IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 5, October 2025, pp. 4017~4031
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i5.pp4017-4031

Detection of chronic kidney disease based on ensemble approach with optimal feature selection using machine learning

Deepika Amol Ajalkar[1], Jyoti Yogesh Deshmukh[2], Mayura Vishal Shelke[3], Shalini Vaibhav Wankhade[4], Shwetal Kishor Patil[5]

[1]Department of Cyber Security and Data Science, G. H. Raisoni College of Engineering and Management, Affilated to Savitribai Phule Pune University, Pune, India
[2]Department of Artificial Intelligence and Data Science, Marathwada Mitramandal's Institute of Technology, Savitribai Phule Pune University, Pune, India
[3]Department Computer Science and Engineering, School of Computing, MIT Art, Design and Technology University, Pune, India
[4]Department of Information Technology, Vishwakarma Institute of Technology, Affilated to Savitribai Phule Pune University, Pune, India
[5]Department of Computer Engineering, Marathwada Mitramandal's Institute of Technology, Savitribai Phule Pune University, Pune, India

Article Info

Article history:

Received Sep 19 2024
Revised Jul 1, 2025
Accepted Aug 6, 2025

Keywords:

Chronic kidney disease
Early detection
Ensemble learning
Hyperparameter tuning
Machine learning

ABSTRACT

Chronic kidney disease (CKD) poses a significant health risk globally, necessitating early and accurate detection to ensure timely intervention and effective treatment. This study presents an advanced ensemble machine learning (ML) approach combined with optimal feature selection to enhance the detection of CKD. Using five baseline ML classifiers like gradient boosting (GB), random forest (RF), K-nearest neighbors (KNN), support vector machine (SVM), and decision tree (DT), and utilizing grid search for hyperparameter tuning, the proposed ensemble model capitalizes on the strengths of each algorithm. Our approach was tested on a public benchmark CKD dataset from Kaggle. The experimental results demonstrate that the ensemble model consistently outperforms individual classifiers and existing methods, achieving 97.5% accuracy, precision, recall, and an F1-score of 97.4%. This superior performance underscores the ensemble model's potential as a reliable early CKD detection tool. Integrating ML into CKD diagnostics enhances accuracy. It facilitates the development of automated, scalable diagnostic tools, aiding healthcare professionals in making informed decisions and ultimately improving patient outcomes.

Corresponding Author:

Mayura Vishal Shelke
Department Computer Science and Engineering, School of Computing
MIT Art, Design and Technology University
Pune, India
Email: mayura.shelke@gmail.com

1. INTRODUCTION

In today's world, chronic kidney disease (CKD) is a rapidly expanding illness that affects thousands of people due to the lack of early indicators and accessible medical treatment. Most people with CKD are from middle-class and lower-class nations [1], [2]. Approximately 1 billion people died from CKD in 2017 [3]. CKD is more common in developed countries. In lower- to middle-income nations, there are 390.6 billion people with CKD overall, with 178 billion men and 211 billion women [4]. These data show that a significant proportion of the population in emerging nations has CKD, and this proportion seems to grow daily. Much research has been conducted on early screening to address CKD at its earliest stages. This paper emphasizes the potential of precision in machine learning (ML) predictive algorithms for CKD. ML, with its ability to enable complex analysis, reduce human error, and improve prediction accuracy, offers hope in the fight

against CKD. When the two kidneys become damaged, a common type of kidney illness known as CKD develops, causing long-term symptoms. This kidney disorder, which may lead to poor kidney functionality, is referred to as kidney failure. A medical condition or a deficiency in necessary nutrients, such as a decrease in the glomerular filtration rate (GFR) [5], might be the result. The proposed forecasting approach uses an ensemble technique with five ML classifiers: gradient boosting (GB), support vector machine (SVM), random forest (RF), decision tree (DT), and K-nearest neighbors (KNN) as baseline learners to forecast outcomes based on medical data input. ML is becoming more critical in identifying medical conditions because it enables complex analysis, reduces human error, and improves prediction accuracy. ML algorithms are considered trustworthy for predicting gastrointestinal disease, cardiovascular disease, type 2 diabetes, and cancers [6].

Various healthcare data affect the produced model's stability and adaptability and lead to deceptive guidelines and repeatable clinical models; it comes with several disadvantages. As a result, the learning procedure in deep learning (DL) may result in a high-variance network and fail to accomplish optimal parameters automatically. A variety of DL frameworks could be used to address this difficulty. We refer to this procedure as ensemble learning, a powerful approach that combines the benefits from conventional and ensemble learning to overcome the limitations of individual models and provide an increased adaptability and broadly applicable approach [7]. Essential learners and variation are the two primary types of ensemble learning [8]. Initially, combining various data sets leads to homogeneous learning. Secondly, using multiple frameworks, diverse development can be accomplished. Among the several configurations used to build ensemble models are stacking [9], boosting [10], and bagging [11]. The stacked ensemble model offers a versatile, resilient, and flexible approach to the investigation. Numerous research has shown that ensemble modeling produces a reliable and efficient framework, reassuring us about the robustness of the approach.

They selected the best feature selection, usually the first step towards creating an effective model. In the field of ML, selecting features was thoroughly studied, showing promise for use in medical fields. Three primary categories of feature selections exist: wrapping, filtering, and embedding [12]. The research team used four feature evaluation techniques to pick the best feature selection. The primary goal was to create an ensemble model to enhance predictive effectiveness while utilizing the best feature subset. Compared with the present methods, the suggested feature shows great promise in the earlier diagnosis of CKD from a medical approach, encouraging us about the potential impact of the research.

Therefore, to address a gap in this area, we will explore multiple ML techniques in this study along with an ensemble strategy to combine these algorithms. Consequently, this paper's primary contributions are as follows:

- To propose an ML-based ensemble model based on a majority voting approach to combine five ML models as a baseline classifier (GB, SVM, KNN, RF, and DT) with fine tuning using grid search to enhance detection and classification performance.
- To examine the benefits and drawbacks of each prediction approach, evaluate its effectiveness through a range of measures. The outcomes are contrasted with those of the current techniques to illustrate the power of the suggested models on the datasets.

The paper's organization is as follows: section 1 presents an overview of CKD. Section 2 discussed the previous CKD prediction and classification literature. The step-by-step methodology is explained in the section 3 in detail. Section 4 measures the performance of the suggested models and compares the final results. Section 5 discusses this study's conclusion and future direction.

2. RELATED WORK

This section discusses algorithm-related investigations and evaluates specific strategies based on their performance. Applying the data mining approach to the specialized examination of healthcare records provides a valuable approach to investigation [13]. Compared with the naïve Bayes (NB) technique, the DT approach achieved a 92% accuracy score, 93% specificity, and 94% sensitivity for classifying diabetic datasets. Additionally, researchers discovered that mining helps recover correlations between traits that are no longer predictive of the outcomes they attempt to forecast. Predictive algorithms using ML approaches, such as logistic regression (LR), SVM, KNN, and DT classification algorithms for CKD forecasting, were discussed by investigators [14]. The study demonstrated that the SVM algorithm had the highest accuracy score, reaching 97%. The proposed technique's learning and testing yielded the highest sensitivity results for SVM. Based on this analysis, it is possible to conclude that chronic kidney failure can be predicted using the SVM algorithm. The research selected and analyzed three distinct techniques [15] to obtain an appropriate prediction rate across the dataset. The study used the GB classifier, which produced the most effective results. While AdaBoost and linear discriminant analysis (LDA) achieved a 96% performance score, the GB classifier achieved a 98% performance score. Additionally, compared to other ML classifiers, the GB classifier requires more time to produce a forecast but provides better-predicted results on both the receiver operating characteristic (ROC) and area under the curve (AUC) scores. Therefore, accurate prediction

depends heavily on the initial processing plan, and preprocessing techniques were used cautiously to achieve the expected outcomes properly. Investigators predicted CKD using a novel selection method [16]. By applying certain classifications and appropriately assessing the overall outcome, CKD is projected to be used in this work. The NB, RF, and artificial neural network (ANN) [17] classifiers were evaluated, and it was found that the RF outperforms the other models. The value of CKD prediction has increased over time. Implementing several feasible adaptive strategies can enhance the recommended classifiers' performance. NB, RF, and KNN were used to forecast CKD. The early identification of CKD aids in promptly treating people afflicted and stops the illness from worsening.

An ML predictive technique for early identification of CKD was created in [18]. The predictive models have been assessed and verified for the initial features provided by the dataset, which contains input features collected through the CKD dataset. DT, RF, and SVM classifiers were built to diagnose CKD. The predictive model's performance score served as the basis for evaluating the models' performance analysis. In contrast, the study's findings demonstrated that the RF model outperforms DT and SVM models regarding CKD prediction. In addition to being necessary for eliminating impurities from the human body, the kidneys also regulate BP, the body's perception of the pH level, and its level of electrolyte. In between malfunctioning in each body organ, dysfunction contributes to minor to fatal disorders. Consequently, scientists from all around humanity have devoted their efforts to developing methods for precisely diagnosing and treating CKD. The number of health conditions that ML classifiers can identify includes CKD, as these algorithms are being utilized more and more in medical research for identification. The process and outcome accuracy have gradually improved due to research into using ML techniques to identify CKD. Out of all the classification methods, investigators suggested that the RF model achieves a 99% accuracy score, which was the most efficient. The research shows how to effectively handle the absence of values in data using four different approaches: statistical procedures. Additionally, it assesses how well ML models work in two scenarios, one in which the hyperparameters are tuned and the other in which they are not, and finds that the algorithms' effectiveness has significantly improved, as shown in [19]. The work aims to investigate the suitability of particular supervised ML models in the biomedical domain and assess their capacity to identify various severe illnesses, including the earlier detection of CKD [20].

Researchers have tried to identify kidney disease earlier on or forecast its emergence. While disease forecasting suggests the underlying disease can occur throughout the future, disease identification suggests the individual now has the illness. Consequently, two lines of research have been established in this field: identification and forecasting. With the first category, there have been a lot of investigations in this area [21]. After examining the previous study, we encountered several research gaps:

- The data on CKD remains insufficient. Medical testing records became the basis for earlier research, but they cover a limited number of instances.
- The earlier studies focused on identifying the disease after it had already manifested.
- The research in this area has never been thoroughly investigated because there is no information.
- A single prior study attempted to forecast illness beforehand. Nonetheless, this study's accuracy was lacking.
- The CDK disease death rate proliferates based on the preceding issues.

3. MATERIAL AND METHODS

This research presents an innovative ensemble ML strategy to identify CKD, employing advanced techniques for optimal feature selection. The selected features significantly impact CKD from a medical perspective. The proposed framework consists of several phases: dataset collection, preprocessing, feature selection, model development, and optimization, as depicted in Figure 1. Each step is described in detail as follows.

3.1. Dataset description

The UC Irvine Machine Learning Library provided the standard CKD dataset used in this investigation [22]. Many researchers utilized this dataset to conduct experiments. The dataset consists of 400 cases, 250 with no CKD and 150 with CKD. Figure 1 shows that each class label contains two values: 1 and 0 for CKD and no CKD, respectively. Figure 2 shows the number of observations in the dataset.

3.2. Data preprocessing

Healthcare information contains normal, abnormal, or null values due to the instrument, network, or record entry operator. Challenges with datasets have a detrimental impact on the development of an ML model. This stage's primary goal is to handle outliers and missing numbers to improve the quality of the healthcare dataset.

3.2.1. Data encoding

The dataset we work with contains both categorical and numeric variables. It's crucial to understand that two feature-selection methods work better with numerical characteristics than categorical ones: ML.

Therefore, all categorical characteristics were encoded using the scikit learning library's labeling encoder package, a key technical step in our process.

3.2.2. Filling missing values

We follow a meticulous process when filling in missing data. Several techniques were proposed, and the choice depends on how much data and features are absent. When the amount of missing data is modest (5% to 10%), traditional statistical approaches like mean, maximum, and mode function well. However, when the percentage of missing values rises (20% to 50%), advanced methods like expectation maximization are needed [23]. In our case, we use the feature averages to impute the missing values, ensuring the quality and integrity of our dataset.

3.2.3. Removing outliers

We take a thorough approach to identify and remove outliers, which are parameters that significantly deviate from the typical range of every feature value. This is a crucial step in the creation of a robust and comprehensive model [24]. In the present investigation, we first examined all data statistically and then verified the findings from a healthcare perspective. Any outliers identified in the data were substituted with the feature average, ensuring the robustness of our model.

3.2.4. Data standardization and normalization

The standard MinMaxScaler() function was employed for scaling feature values. In (1) was used for scaling the numerical values for batch normalization and standardization. In this case, the standard deviation has been configured to six, and the data is assumed to be zero.

$$S(D) = \frac{\sum_{i=1}^{S} d_i - d_{min}}{d_{max} - d_{min}} \quad (1)$$

where, S, D, d_i, d_{max}, d_{min} represents the number of data instances in dataset, is the average of the characteristics, the lowest and maximal instances values, respectively.

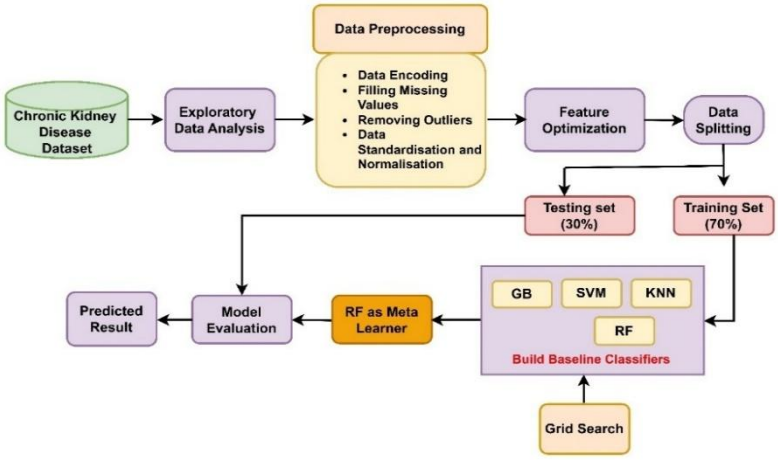

Figure 1. Proposed framework to predict CKD

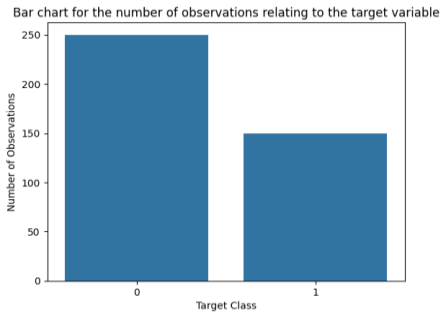

Figure 2. Distribution of dataset

3.3. Feature optimization

Figure 3 demonstrates positive relationships between each feature. Conversely, there are some negative correlations with hemoglobin, potassium, red blood cell count, and white blood cell count. The associations between each feature can be seen by examining the heatmap of the patient information displayed in Figure 3. Figure 4 analysis of the histograms is essential for feature selection, as it helps identify outliers, the need for normalization and also helps to shows the overall importance of each feature in detecting CKD.

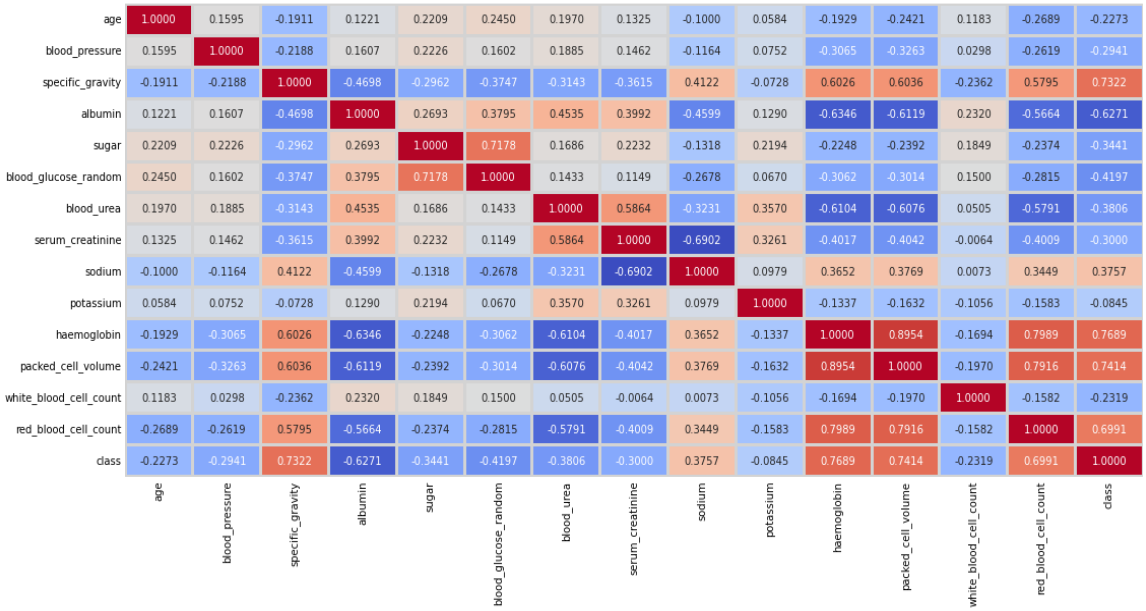

Figure 3. Heatmap of the patient information

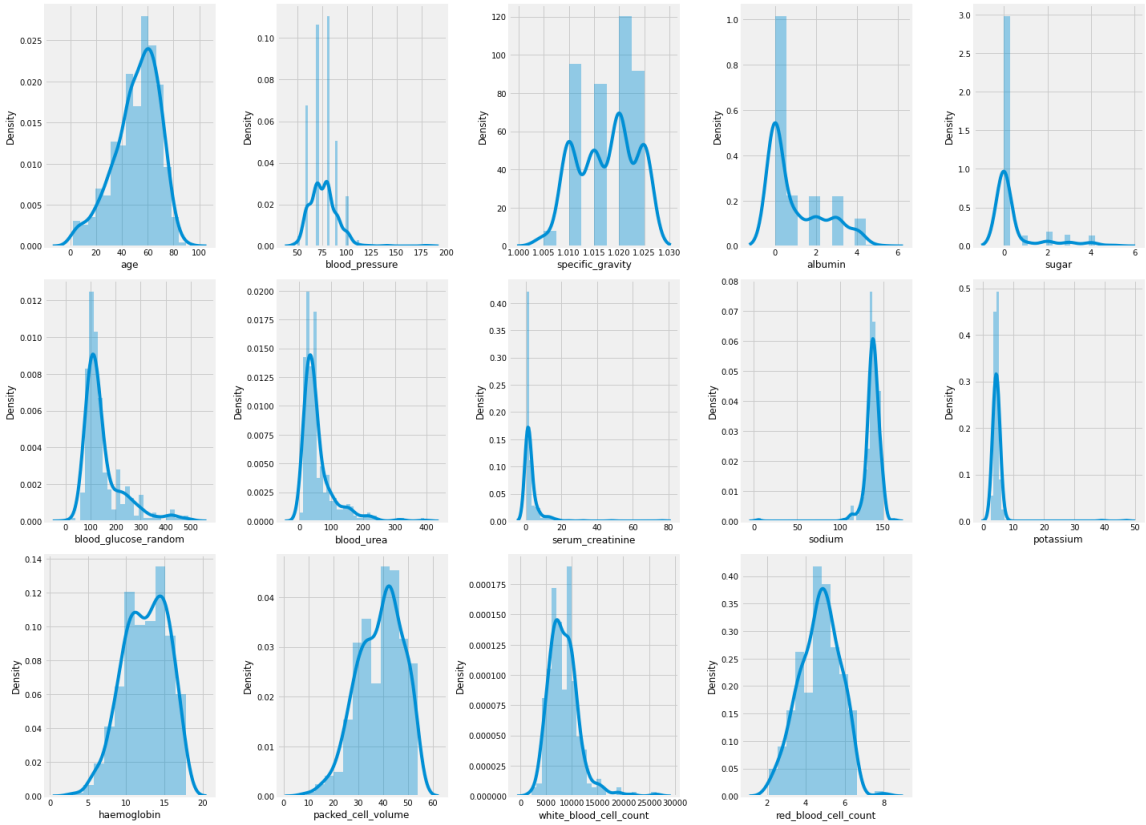

Figure 4. Heatmap of each feature

3.4. Build machine learning models

Classification techniques are crucial for training and testing supervised learning systems. Baseline ML classifiers utilize data from the training and testing sets to achieve the desired results. The selection of an appropriate classification method is crucial, as it significantly impacts the accuracy, generalization, and overall performance of the ML system.

3.4.1. Decision tree

The DT algorithm operates like a graph or tree-like structure containing the root nodes and sub-nodes, like leaves. The characteristics in this case are the sub-nodes, while the subdivisions represent the results of each examination on each node. It is among the most widely used algorithms for categorization since it may function without requiring location barriers or a wealth of field data [25].

3.4.2. Random forest

The RF technique has become the most efficient among the several ML techniques. It has been applied to forecast and probabilistic calculations. Numerous DTs make up the RF classifier.

3.4.3. K-nearest neighbor

A popular supervised learning technique for handling regression and categorization problems includes the kNN. Selecting the numerical value of the instance (k), which is nearest to the query, and determining the Euclidian distances among every example with the data are the steps involved in kNN decision-making. From there, the more frequent label with categorization for regression is selected. The kNN approach's performance is increased by randomly selecting the value of k.

3.4.4. Gradient boosting

GB uses DT all the time. It is based on the idea that the optimal future model can decrease the total error in forecasting when combined with previous models. The critical notion becomes to reduce errors by specifying the expected outcomes for this next model. The desired results for every instance are ascertained using the gradient of the error rate with respect to the forecast.

3.4.5. Support vector machine

A hyperplane has been employed by the SVM to divide the data set into two groups. The technique looks for a decision threshold to optimize the margin and reduce categorization errors among the two groups. The process of finding SVM accomplishes this. The data elements nearest to the decision border are these. Additional data elements can be categorized by identifying which part of the hyperplane it corresponds to after the hyperplane has been discovered. When projecting the input within a high-dimensional plan, the kernel within support vector classification (SVC) is utilized for handling non-linearly separated data [26], [27].

3.4.6. Ensemble model

To build the ensemble model used, five baseline ML classifiers, GB, RF, KNN, SVM, and DT, were trained using grid search to optimize hyperparameters CKD prediction models and achieved the best performance. Each classifier was trained on the CKD dataset, and their predictions were used to create a new dataset that contained these predictions as features. This new dataset was then used to train a meta-learner, specifically an RF model optimized by grid search. Meta learner combined the predictions of the baseline classifiers to produce a final, more accurate prediction. This ensemble approach leverages the strengths of multiple models to improve prediction accuracy for CKD detection.

The proposed Pseudocode 1 outlines the construction of an ensemble learning model using CKD dataset D. Initially, multiple baseline models (GB, RF, KNN, SVM, DT) are optimized using grid search and trained on D. Their predictions are then stacked to form a meta-dataset D', on which a meta-learner (RF) is trained to generate the final ensemble model, improving classification performance.

Pseudocode 1: Ensemble model

```
Input: CKD Dataset D = {(x1, y1), (x2, y2)..... (xn, yn)}
       ML Models GB, RF, KNN, SVM, DT
Output: Ensemble E
Begin
Step-1: Grid Search and Train Baseline ML Models
    Initialize hyperparameter grids for GB, RF, KNN, SVM, DT

    Perform grid search and cross-validation for each model:
```

```
    For model in [GB, RF, KNN, SVM, DT]:
        Perform Grid Search on model with dataset D
        Select best model M_a based-on grid search results
        Train best model M_a on dataset D
        Save trained model M_a
E = [M_GB, M_RF, M_KNN, M_SVM, M_DT]
end for
Step-2: Create New Dataset D' for Meta Learner
    Initialize D' = []
    For each sample (x_i, y_i) in dataset D:
        Initialize feature vector s_i = []
        For each model M_a in E:
            r_ai = Predict class label of x_i using M_a
            Append r_ai to s_i
        Append (s_i, y_i) to D'
   end for
Step-3: Train Meta Learner (RF) on D'
    Perform Grid Search on RF with dataset D'
    Select best RF model M_RF_meta based on grid search results
    Train best RF model M_RF_meta on dataset D'
Step-4: Return the trained Ensemble Model
  Return M_RF meta
End
```

3.5. Loss function

The loss function for an ensemble model, especially one using a meta-learner like RF, typically involves the combined error from all the base learners. Simplified representation of the loss function for such an ensemble model:

- Base learner loss function: for each base learner M_a, the loss functions L_a is computed on the training dataset D.
- Meta learner loss function: the meta learner M_{meta} uses the predictions from all base learners to create a new dataset D'. The loss function L_{meta} is computed on this new dataset.
- Loss function for base learners: for each base learner Ma.

$$L_a = \frac{1}{2} \sum_{i=1}^{n} L(y_i, M_a(x_i)) \quad (2)$$

Where n represent number of features, $y_i$is the true label for features x_i, $M_a(x_i)$ is the prediction of the base learner. M_a for sample x_i, and L is the cross-entropy loss function for classification.

- Loss function for meta learner: for the meta learner M_{meta}:

$$L_{meta} = \frac{1}{m} \sum_{j=1}^{m} L\left(y_i, M_{meta}(S_j)\right) \quad (3)$$

Where m represent number of features in the new dataset D'. y_i is the true label for the new feature vector S_j, $M_{meta}(S_j)$ is the prediction of the meta learner M meta for the new feature vector S_j, and S_j is the new feature vector consisting of predictions from all base learners for the original sample x_j

- Combined loss function: the overall loss function for the ensemble model can be seen as the combination of the losses from the base learners and the meta learner:

$$L_{ensemble} = \sum_a \alpha_a L_a + \beta L_{meta} \quad (4)$$

Where α_a and β represents the weights that balance the importance of each base learner's loss and the meta learner's loss.

3.6. Hyperparameter settings

The random grid search technique was employed for hypermeter configuration to attain optimal performance regarding the computation efficiency of the suggested baseline ML classifiers and ensemble model as shown on Table 1. Grid search allows for systematically examining different variations of hyperparameters by providing a sequence of values corresponding to every parameter. This guarantees that every possibility is explored to determine the hyperparameters' desired values. Because grid search seems predictable, it consistently produces identical results with similar information and parameters. This feature makes it easier to compare data repeatedly, promoting accurate analysis and evaluation. Grid search is simple to use and is one of its main benefits.

Table 1. Hyperparameter setting of proposed baseline ML classifiers and ensemble model

Classifiers	Parameters
GB	parameters_grid_gb = {'n_estimators': [50, 100, 200],'learning_rate': [0.01, 0.1, 0.2], 'max_depth': [3, 4, 5],'subsample': [0.7, 0.8, 0.9]}
SVM	parameters_grid = {'kernel': ['poly', 'rbf', 'linear', 'sigmoid'], 'C': [0.1, 1, 10, 100, 1000], 'gamma': ['scale', 'auto'], 'shrinking': [True, False]}
KNN	knn_parameters_grid = {'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], 'weights': ['uniform', 'distance'],'algorithm':['auto', ball_tree', 'kd_tree', 'brute'], 'n_jobs':[1, -1]}
DT	dt_parameters_grid = {'criterion': ['gini', 'entropy'], 'splitter': ['best', 'random'], 'min_samples_leaf': [1, 2, 3, 4, 5], 'max_features': ['auto', 'sqrt', 'log2']}
RF	rf_parameters_grid = {'n_estimators': [10, 30, 40, 50, 60, 70, 80, 90, 100], 'criterion': ['gini', 'entropy'], 'min_samples_split': [1.0, 2, 3, 4, 5], 'max_features': ['auto', 'sqrt', 'log2']}
Ensemble	param_grid = {'n_estimators': [10, 50, 100], 'criterion': ['gini', 'entropy'],'min_samples_split': [2, 5, 10],'max_features': ['auto', 'sqrt', 'log2']}

4. RESULT AND DISCUSSION

This section presents the effectiveness of the suggested baseline ML classifiers and the suggested ML-based ensemble classifier in detecting and classifying the CKD on the dataset based on the feature optimization technique. All the baseline ML classifiers were tuned using random grid search, and all classifiers were built using Keras and the TensorFlow library. Grid search was utilized to optimize the meta-learner classifier. Google Colaboratory was used to conduct all of the tests. We used the CKD dataset for these experiments. Using stratified sampling, divide the dataset into two sets: 70% training with 280 samples and 30% testing with 120 samples. The performance of ML models was measured using several evaluation parameters such as accuracy, precision, recall, and F1-score. True positive (TP), false positive (FP), true negative (TN), and false negative (FN) represent the units of calculation used for all of them.

$$Accuracy = \frac{T\_P + T\_N}{T\_P + T\_N + F\_P + F\_N} \quad (5)$$

$$Precision = \frac{T\_P}{T\_P + F\_P} \quad (6)$$

$$Recall = \frac{T\_P}{T\_P + F\_N} \quad (7)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Table 2 shows the performance analysis of various baseline ML classifiers and an ensemble model based on their metrics. The GB and RF both achieve a high accuracy of 97.5%, with GB having a precision of 95.4% and recall of 97.6%, resulting in an F1-score of 96.5%. RF demonstrates perfect recall at 100%, slightly lower precision at 94.7%, and the highest F1 score at 97.3%. KNN shows the lowest accuracy at 92.5%, precision of 93.0%, recall of 91.4%, and an F1-score of 92.1%. SVM performs well with 95.8% accuracy, 94.7% precision, 96.4% recall, and a 95.6% F1 score. DT achieves 93.3% accuracy, 89.5% precision, 96.2% recall, and a 92.7% F1 score. The ensemble model consistently performs consistently across all metrics, achieving 97.5% accuracy, precision, and recall and a nearly perfect F1-score of 97.4%. Tables 3 to 8 show the classification report of each baseline ML model. It is indicated that each model performed better, with a maximum accuracy score of 97.00%. Table 8 also shows the classification report of the proposed ensemble model.

Table 2. Performance analysis of proposed baseline ML and ensemble model

Classifiers	Accuracy	Precision	Recall	F1-score
GB	97.5	95.4	97.6	96.5
RF	97.5	94.7	100	97.3
KNN	92.5	93.0	91.4	92.1
SVM	95.8	94.7	96.4	95.6
DT	93.3	89.5	96.2	92.7
Ensemble	97.5	97.5	97.5	97.4

Table 3. Classification report of GB

	Precision	Recall	F1-score	Support
0	0.97	1.00	0.98	84
1	1.00	0.92	0.96	36
Accuracy			0.97	120
Macro Avg	0.98	0.96	0.97	120
Weighted Avg	0.98	0.97	0.97	120

Table 4. Classification report of SVM

	Precision	Recall	F1-score	Support
0	0.97	1.00	0.98	84
1	1.00	0.92	0.96	36
Accuracy			0.97	120
Macro Avg	0.98	0.96	0.97	120
Weighted Avg	0.98	0.97	0.97	120

Table 5. Classification report of KNN

	Precision	Recall	F1-score	Support
0	0.95	0.98	0.96	84
1	0.94	0.89	0.91	36
Accuracy			0.95	120
Macro Avg	0.95	0.93	0.94	120
Weighted Avg	0.95	0.95	0.95	120

Table 6. Classification report of RF

	Precision	Recall	F1-score	Support
0	0.97	1.00	0.98	84
1	1.00	0.92	0.96	36
Accuracy			0.97	120
Macro Avg	0.98	0.96	0.97	120
Weighted Avg	0.98	0.97	0.97	120

Table 7. Classification report of DT

	Precision	Recall	F1-score	Support
0	0.97	1.00	0.98	84
1	1.00	0.92	0.96	36
Accuracy			0.97	120
Macro Avg	0.98	0.96	0.97	120
Weighted Avg	0.98	0.97	0.97	120

Table 8. Classification report of ensemble model

	Precision	Recall	F1-score	Support
0	0.97	1.00	0.98	84
1	1.00	0.92	0.96	36
Accuracy			0.97	120
Macro Avg	0.98	0.96	0.97	120
Weighted Avg	0.98	0.97	0.97	120

Figure 5 shows the confusion matrix of the proposed baseline ML and ensemble model for calculating the TP, FN, TN, and FP values. The proposed GB, SVM, DT, RF, and ensemble models all exhibit similar predictive values with 84 TPs, 0 FNs, 33 TNs, and 3 FPs, indicating perfect recall due to the absence of FNs and high specificity with only 3 FPs out of 36 TNs. KNN shows slightly lower performance with 82 TPs, 2 FNs, 32 TNs, and 4 FPs, still maintaining high sensitivity but slightly decreasing specificity. This analysis indicates that all models, except KNN, provided almost identical and superior performance in identifying positive and negative cases, with KNN performing slightly lower in comparison.

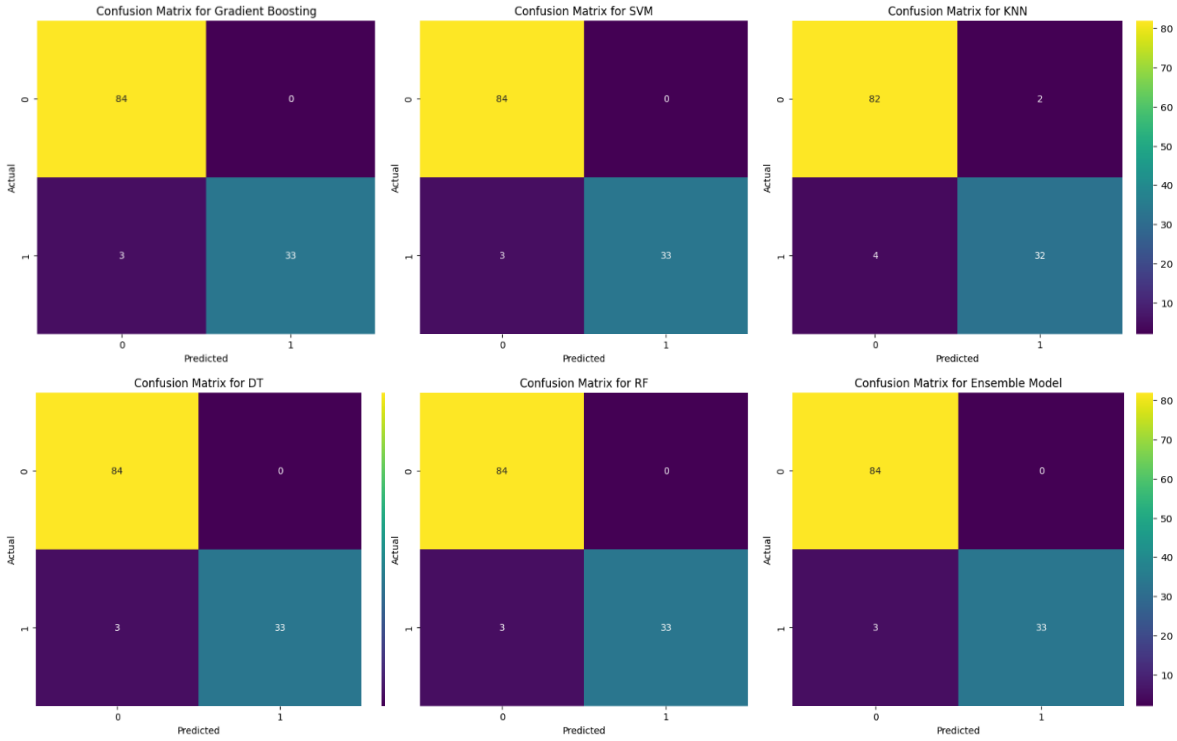

Figure 5. Confusion matrix for proposed classifiers

Figures 6 to 11 show the 3D plots for GB, SVM, KNN, DT, RF, and ensemble, which indicate the relationship between various hyperparameters and model accuracy. There are three 3D plots of each model. The first plot examines the effect of varying the number of estimators (n_estimators) of each model and the subsample ratio on accuracy, with a specific configuration of 200 estimators and a subsample of 0.7; C parameter is one and shrinking was set to true; minimum sample leaf set to 4 for DT; and minimum sample splitting values set to 10 for RF; minimum sample splitting values set to 2 and maximum depth is

10 for ensemble. It illustrates how the accuracy changes as the parameters are adjusted, showing potential peaks at specified values. The second plot explores the effect of changing the learning rate and subsample ratio on the accuracy set to a learning rate of 0.1 and a subsample of 0.7 for GB. The best parameter gamma is one, and shrinking is true for SVM, emphasizing how accuracy responds to parameter changes, potentially revealing optimal regions for these settings. The third plot focused on the interaction between the learning rate and the number of estimators at learning rates between 0.1 and 200 estimators and how these parameters influence the model's performance, helping to identify the best combination for maximum accuracy.

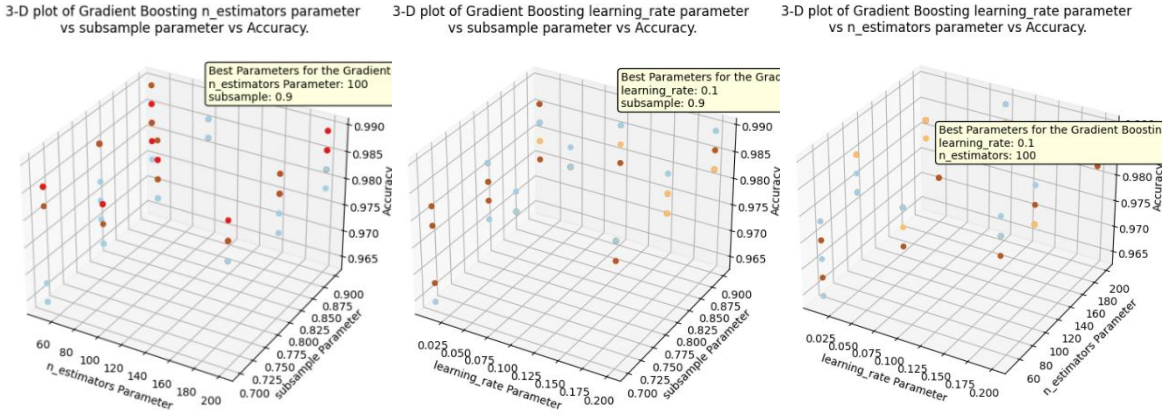

Figure 6. 3D plotting of GB

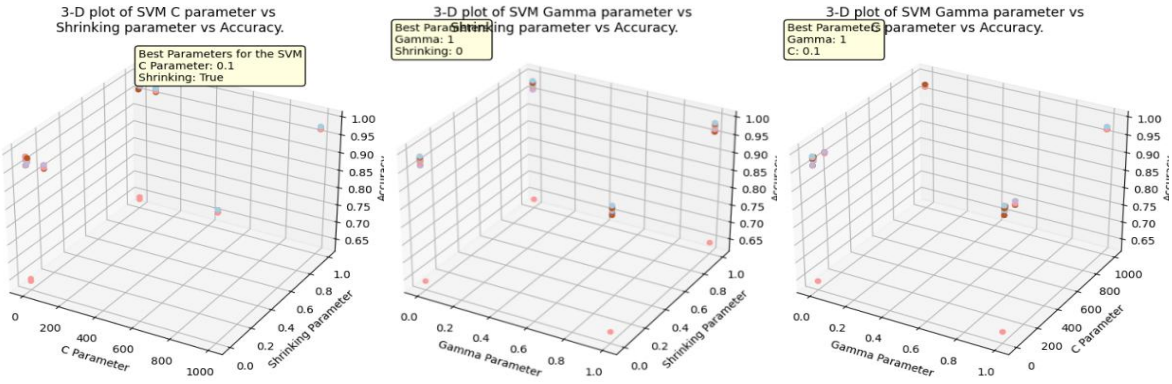

Figure 7. 3D plotting of SVM

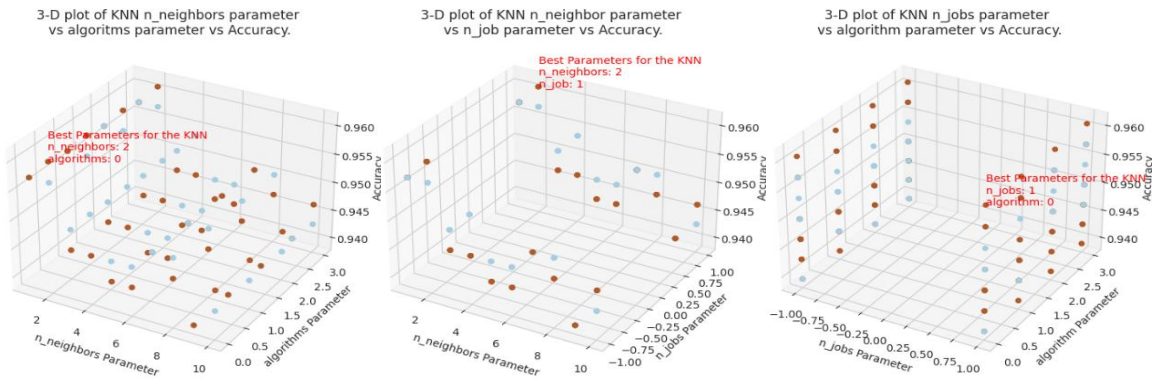

Figure 8. 3D plotting of KNN

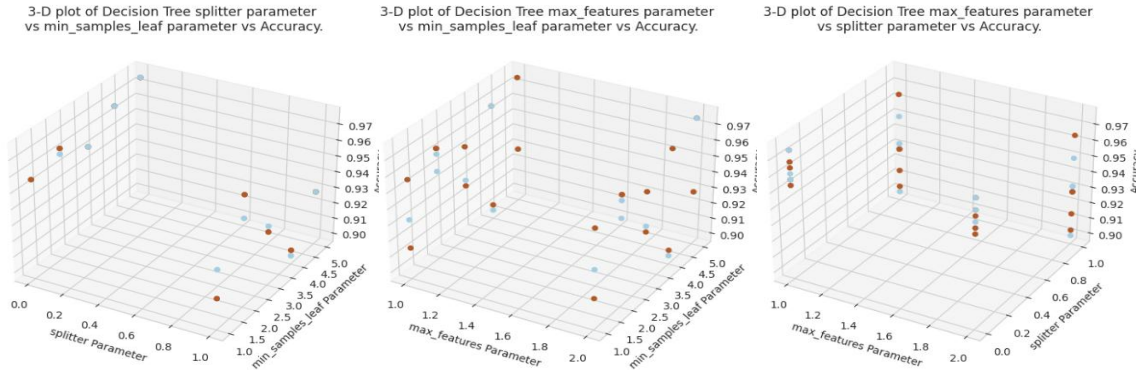

Figure 9. 3D plotting of DT

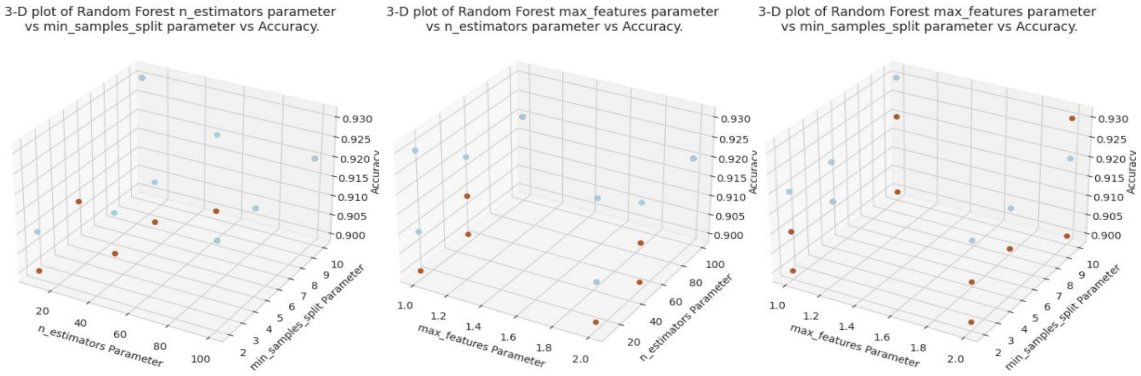

Figure 10. 3D plotting of RF

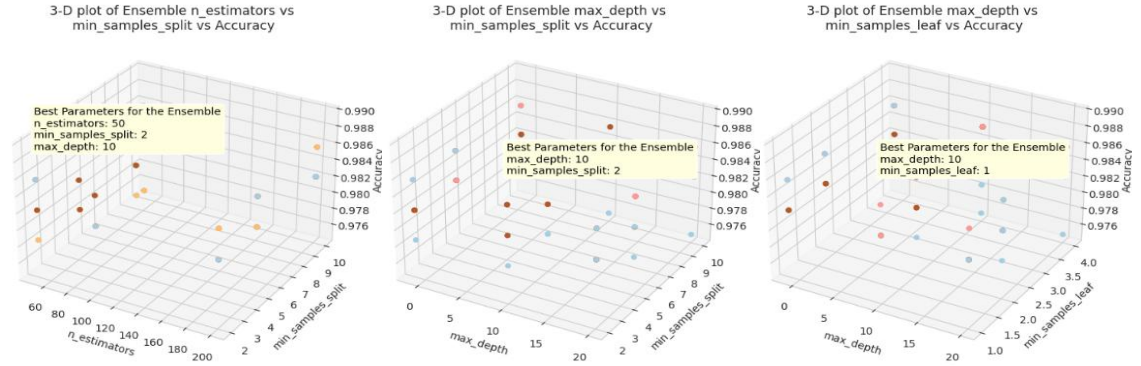

Figure 11. 3D plotting of ensemble model

Figure 12 shows the ROC-AUC curve that indicates the performance of all proposed ML and ensemble models in distinguishing between classes based on their true positive rate (TPR) and false positive rate (FPR) scores. The ROC-AUC curves for GB, RF, and the ensemble model show excellent performance with an AUC of 0.982. The SVM also performs very well with an AUC of 0.963, followed by the DT with an AUC of 0.941, and KNN with an AUC of 0.934, all significantly outperforming the baseline random performance represented by an AUC of 0.5.

Table 9 shows the comparative analysis of the proposed ensemble model with various existing methods used for CKD detection in terms of accuracy score. Kumar *et al.* [28] utilized SVM, DT, and ANN, achieving 92.76% accuracy, while Walse *et al.* [29] used RF, NB, and DT with 91.00% accuracy. Pal [30] employed LR, DT, and bagging, attaining 95.92%. Raza [31] and Atallah and Mousa [32] used majority voting ensemble (MVE), recording 88.00% and 90.00% accuracy, respectively. Neloy *et al.* [33] with weighted average ensemble (WAE) reached 93.00%, and Revathy *et al.* [20] reported lower accuracies for RF (74.00%),

DT (73.5%), and LR (72.2%). Hassan *et al.* [34] achieved a high accuracy of 97.33% using an ensemble approach, Islam *et al.* [35] reported an accuracy of 97.00% using XGBoost. Saif *et al.* [36] utilized convolutional neural network (CNN) and long short-term memory (LSTM), achieving 94.00%. The proposed ensemble model with grid search outperforms all these methods, achieving the highest accuracy of 97.5%, demonstrating its superior efficacy in CKD detection.

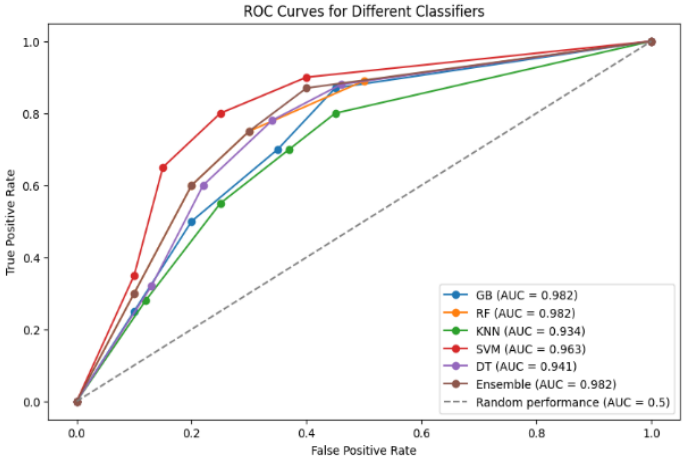

Figure 12. ROC-AUC curve

Table 9. Comparative analysis of proposed the ensemble model with existing methods

References	Methods used	Accuracy in %
Kumar *et al.* [28]	SVM, DT, ANN	92.76
Walse *et al.* [29]	RF, NB, DT	91.00
Pal [30]	LR, DT, Bagging	95.92
Raza [31]	MVE	88.00
Atallah and Mousa [32]	MVE	90.00
Neloy *et al.* [33]	WAE	93.00
Revathy *et al.* [20]	RF	74.00
	DT	73.50
	LR	72.20
Hassan *et al.* [34]	Ensemble	97.33
Islam *et al.* [35]	XGBoost	97.00
Saif *et al.* [36]	CNN, LSTM	94.00
Proposed	Ensemble with grid search	97.5

5. CONCLUSION AND FUTURE SCOPE

In many instances, combining the results of multiple independent ML models can reduce generalization errors and provide improved results. As a result, the ensemble strategy has emerged as a reliable and industry-leading approach in many domains. The basic concept behind ensemble modeling is to train multiple models and combine their findings with one of several ensemble strategies. Researchers have tried to detect or predict the onset of kidney disease. The practical implications of the investigation include the fact that previous investigations have been interested in diagnosing diseases. Additionally, previous algorithms have often shown poor performance. The central part of this study is to predict CKD using an ML-based ensemble. This research was conducted on a standard CKD dataset collected from Kaggle. We propose five predictive ML models as baseline classifiers in this study: GB, SVM, KNN, RF, and DT, and configure the grid search hyperparameters of each classifier. Then, we combine all the baseline classifiers to form an ensemble model with fine-tuning using grid search hyperparameters and classification majority voting techniques to improve performance. The final experimental results indicate that the combined model consistently performs better, with 97.5% accuracy, precision, and recall and an almost perfect F1-score of 97.4%, compared to individual classifiers and existing methods. This robust performance of the ensemble model is a reliable tool for earlier CKD detection, which is critical for timely intervention and treatment. The integration of ML into CKD detection not only provides higher levels of accuracy but also paves the way for automated, scalable diagnostic tools that can help medical practitioners make accurate decisions and patient conditions eventually improve. There are numerous avenues for future research direction in CKD prediction. Expanding the dataset to include diverse populations and more varied clinical settings could enhance the

model's generalizability and robustness. Additional advanced ML algorithms, such as pre-trained DL models, might improve accuracy and predictive capabilities. Moreover, integrating multi-modal data, including environmental and lifestyle factors, could provide a more reliable evaluation of CKD risk factors and early detection. Additionally, explainable AI techniques can help demystify the ensemble model's decision-making process, thereby increasing trust and adoption among medical practitioners. Collaborations with healthcare institutions for pilot studies and longitudinal research can validate and refine the model, ensuring its practical utility and impact. Finally, exploring integrating CKD detection systems with electronic health records (EHR) and other healthcare IT infrastructures can streamline workflows and improve patient management outcomes.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Deepika Amol Ajalkar	✓	✓			✓			✓	✓	✓		✓		✓
Jyoti Yogesh Deshmukh	✓		✓	✓		✓	✓		✓	✓	✓			
Mayura Vishal Shelke	✓				✓	✓			✓	✓		✓	✓	
Shalini Vaibhav Wankhade	✓	✓		✓	✓				✓		✓		✓	
Shwetal Kishor Patil		✓	✓			✓	✓	✓		✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] H. Khalid, A. Khan, M. Z. Khan, G. Mehmood, and M. Shuaib Qureshi, "Machine learning hybrid model for the prediction of chronic kidney disease," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/9266889.
[2] Y. Dubey, P. Mange, Y. Barapatre, B. Sable, P. Palsodkar, and R. Umate, "Unlocking precision medicine for prognosis of chronic kidney disease using machine learning," *Diagnostics*, vol. 13, no. 19, Oct. 2023, doi: 10.3390/diagnostics13193151.
[3] E. F. Carney, "The impact of chronic kidney disease on global health," *Nature Reviews Nephrology*, vol. 16, no. 5, pp. 251–251, May 2020, doi: 10.1038/s41581-020-0268-7.
[4] A. Giandalia *et al.*, "Gender differences in diabetic kidney disease: focus on hormonal, genetic and clinical factors," *International Journal of Molecular Sciences*, vol. 22, no. 11, May 2021, doi: 10.3390/ijms22115808.
[5] F. Viazzi *et al.*, "Natural history and risk factors for diabetic kidney disease in patients with T2D: lessons from the AMD-annals," *Journal of Nephrology*, vol. 32, no. 4, pp. 517–525, Aug. 2019, doi: 10.1007/s40620-018-00561-3.
[6] A. Triantafyllidis *et al.*, "Deep learning in mHealth for cardiovascular disease, diabetes, and cancer: systematic review," *JMIR mHealth and uHealth*, vol. 10, no. 4, Apr. 2022, doi: 10.2196/32344.
[7] S. Kumari and S. K. Singh, "An ensemble learning-based model for effective chronic kidney disease prediction," in *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, IEEE, Nov. 2022, pp. 162–168, doi: 10.1109/ICCCIS56430.2022.10037698.
[8] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Early prediction of chronic kidney disease based on ensemble of deep learning models and optimizers," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 1, Apr. 2024, doi: 10.1186/s43067-024-00142-4.
[9] E. Dritsas and M. Trigka, "Machine learning techniques for chronic kidney disease risk prediction," *Big Data and Cognitive Computing*, vol. 6, no. 3, Sep. 2022, doi: 10.3390/bdcc6030098.
[10] S. M. Ganie, P. K. Dutta Pramanik, S. Mallik, and Z. Zhao, "Chronic kidney disease prediction using boosting techniques based on clinical parameters," *PLOS ONE*, vol. 18, no. 12, Dec. 2023, doi: 10.1371/journal.pone.0295234.
[11] O. A. Jongbo, A. O. Adetunmbi, R. B. Ogunrinde, and B. Badeji-Ajisafe, "Development of an ensemble approach to chronic kidney disease diagnosis," *Scientific African*, vol. 8, Jul. 2020, doi: 10.1016/j.sciaf.2020.e00456.
[12] S. A. E.-Mienye, T. G. Swart, E. Esenogho, and I. D. Mienye, "A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease," *Bioengineering*, vol. 9, no. 8, Jul. 2022, doi: 10.3390/bioengineering9080350.

[13] P. Ghosh, F. M. J. M. Shamrat, S. Shultana, S. Afrin, A. A. Anjum, and A. A. Khan, "Optimization of prediction method of chronic kidney disease using machine learning algorithm," in *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, IEEE, 2020, pp. 1–6, doi: 10.1109/iSAI-NLP51646.2020.9376787.
[14] K. R. A. Padmanaban and G. Parthiban, "Applying machine learning techniques for predicting the risk of chronic kidney disease," *Indian Journal of Science and Technology*, vol. 9, no. 29, Aug. 2016, doi: 10.17485/ijst/2016/v9i29/93880.
[15] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," in *2016 Management and Innovation Technology International Conference (MITicon)*, IEEE, Oct. 2016, pp. 80–83, doi: 10.1109/MITICON.2016.8025242.
[16] A. Levin *et al.*, "Guidelines for the management of chronic kidney disease," *Canadian Medical Association Journal*, vol. 179, no. 11, pp. 1154–1162, Nov. 2008, doi: 10.1503/cmaj.080351.
[17] S. Akter *et al.*, "Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease," *IEEE Access*, vol. 9, pp. 165184–165206, 2021, doi: 10.1109/ACCESS.2021.3129491.
[18] G.-S. Fu, Y. L.-Schwartz, Q.-H. Lin, and D. Zhang, "Machine learning for medical imaging," *Journal of Healthcare Engineering*, vol. 2019, pp. 1–2, Apr. 2019, doi: 10.1155/2019/9874591.
[19] R. Devika, S. V. Avilala, and V. Subramaniyaswamy, "Comparative study of classifier for chronic kidney disease prediction using naive Bayes, KNN and random forest," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Mar. 2019, pp. 679–684, doi: 10.1109/ICCMC.2019.8819654.
[20] S. Revathy, B. Bharathi, P. Jeyanthi, and M. Ramesh, "Chronic kidney disease prediction using machine learning models," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 6364–6367, Oct. 2019, doi: 10.35940/ijeat.A2213.109119.
[21] S. Krishnamurthy *et al.*, "Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan," *Healthcare*, vol. 9, no. 5, May 2021, doi: 10.3390/healthcare9050546.
[22] L. Rubini, P. Soundarapandian, and P. Eswaran, "Chronic kidney disease," *UCI Machine Learning Repository*. 2015, doi: 10.24432/C5G020.
[23] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996, doi: 10.1109/79.543975.
[24] L. Greco, G. Luta, M. Krzywinski, and N. Altman, "Analyzing outliers: robust methods to the rescue," *Nature Methods*, vol. 16, no. 4, pp. 275–276, Apr. 2019, doi: 10.1038/s41592-019-0369-z.
[25] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, "A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, IEEE, Jul. 2017, pp. 204–207, doi: 10.1109/ISCC.2017.8024530.
[26] A. Vijayalakshmi and V. Sumalatha, "Survey on diagnosis of chronic kidney disease using machine learning algorithms," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Dec. 2020, pp. 590–595, doi: 10.1109/ICISS49785.2020.9315880.
[27] N. V. G. Raju, K. P. Lakshmi, K. G. Praharshitha, and C. Likhitha, "Prediction of chronic kidney disease (CKD) using data science," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, IEEE, May 2019, pp. 642–647, doi: 10.1109/ICCS45141.2019.9065309.
[28] N. K. Kumar, R. L. Tulasi, and D. Vigneswari, "An ensemble multi-model technique for predicting chronic kidney disease," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 2, Apr. 2019, doi: 10.11591/ijece.v9i2.pp1321-1326.
[29] R. S. Walse, G. D. Kurundkar, S. D. Khamitkar, A. A. Muley, P. U. Bhalchandra, and S. N. Lokhande, "Effective use of naïve Bayes, decision tree, and random forest techniques for analysis of chronic kidney disease," in *Information and Communication Technology for Intelligent Systems (ICTIS 2020)*, 2021, pp. 237–245, doi: 10.1007/978-981-15-7078-0_22.
[30] S. Pal, "Chronic kidney disease prediction using machine learning techniques," *Biomedical Materials & Devices*, vol. 1, no. 1, pp. 534–540, Mar. 2023, doi: 10.1007/s44174-022-00027-y.
[31] K. Raza, "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule," in *U-Healthcare Monitoring Systems*, Elsevier, 2019, pp. 179–196, doi: 10.1016/B978-0-12-815370-3.00008-6.
[32] R. Atallah and A. A.-Mousa, "Heart disease detection using machine learning majority voting ensemble method," in *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, IEEE, Oct. 2019, pp. 1–6, doi: 10.1109/ICTCS.2019.8923053.
[33] M. A. I. Neloy, N. Nahar, M. S. Hossain, and K. Andersson, "A weighted average ensemble technique to predict heart disease," in *Proceedings of the Third International Conference on Trends in Computational and Cognitive Engineering*, 2022, pp. 17–29, doi: 10.1007/978-981-16-7597-3_2.
[34] M. M. Hassan, T. Ahamad, and S. Das, "An ensemble learning approach for chronic kidney disease prediction using different machine learning algorithms with correlation based feature selection," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2022, pp. 242–247, doi: 10.1109/ICCIT57492.2022.10055471.
[35] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," *Journal of Pathology Informatics*, vol. 14, 2023, doi: 10.1016/j.jpi.2023.100189.
[36] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Deep-kidney: an effective deep learning framework for chronic kidney disease prediction," *Health Information Science and Systems*, vol. 12, no. 1, Dec. 2023, doi: 10.1007/s13755-023-00261-8.

BIOGRAPHIES OF AUTHORS

Dr. Deepika Amol Ajalkar working as an Assistant Professor, Head of the Department of Cyber Security and Data Science at G. H. Raisoni College of Engineering and Management, Pune, Maharashtra. She received her B.E. degree (Computer Engineering) from MIT College of Enginerring, Pune, M.E. degree (Computer Engineering) from Savitribai Phule University, Pune, and Ph.D. (Computer Engineering) from Jagdishprasad Jhabarmal Tibrewala University (JJTU), Jhunjhunu, India. She has 12 years of teaching experience. Her area of research is artificial intelligence, machine learning, software engineering, and WSN. She has more than 25 publications in international/national journals and conferences, 4 patents, and 2 books. She can be contacted at email: dipikaus@gmail.com.

Dr. Jyoti Yogesh Deshmukh has over 14 years of experience. Presently working as an Associate Professor and HOD in the Department Artificial Intelligence and Data Science, Marathwada Mitramandal's Institute of Technology, Lohgaon, Pune-411047. She was awarded her Ph.D. in cloud computing and data security on research titled "Message privacy with load balancing using attribute-based encryption". She has completed her BEIT from STES's Smt. Kashibai Navale College of Engineering, Pune. She is the Ph.D. Co-guide at JJTU, Rajasthan. Her research areas are network security, artificial intelligence, machine learning, and data science. She has published 40+ papers in national and international conferences and journals, published 3 book chapters in CRC Press Taylor and Fransis group and IGI Global, and published 7 patents. She is the international co-author of one book titled "Foundations of data science for engineering problem solving" in Springer publication, Singapore, and submitted 3 research proposals to SEED-DST, SERB, and CSIR- ASPIRE research funding agencies. She is Lifetime member of ISTE Professional membership. She is a reviewer for Engineered Science, International Journal of Advanced Technology and Engineering Exploration, Frontiers in Blockchain, Concurrency and Computation: Practice and Experience, IGI Global, IEEE conferences, Applied Artificial Intelligence, and Springer ICICBDA. She can be contacted at email: jyoti1584@gmail.com.

Dr. Mayura Vishal Shelke received a Ph.D. degree in Computer Science and Engineering from VTU, Belagavi, India. She is currently working as an Assistant Professor at MIT Art, Design, and Technology University, School of Computing, Pune, Maharashtra, India. She has more than 13 years of teaching experience. She has published more than 20 papers in various national/international journals and conferences, which are indexed by Scopus. She also published 4 books and 6 patents in various computer engineering domains. She is a lifetime member of ISTE and IAENG. Her research interests include artificial intelligence, machine learning, deep learning, and computer networks. She can be contacted at email: mayura.shelke@gmail.com.

Prof. Shalini Vaibhav Wankhade received her B.Tech. from Government College of Engineering Amravati, M.Tech. from Dr. Babasaheb Ambedkar Technological University, Lonere and pursuing her Ph.D. from Sathyabama University Chennai. She is currently working as an Assistant Professor at VIIT Pune, and having more than 15 years of teaching experience. She has published more than 25 papers in various national/international journals and conferences, which are indexed by Scopus. She had published 6 patents in various computer engineering domains and worked on research project. Her research interests are ML, DL, and CNN. She can be contacted at email: shalini.wankhade@viit.ac.in.

Dr. Shwetal Kishor Patil has over 19 years of experience and is presently working as an Assistant Professor in the Computer Engineering department at Marathwada Mitramandal's Institute of Technology, Lohgaon, Pune. She holds a Ph.D. in Healthcare Machine Learning with a focus on ischemia heart disease prediction using machine learning. Additionally, she completed her M.E. from MIT AOE Alandi, Pune, and her B.E. from North Maharashtra University. She is also a Ph.D. Co-guide at JJTU, Rajasthan. She has extensive teaching experience in theory of computation (TOC), design and analysis of algorithms (DAA), data structures, and database management systems (DBMS). Her research interests include artificial intelligence, machine learning, and data science, and she has published over 35 papers in national and international conferences and journals. She is a lifetime member of IEI Professional Membership. She can be contacted at email: patilshwetal.1311@gmail.com.