML techniques known as SVM are often used in regression and classification applications. They work by determining the best hyperplane in a higher dimensional space which best separates data points of different kinds. The primary objective is to achieve better generalization by putting in as much of a gap between classes as possible. SVM utilizes many different kernel functions such as linear, polynomial, sigmoid, Radial Basis Function (RBF) kernel, etc. SVM is used for complicated, non-linear patterns. These kernel functions are used to convert the input data to higher dimensions where separation is more practical. Support vectors (nearest data points from each class, they are essential in assessing model's performance) determine the decision boundary [31, 32].

By being able to paper the data successfully into a high-dimensional space, the Support Vector Machine (SVM) model is able to find complex patterns and class boundaries. The accuracy of the system to predict the label of known and unknown data depends on the training of the system. Making sure that the predictions of the model are very similar to the real class labels is the primary goal. The confusion matrix in figure 13 shows the performance of SVM algorithm and how much it can distinguish between classes and yield accurate classification results [45, 46].

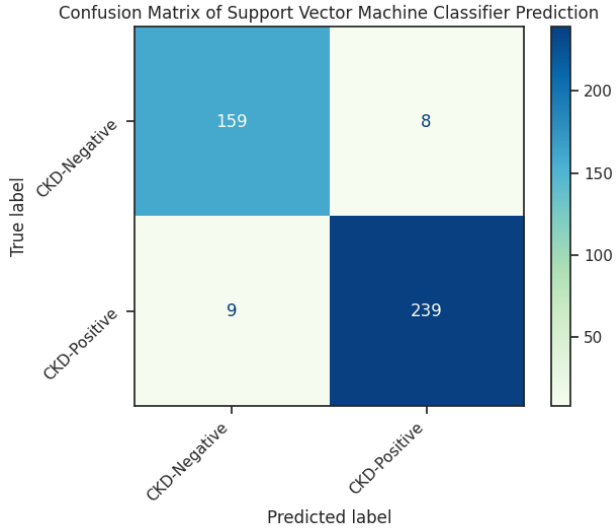


Figure 13: SVM Algorithm Confusion Matrix

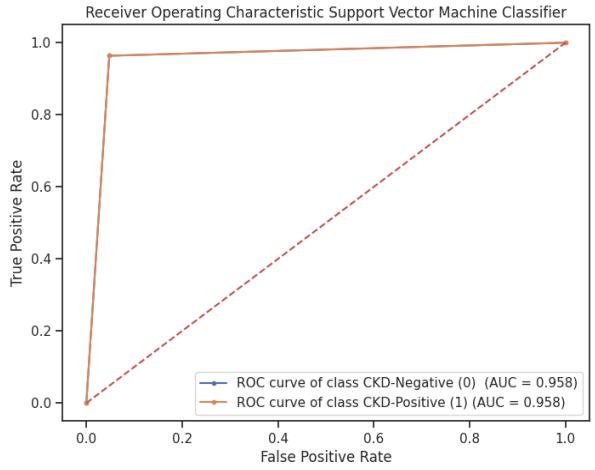


Figure 14: Illustrates the (ROC) Curve for SVM

The confusion matrix played a very important role in generating accurate labels for detection and prediction tasks, given that Support Vector Machine (SVM) is the suitable model for our specific case. The SVM method gave the results shown in Figure 14, applied to a made-up dataset. Receiver Operating Characteristic (ROC) curves were used to explore the model in finer detail. These curves give the visual representation of correctness of the model based on the evaluation defined by a user. In addition to displaying patterns of prediction, these insights from the ROC are useful in the improvement of estimation procedure, which would lead to increased overall accuracy and dependability of the model [47].

H. Quadratic Discriminant Analysis (QDA) Algorithm

A strong supervised classification method, QDA assumes that each of the classes in the dataset has a Gaussian distribution of its own. Class specific covariance matrices are modelled by QDA (as

opposed to LDA which results in flexible, curved decision boundaries that can better capture more complex data patterns. Because of this, QDA is especially well suited to the case where the equal class covariances assumption of LDA is violated. It is better at dealing with imbalanced classes and requires less data for training. But when there are a lot of predictors, QDA can be computationally taxing as the covariance matrix for each class has to be estimated separately and it can be overfit when the number of predictor variables is greater than the number of training data [33].

Classification accuracy is enhanced by this probabilistic method, which is more so when you are dealing with complex class distribution. The resulting confusion matrix of the performance of the QDA model will be shown in graphical form in Figure 15 where it is clear how the QDA model predicted class label.

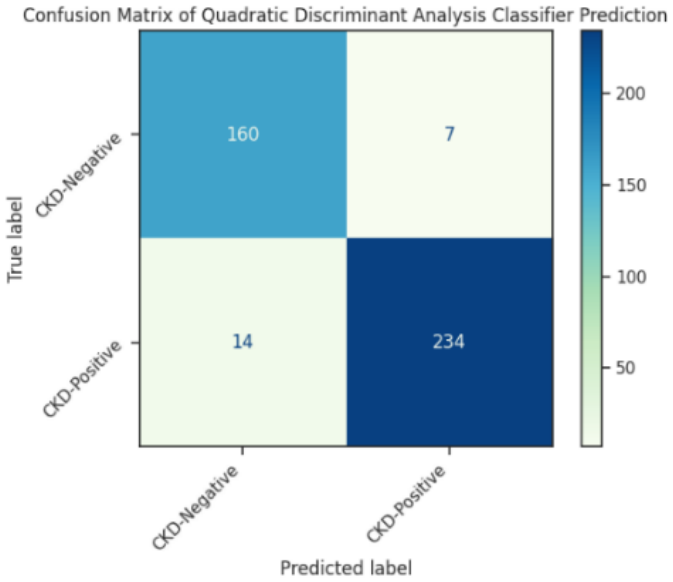


Figure 15: Confusion Matrix QDA Algorithm

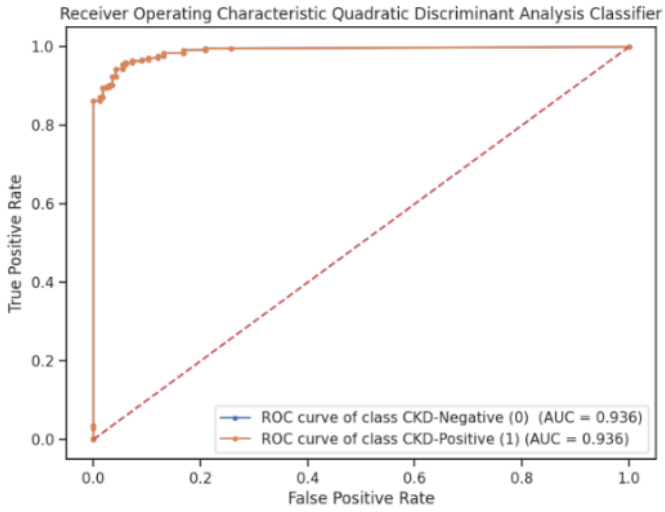


Figure 16: Illustrates the ROC Curve for QDA

By changing the discriminating threshold, ROC (Receiver Operating Characteristic) analysis is a powerful method of assessing the ability of a classifier model to distinguish between classes. By balancing the true positive and false positive rates, it makes it possible to make decisions based on data and relate heavily with cost-benefit analysis. This trade-off is essentially shown by the ROC curve, which is useful for evaluating the model performance for different thresholds. The ROC curve for our model is depicted in Figure 16 which gives us a very good idea of the efficacy of classification and highlights its ability to give correct predictions.

I. Decision Tree (DT) Algorithm

Decision Tree (DT) ALGORITHM: Due to its interpretability, organized decision-making process, DT approach is a popular supervised ML strategy for the categorization of tasks. Using a set of learnt decision rules, derived from training data, builds a model in the shape of a tree that predicts the value or class of a target variable. Every instance of data is evaluated based on the importance of the features in relation to the target attribute at the root node, where the procedure begins. Decision branches are made by selecting relevant attributes and comparisons are made at each node to guide the flow in the direction of the desired result. The data splits into more specialized subgroups as it proceeds until it reaches a decision at the leaf nodes. Because of their large adaptability, decision trees can be combined with various scientific techniques and algorithms to increase the accuracy of the prediction [34, 35].

The DT classifier works well to decide the probabilities of different classes by following these structured phases, which can be categorized with high accuracy according to the probabilities with the highest probability. The DT algorithm was employed in our study for the analysis of historic data, which made an accurate prediction followed by suitable labels. In order to make the predictions fall under the right categories, the model learned decision rules during the training phase. The confusion matrix as shown in Figure 17 illustrates the result of the performance, and it amply depicts how good the Decision Tree technique is in classifying the data points.

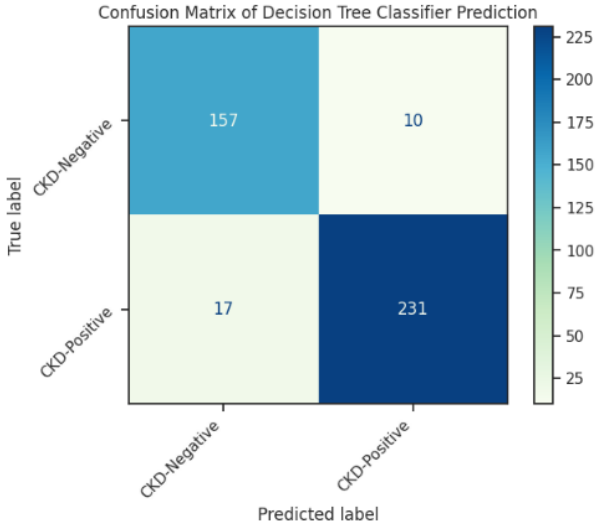


Figure 17: Confusion Matrix (DT) Algorithm

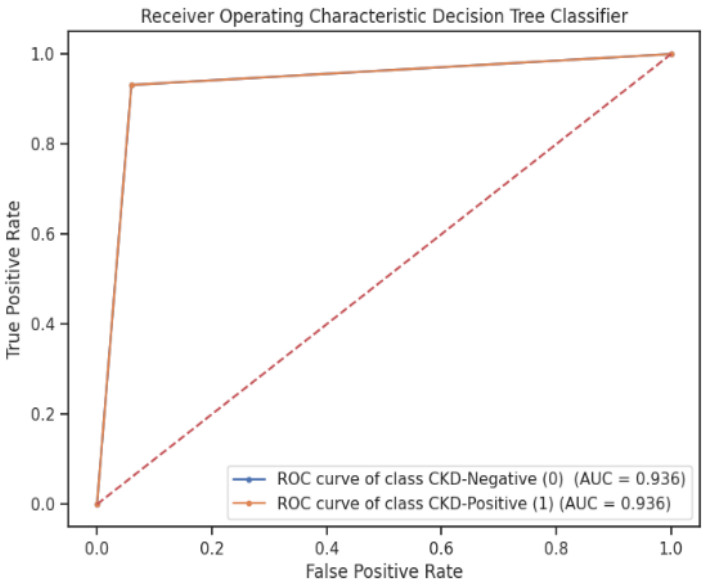


Figure 18: Illustrates the (ROC) Curve DT

J. Experimental Setup

Python is a powerful and versatile programming language that is often used in areas such as web applications, machine learning, software development and data analysis. In this work, we used python to develop several ML algorithms and expedite data processing utilizing the Anaconda Navigator and Jupyter Notebook interface.

- It has 16.0 GB of RAM, which guarantees that memory-intensive operations run smoothly.
- The machine has an x64-based processor and runs Windows 11 Home (64-bit), providing a contemporary and compatible programming environment.
- A 1TB Solid-State Drive (SSD) is provided, offering plenty of storage space for datasets, models, and results as well as quick data access.

IV. RESULTS AND DISCUSSION

We carried out a thorough analysis of several hybrid approaches to determine how well they might predict and categorize CKD. To improve diagnosis accuracy, each approach incorporated classification and clustering algorithms. The hybrid algorithm combinations and their accompanying accuracy rates are shown in the table below, providing a clear comparison of model performance and emphasizing the best method for CKD prediction (as illustrated in Table 6).

Table 6: Shows the precision of models that use a variety of techniques to predict CKD

Hybrid methods	Accuracy of methods
Agglomerative Hierarchical, Stacking Ensemble & Logistic Regression, Decision Tree (DT) Proposed Method	93.4939 %
Agglomerative Hierarchical, Stacking Ensemble & Logistic Regression, Quadratic Discriminant Analysis (QDA) Proposed Method	94.9397 %
Agglomerative Hierarchical, Stacking Ensemble & Logistic Regression, Support Vector Machine (SVM) Proposed Method	95.9036 %
Agglomerative Hierarchical, Stacking Ensemble & Logistic Regression, Multi-layer Perceptron (MLP) Proposed Method	96.3855 %
Agglomerative Hierarchical, Stacking Ensemble & Logistic Regression, K-Nearest Neighbors (K-NN) Proposed Method	97.5903 %

The combination of AHC, SEC-LR, and K-NN produced the best accuracy of 97.5903% among them, suggesting that it was a powerful model for predicting CKD. AHC, SEC-LR, and MLP together produced an impressive accuracy of 96.3855%, whereas AHC, SEC-LR, and SVM together produced an impressive accuracy of 95.9036%. Additional combinations, such as AHC, SEC-LR, and QDA, produced an impressive accuracy of 94.9397 percent. Moreover, a respectable accuracy of 93.4939% was obtained by combining AHC, SEC-LR, and DT. These findings highlight how well hybrid ensemble approaches can improve the accuracy and dependability of CKD classification algorithms (as illustrated in Table 7 and Table 8).

Table 7: Performance Scores of Hybrid Algorithms in CKD Diagnosis

S/No.	Parameter Score	Agglomerative Hierarchical, SEC & LR, DT Algorithm	Agglomerative Hierarchical, SEC & LR, QDA Algorithm	Agglomerative Hierarchical, SEC & LR, SVM Algorithm
1	Precision	0.93040253	0.94524729	0.95701995
2	Recall	0.93578568	0.95081610	0.95790274
3	F1-Score	0.93280319	0.94773581	0.95745514
4	Sensitivity	0.94011976	0.95808383	0.95209580
5	Specificity	0.93145161	0.94354838	0.96370967

Table 8: Evaluation Metrics of Hybrid Algorithms in CKD Diagnosis

S/No.	Parameter Score	Agglomerative Hierarchical, SEC & LR, MLP Algorithm	Agglomerative Hierarchical, SEC & LR, K-NN Algorithm
1	Precision	0.96480505	0.97690865
2	Recall	0.95997923	0.97299352
3	F1-Score	0.96223358	0.97484848
4	Sensitivity	0.94011976	0.95808383
5	Specificity	0.97983870	0.98790322

The parameter values show how well various algorithm setups work in predicting breast cancer. The AHC, SEC-LR, and DT algorithm combinations had the following respective precision, recall, F1-score, sensitivity, and specificity: 0.93040253, 0.93578568, 0.93280319, 0.94011976, and 0.93145161. However, combining AHC, SEC-LR, and QDA resulted in the following results: precision of 0.94524729, recall of 0.95081610, F1-score of 0.94773581, sensitivity of 0.95808383, and specificity of 0.94354838. Furthermore, the combination of AHC, SEC-LR, and MLP produced a precision of 0.95701995, a recall of 0.95790274, an F1 score of 0.95745514, a sensitivity of 0.95209580, and a specificity of 0.96370967. Additionally, a precision of 0.96480505, recall of 0.95997923, F1-score of 0.96223358, sensitivity of

0.94011976, and perfect specificity of 0.97983870 were attained by the combination of AHC, SEC-LR, and MLP. Finally, a precision of 0.97690865, recall of 0.97299352, F1-score of 0.97484848, sensitivity of 0.95808383, and perfect specificity of 0.98790322 were attained by the combination of AHC, SEC-LR, and K-NN.

The effectiveness of hybrid combinations of algorithms in terms of accurate breast cancer predictions is shown by these parameter values. The ability of the models to filter out false results and find true positives is proven by their consistently excellent scores across important evaluation criteria, such as precision and recall. This is a very impressive result that suggests that these hybrid techniques have a lot of promise for better clinical decision making and prognostic accuracy in breast cancer research.

The accuracy performance of various combinations of the hybrid algorithms used in the (CKD) prediction system is as shown in Figures 19 and 20. Four important combinations are shown in the bar chart with the accuracy value ranging from 93.49% to 97.59%. Interestingly, the combination of AHC, K-NN and Stacking Ensemble Classifier with Logistic Regression (SEC-LR) gave the best possible accuracy of 97.59%, showing its tremendous potential in predicting CKD. Closely following, the results of combination of AHC, SEC-LR and MLP gave an outstanding accuracy of 96.39%, while the results of combination of AHC, SEC-LR and SVM achieved a strong 95.90%. Accuracy was found to be 94.94% and 93.49% when AHC, SEC-LR and QDA and AHC, SEC-LR and DT were used in combination respectively. These outcomes are unequivocal proof of the working of hybrid ensemble approaches to enhance prediction accuracy. The potential and reliability of CKD prediction systems are greatly improved by using many ML models as a single system.

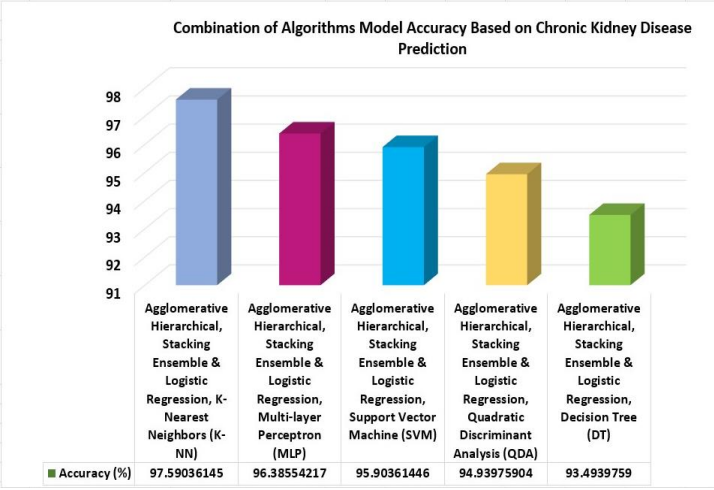


Figure 19: Combination of Algorithms Model Accuracy of (CKDP)

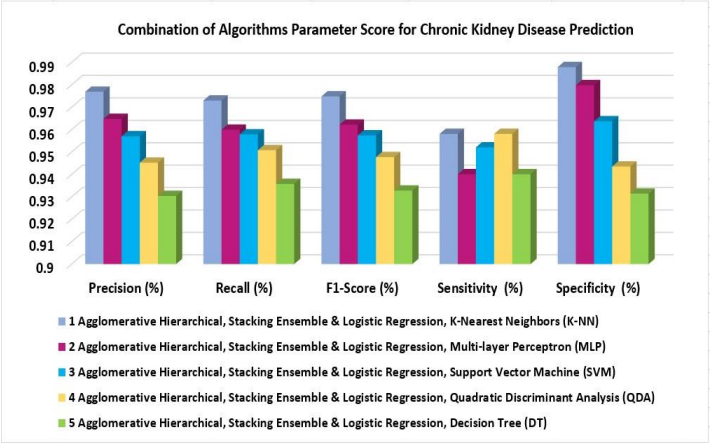


Figure 20: Combination of Algorithms Parameter Score of (CKDP)

The predictive performance of the combinations of hybrid models is maximized as shown in the graph of accuracy level. This means that the precision of the system is at its greatest at the current time. High - performance alignments of critical assessment metrics, such as "Score Precision", "F1-Score", "Recall", "Sensitivity" and "Specificity", further indicate a balanced and trustworthy prediction model. Even though the system generates good results, these measures can still be improved upon or altered in accordance with needs or clinical goals in order to create more efficacy of diagnosis.

The performance of multiple combination strategies for Chronic Kidney Disease Prediction (CKDP) models is compared in Table 9, that highlights the accuracy gained by different approaches. The analysis incorporates the proposed models and research that has already been published to show how the new approach improves on them. Ghosh et al. [18] created a stacked ensemble of Naive Bayes, Decision Tree, and Random Forest, which got a good accuracy of 94.99%. Previous research including the work done by Mohammed & Beshah et al. [11] used Knowledge-Based System with the use of Decision Tree which reached 91% accuracy. While Tsai et al. [16] study used Random Forest with SHAP and accuracy reached 92.1%, while Amirgaliyev et al. [15] demonstrated 93% accuracy using SVM with clinical characteristics. The performance of the proposed methods is much better than these previous studies. The hybrid models proposed in this study, however, are noticeably better than existing methods. Notably, the highest accuracy of 97.59% was achieved through a combination of Agglomerative Hierarchical Clustering (AHC), Stacking Ensemble with Logistic Regression (SEC-LR) and K-Nearest Neighbors (K-NN). With accuracy between 93.49% and 96.39%, other combinations such as SEC-LR using MLP, SVM, QDA and DT also demonstrated excellent predictive performance. These findings represent the degree of improvement in the precision and robustness of CKD prediction systems that can be offered through the suggested hybrid ensemble approaches. The performance of the proposed hybrid algorithms is shown in this comparative study, and it exhibits great potential to enhance the accuracy of prediction of health data much better than the established methods (as illustrated in Table 9).

V. COMPARATIVE ANALYSIS

The performance of multiple combination strategies for Chronic Kidney Disease Prediction (CKDP) models is compared in Table 9, that highlights the accuracy gained by different approaches. The

analysis incorporates the proposed models and research that has already been published to show how the new approach improves on them. Ghosh et al. [18] created a stacked ensemble of Naive Bayes, Decision Tree, and Random Forest, which got a good accuracy of 94.99%. Previous research including the work done by Mohammed & Beshah et al. [11] used Knowledge-Based System with the use of Decision Tree which reached 91% accuracy. While Tsai et al. [16] study used Random Forest with SHAP and accuracy reached 92.1%, while Amirgaliyev et al. [15] demonstrated 93% accuracy using SVM with clinical characteristics. The performance of the proposed methods is much better than these previous studies. The hybrid models proposed in this study, however, are noticeably better than existing methods. Notably, the highest accuracy of 97.59% was achieved through a combination of Agglomerative Hierarchical Clustering (AHC), Stacking Ensemble with Logistic Regression (SEC-LR) and K-Nearest Neighbors (K-NN). With accuracy between 93.49% and 96.39%, other combinations such as SEC-LR using MLP, SVM, QDA and DT also demonstrated excellent predictive performance. These findings represent the degree of improvement in the precision and robustness of CKD prediction systems that can be offered through the suggested hybrid ensemble approaches. The performance of the proposed hybrid algorithms is shown in this comparative study, and it exhibits great potential to enhance the accuracy of prediction of health data much better than the established methods (as illustrated in Table 9).

Table 9: Comparative Study of Hybrid Methods for CKD

Study / Authors	Hybrid Method	Accuracy
Mohammed & Beshah et al. [11]	Knowledge-Based System (KBS) with Decision Tree	91.00%
Xie et al. [12]	Naïve Bayes, K-NN, SVM, DT, ANN	92.10%
Amirgaliyev et al. [15]	SVM with Clinical Features	93.00%
Tsai et al. [16]	Random Forest with SHAP	92.10%
Ghosh et al. [18]	Hybrid Model: Naïve Bayes + Decision Tree + Random Forest (Stacked Ensemble)	94.99%
Zhu et al. [19]	XGBoost for CVD Prediction in CKD Patients + SHAP	80.60%
Proposed Method (AHC + SEC-LR + DT)	Agglomerative Hierarchical Clustering + Stacking Ensemble + Decision Tree	**93.49%**
Proposed Method (AHC + SEC-LR + QDA)	Agglomerative Hierarchical Clustering + Stacking Ensemble + QDA	**94.94%**
Proposed Method (AHC + SEC-LR + SVM)	Agglomerative Hierarchical Clustering + Stacking Ensemble + SVM	**95.90%**
Proposed Method (AHC + SEC-LR + MLP)	Agglomerative Hierarchical Clustering + Stacking Ensemble + Multi-layer Perceptron	**96.39%**
Proposed Method (AHC + SEC-LR + K-NN)	Agglomerative Hierarchical Clustering + Stacking Ensemble + K-Nearest Neighbors	**97.59%**

VI. STUDY LIMITATIONS AND FUTURE RECOMMENDATIONS

Although the proposed Ensemble Machine Learning (EML)-based framework demonstrates promising results for early CKD prediction, several limitations should be considered. First, the performance of the proposed models depends on the quality, size, and diversity of the available dataset. Since the study relies on existing clinical data, variations in patient characteristics, healthcare practices, and data collection procedures may affect the generalizability of the proposed approach. Second, the current framework primarily focuses on structured clinical attributes, while other important sources of medical information, such as electronic health records, medical imaging, and longitudinal patient data, are not incorporated. Additionally, although ensemble learning techniques improve prediction performance, the complexity of these models may reduce interpretability, which remains an important consideration for clinical adoption.

Future research can address these limitations by evaluating the proposed framework on larger, multi-center, and real-world clinical datasets to improve reliability and generalization. The integration of additional healthcare data sources, including electronic health records, genomic information, and medical imaging data, can provide a more comprehensive understanding of CKD progression. Furthermore, advanced explainable artificial intelligence (XAI) techniques can be incorporated to improve model transparency and help healthcare professionals better understand prediction outcomes. Future studies may also explore deep learning and hybrid AI approaches to enhance predictive capability and develop more robust clinical decision-support systems for early CKD diagnosis and personalized patient care.

VII. CONCLUSION

Machine learning (ML) and artificial intelligence (AI) methods have demonstrated great promise in the detection of high-risk individuals and the accurate prediction of the development of chronic kidney disease (CKD). By combining advanced analytical models with Clinical Decision Support Systems (CDSS) and Electronic Health Records (EHRs), healthcare providers have the potential to improve early disease detection, contribute to the timely interventions and improve overall patient management strategies. The use of ML algorithms in computer-assisted decision-making systems not only leads to the improvement of diagnostic accuracy, but also the efficiency of clinical workflow, which will ultimately lead to cost-effective healthcare delivery. In this research, a strong hybrid framework is proposed in order to support accurate diagnosis and prediction of CKD. The technique exploits a wide array of machines learning algorithms such as Decision Tree (DT), Agglomerative Hierarchical Clustering (AHC), K-Nearest Neighbors (K-NN), Multi-Layer Perceptron (MLP), Ensemble machine learning (EML), Logistic regression (LR), Stacking ensemble classifier (SEC), and Quadratic discriminants (QDA) to construct a comprehensive and intelligent model for the diagnostic purpose. The maximum accuracy of 97.5903% was achieved by AHC, SEC-LR, and K-NN first, AHC, SEC-LR, and MLP (96.3855%), AHC, SEC-LR, and SVM (95.9036%), AHC, SEC-LR, and QDA (94.9397%) then AHC, SEC-LR, and DT (93.4939%). These results demonstrate how successful hybrid ensemble methods are at predicting CKD. Future studies could look into the suggested system's scalability and generalizability by using even more advanced algorithms and including even more varied data. Improvements in the interpretability and adaptability of the system could potentially help physicians make more precise and data-driven judgements, which would ultimately lead to improvements in the standard of healthcare delivery.

REFERENCES

[1] K. T. Mills, Y. Xu, W. Zhang, J. D. Bundy, C. S. Chen, T. N. Kelly, et al., “A systematic analysis of worldwide

population-based data on the global burden of chronic kidney disease in 2010," Kidney International, vol. 88, pp. 950–957, 2015, doi: 10.1038/ki.2015.230.

[2] K. Matsushita, S. H. Ballew, A. Y. Wang, R. Kalyesubula, E. Schaeffner, and R. Agarwal, "Epidemiology and risk of cardiovascular disease in populations with chronic kidney disease," Nature Reviews Nephrology, vol. 18, pp. 696–707, 2022, doi: 10.1038/s41581-022-00616-6.

[3] S. R. Vaidya and N. R. Aeddula, "Chronic kidney disease," StatPearls [Internet], StatPearls Publishing, Treasure Island, FL, USA, 2024.

[4] A. S. Chouhan, M. Kaple, and S. Hingway, "A brief review of diagnostic techniques and clinical management in chronic kidney disease," Cureus, vol. 15, Art. no. e49030, 2023, doi: 10.7759/cureus.49030.

[5] A. Whaley-Connell, R. Nistala, and K. Chaudhary, "The importance of early identification of chronic kidney disease," Missouri Medicine, vol. 108, no. 1, pp. 25–28, 2011.

[6] P. Chittora, S. Chaurasia, P. Chakrabarti, G. Kumawat, T. Chakrabarti, Z. Leonowicz, et al., "Prediction of chronic kidney disease—a machine learning perspective," IEEE Access, vol. 9, pp. 17312–17334, 2021.

[7] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," Journal of Big Data, vol. 9, no. 1, Art. no. 109, 2022.

[8] E. Dritsas and M. Trigka, "Machine learning techniques for chronic kidney disease risk prediction," Big Data and Cognitive Computing, vol. 6, no. 3, Art. no. 98, 2022.

[9] B. Boukenze, A. Haqiq, and H. Mousannif, "Predicting chronic kidney failure disease using data mining techniques," in Advances in Ubiquitous Networking 2: Proceedings of UNet'16, Springer, Singapore, 2017, pp. 701–712.

[10] L. Jena, B. Patra, S. Nayak, S. Mishra, and S. Tripathy, "Risk prediction of kidney disease using machine learning strategies," in Intelligent and Cloud Computing: Proceedings of ICICC 2019, vol. 2, Springer, Singapore, 2021, pp. 485–494.

[11] M. Almasoud and T. E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors," International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, pp. 89–96, 2019.

[12] R. Mahmood, S. Mustafa, H. Asif, A. Raza, and M. Salahuddin, "Leveraging artificial intelligence to optimize software project management: Enhancing efficiency, risk mitigation, and decision-making," Contemporary Journal, doi: 10.12345/f42z1z57.

[13] F. U. Zaman, M. Z. Khan, A. Imroz, A. A. Khan, M. Salahuddin, and M. Kainat, "Student performance analysis in higher education using integrated approach of machine learning techniques," in *Proc. 2024 International Conference on Sustainable Technology and Engineering (i-COSTE)*, Dec. 2024, pp. 1–6.

[14] A. Abbas, M. Salahuddin, M. Z. Khan, et al., "Machine learning-based hybrid technique to enhance cyber-attack perspective," *Journal of Cloud Computing*, vol. 14, Art. no. 57, 2025, doi: 10.1186/s13677-025-00782-5.

[15] M. Z. Khan, S. A. Shaikh, A. A. Khan, A. Imroz, M. Salahuddin, P. Bhatti, and S. D. Bhatti, "Optimizing heart disease forecasting: Bridging gaps in interpretability, efficiency, and scalability using machine learning," *Biomedical Materials & Devices*, 2025, doi: 10.1007/s44174-025-00504-0.

[16] M. Z. Khan, S. A. Shaikh, A. A. Khan, A. Imroz, M. Salahuddin, P. Bhatti, and S. D. Bhatti, "Artificial intelligence in dermatology: Current applications and future innovations," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/396020302_ARTIFICIAL_INTELLIGENCE_IN_DERMATOLOGY_CURRENT_APPLICATIONS_AND_FUTURE_INNOVATIONS

[17] P. A. Moreno-Sánchez, "Data-driven early diagnosis of chronic kidney disease: Development and evaluation of an explainable AI model," *IEEE Access*, vol. 11, pp. 38359–38369, 2023, doi: 10.1109/ACCESS.2023.3264270.

[18] B. P. Ghosh et al., "Advancing chronic kidney disease prediction: Comparative analysis of machine learning algorithms and a hybrid model," *Journal of Computer Science and Technology Studies*, vol. 6, no. 3, pp. 15–21, 2024, doi: 10.32996/jcsts.2024.6.3.2.

[19] H. Zhu et al., "Machine learning model for cardiovascular disease prediction in patients with chronic kidney disease," *Frontiers in Endocrinology*, vol. 15, Art. no. 1390729, 2024, doi: 10.3389/fendo.2024.1390729.

[20] "Chronic Kidney Disease Prediction (CKDP) Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/rabieelkharoua/chronic-kidney-disease-dataset-analysis/data

[21] "Feature Importance," Built In, 2011. [Online]. Available: https://builtin.com/data-science/feature-importance

[22] W. U. Zahra, F. U. Zaman, G. Mumtaz, M. Salahuddin, M. Z. Khan, S. A. Sultan, S. Hira, and F. Parveen, "Approaches to predict cardiovascular issue using machine learning method," *Spectrum of Engineering Sciences*, vol. 3, no. 4, pp. 417–429, 2025.

[23] M. Muqaddas, S. Majeed, S. Hira, and G. Mumtaz, "A systematic literature review on performance evaluation of SQL and NoSQL database architectures," *Journal of Computing & Biomedical Informatics*, vol. 7, no. 2, 2024. [Online]. Available: https://jcbi.org/index.php/Main/article/view/548/502

[24] M. Moeez, R. Mahmood, H. Asif, M. W. Iqbal, K. Hamid, U. Ali, and N. Khan, "Comprehensive analysis of DevOps: Integration, automation, collaboration, and continuous delivery," *Bulletin of Business and Economics*, vol. 13, no. 1, 2024.

[25] M. Salahuddin, F. U. Zaman, G. Mumtaz, M. Z. Khan, M. Kainat, S. Hira, F. Parveen, and R. Mahmood, "Integrating network intrusion detection with machine learning techniques for enhanced network security," Spectrum of Engineering Sciences, vol. 3, no. 4, pp. 612–625, 2025. [Online]. Available: https://sesjournal.com/index.php/1/article/view/286

[26] M. S. Ramzan, F. Nasim, H. N. Ahmed, U. Farooq, and H. Khan, "An innovative machine learning-based end-to-end data security framework in emerging cloud computing databases and integrated paradigms: Analysis on taxonomy, challenges, and opportunities," Spectrum of

Engineering Sciences, vol. 3, no. 2, pp. 90–125, 2025. [Online]. Available: https://sesjournal.com/index.php/1/article/view/106

[27] F. Parveen, S. Iqbal, G. Mumtaz, and M. Salahuddin, "Real-time intrusion detection with deep learning: Analyzing the UNR intrusion detection dataset," Journal of Computing & Biomedical Informatics, vol. 7, no. 2, 2024. [Online]. Available: https://jcbi.org/index.php/Main/article/view/554

[28] M. Afzal, M. Salahuddin, S. Hira, M. F. Sultan, S. Z. Ahmad, and M. W. Iqbal, "A systematic literature review of understanding the human-computer interaction collaboration with user experience design," Bulletin of Business and Economics, vol. 13, no. 2, pp. 723–729, 2024, doi: 10.61506/01.00386.

[29] H. Ahmad, G. Mumtaz, and M. Salahuddin, "A hybrid deep learning model for high-accuracy brain tumor," Al-Aasar, vol. 2, no. 3, pp. 1–15, Aug. 2025, doi: 10.63878/aaj737.

[30] "Agglomerative Hierarchical Clustering," Built In, 2011. [Online]. Available: https://builtin.com/machine-learning/agglomerative-clustering "Agglomerative Hierarchical Clustering," Solver, 1996. [Online]. Available: https://www.solver.com/xlminer/help/hierarchical-clustering-intro

[31] "K-Nearest Neighbors (K-NN) Algorithm," IBM. [Online]. Available: https://www.ibm.com/topics/knn

[32] S. Y. Shaikh, N. A. Qureshi, M. Z. Khan, M. A. Khan, A. Imroz, and M. A. Kalwar, "Performance analysis of classification algorithms for software defects prediction by mathematical modelling and simulations," Sindh Journal of Headways in Software Engineering, vol. 2, no. 1. [Online]. Available: http://sjhse.smiu.edu.pk/sjhse/index.php/SJHSE/article/view/54

[33] A. Ilyas, F. U. Zaman, M. Salahuddin, F. A. Zia, M. Z. Khan, S. Hira, and M. A. Ameen, "AI-driven intrusion detection using machine learning: An anomaly-based analysis of network traffic," Spectrum of Engineering Sciences, vol. 3, no. 12, pp. 664–679, 2025. [Online]. Available: https://thesesjournal.com/index.php/1/article/view/1708/1288

[34] "Quadratic Discriminant Analysis," Society for Elimination of Rural Poverty (SERP), 2014. [Online]. Available: https://serp.ai/quadratic-discriminant-analysis/

[35] M. Z. Khan, S. A. Shaikh, M. A. Shaikh, K. Khatri, M. A. Rauf, A. Kalhoro, and M. Adnan, "The performance analysis of machine learning algorithms for credit card fraud detection," International Journal of Online Engineering (iJOE), vol. 19, no. 3, 2022, doi: 10.3991/ijoe.v19i03.35331.

[36] "Decision Tree Algorithm," KDnuggets, 2020. [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

[37] M. Z. Khan, A. A. Khan, A. A. Laghari, Z. A. Shaikh, M. A. K. Khani, D. Morkovkin, O. Gavel, S. Shkodinsky, S. Makar, and D. Taburov, "Comparative case study: An evaluation of performance computation between support vector machine, K-nearest neighbors, K-mean, and principal component analysis," Journal of Tianjin University Science and Technology, 2022. [Online]. Available: https://tianjindaxuexuebao.com/details.php?id=DOI:10.17605/OSF.IO/HK3SF

[38] M. A. Baig, S. A. Shaikh, K. K. Khatri, M. A. Shaikh, M. Z. Khan, and A. Rauf, "Prediction of students performance level using integrated approach of ML algorithms," International Journal of Emerging Technologies in Learning (iJET), vol. 18, no. 1, pp. 216–234, 2023, doi: 10.3991/ijet.v18i01.35339.

[39] M. Z. Khan, F. U. Zaman, M. Adnan, A. Imroz, M. Abdul Rauf, and Z. Phul, "Comparative case study: An evaluation of performance computation between SQL and NoSQL database," Sindh Journal of Headways in Software Engineering, 2023. [Online]. Available: http://sjhse.smiu.edu.pk/sjhse/index.php/SJHSE/article/view/42

[40] H. Farahbakhsh, N. Qureshi, M. Z. Khan, M. A. Khan, A. S. Soomro, A. Imroz, and H. B. Marri, "Performance evolution for sentiment classification using machine learning algorithm," Journal of Applied Research in Technology & Engineering, vol. 4, no. 2, pp. 97–110, 2023, doi: 10.4995/jarte.2023.19306.

[41] M. Siddiqui, H. A. Kalwar, M. Z. Khan, M. A. Khan, A. Imroz, M. A. Kalwar, and H. B. Marri, "Performance analysis for the diagnosis of COVID-19 prediction by mathematical modeling and simulation," International Journal of Artificial Intelligence & Mathematical Sciences, 2023. [Online]. Available: http://ijaims.smiu.edu.pk/ijaims/index.php/AIMS/article/view/47

[42] M. Z. Khan, F. U. Zaman, P. Bhatti, A. A. Khan, S. A. Sultan, and S. D. Bhatti, "A comparative analysis of deep learning architectures for efficient brain tumor detection," in Proc. 2024 International Conference on Sustainable Technology and Engineering (i-COSTE), Dec. 2024, pp. 1–6.

[43] P. Bhatti, A. Alim, F. U. Zaman, S. A. Sultan, M. Z. Khan, and S. M. N. Mustafa, "Machine learning-based liver disease prediction: Enhancing diagnosis and prognosis," in Proc. 2024 International Conference on Sustainable Technology and Engineering (i-COSTE), Dec. 2024, pp. 1–7.

[44] S. D. Bhatti, F. U. Zaman, A. A. Khan, M. Z. Khan, S. A. Sultan, and A. Imroz, "A machine learning model for the early detection of adults' serious asthma," in Proc. 2024 International Conference on Sustainable Technology and Engineering (i-COSTE), Dec. 2024, pp. 1–6.

[45] F. U. Zaman, M. Z. Khan, A. Imroz, A. A. Khan, M. Salahuddin, and M. Kainat, "Student performance analysis in higher education using integrated approach of machine learning techniques," in Proc. 2024 International Conference on Sustainable Technology and Engineering (i-COSTE), Dec. 2024, pp. 1–6.

[46] J. Du, J. Gao, J. Guan, B. Jin, N. Duan, L. Pang, H. Huang, Q. Ma, C. Huang, and H. Li, "Applying stacking ensemble method to predict chronic kidney disease progression in Chinese population based on laboratory information system: A retrospective study," PeerJ, vol. 12, Art. no. e18436, 2024.