Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 5 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i05.58661

PAPER

Early CKD Prediction Using Ensemble and Basic Machine Learning Models

Raghavendra Srinivasaiah[1] (✉), Santosh Kumar Jankatti[2], Niranjana Shravanabelagola Jinachandra[1], Manjunath Ramanna Lamani[3], Ravikumar Hodikehosahally Channegowda[4]

[1]CHRIST University, Bangalore, India

[2]Dayananda Sagar University, Bangalore, India

[3]Moodlakatte Institute of Technology, Kundapura, India

[4]Dayananda Sagar Academy of Technology and Management, Bengaluru, India

s.raghavendra@christuniversity.in

ABSTRACT

Chronic kidney disease (CKD) is a progressive illness that often remains undiagnosed until advanced stages and represents a significant global health burden. Proper and timely diagnosis of CKD can significantly improve patient prognosis and reduce treatment costs. This study evaluates several machine learning (ML) models, including support vector machine (SVM), random forest (RF), gradient boosting (GB), Naïve Bayes (NB), AdaBoost, and a multilayer perceptron (MLP) neural network. Additionally, it proposes a stacking ensemble model combining RF and GB for accurate CKD prediction using a publicly available Kaggle dataset. Missing value handling and feature normalisation are performed during data preprocessing, and model performance is evaluated using an 80:20 train–test split with metrics such as the area under the curve (AUC), classification accuracy (CA), F1-score, precision, recall, and Matthews Correlation Coefficient (MCC). Experimental results indicate that RF and GB achieve the strongest individual performance, while the proposed stacking ensemble attains the highest CA of 99.4%. These findings highlight the potential of artificial intelligence (AI)-driven predictive models to support proactive CKD diagnosis and enhance clinical decision-making in healthcare systems.

KEYWORDS

chronic kidney disease (CKD), machine learning (ML), ensemble learning, prediction, gradient boosting (GB)

1 INTRODUCTION

About 10% of the global population has chronic kidney disease (CKD), which is a global health issue. The complex issue of CKD is increasing slowly, especially in developing nations. Public awareness of such a high-priority issue is recent and minor [1].

Early detection of CKD progression within a given period is crucial for effective treatment, as it improves quality of life and prolongs survival duration.

Srinivasaiah, R., Jankatti, S. K., Jinachandra, N. S., Lamani, M. R., Channegowda, R. H. (2026). Early CKD Prediction Using Ensemble and Basic Machine Learning Models. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(5), pp. 171–184. https://doi.org/10.3991/ijoe.v22i05.58661

Article submitted 2025-09-14. Revision uploaded 2026-02-13. Final acceptance 2026-02-13.

Early detection prevents the cost of the final stage [2]. Demographic ageing and a higher prevalence of long-term diseases like hypertension, diabetes, and obesity will lead to a growth in the prevalence of CKD in the coming years. Due to the variability of victims of CKD and the impact of the elusive concerns, predicting when CKD will cause renal failure is difficult. Furthermore, false predictions and unjustified interventions can create a gap in intervention, leading to kidney failure [3] [4].

CKD stages 1 to 3 are asymptomatic, and most individuals with CKD are unaware of their status and remain undiagnosed. Symptoms will be experienced later after the onset of complications, such as kidney failure, and the development of other comorbidities becomes linked with the disease. CKD is a disease of multimorbidity, and the most prevalent among these are cardiovascular disease, hypertension, and type 2 diabetes. For high-risk CKD patients, measurement of albumin-creatinine ratio in urine, blood pressure, and glomerular filtration rate should be noted and monitored at regular intervals [5] [6].

Machine learning (ML) is a subfield of artificial intelligence (AI) that is capable of teaching computers to act like human beings by learning from and improving them. ML is less dependent on human intervention and has numerous applications in various fields, including societal, food, transport, education, industry, and health, among others [7]. ML developments have equipped researchers with the tools necessary to carry out the tasks involved in data generation, noise removal in data, feature extraction, classification, and dimension reduction [8]–[16]. ML algorithms allow us to detect the intangible dependencies of data. ML can be utilised to train the algorithm for it to identify the input-output feature relationship. Additionally, ML helps us understand the interdependency among data features. [17]. Applying ML methods to healthcare data will assist in predicting whether a patient has long-term diseases like cancer [18] [19], kidney failure [20] [21], heart disease [22] [23], stroke [24] [25], diabetes [26] [27], and hypertension [28] [29].

Health information predictions are based on the datasets, lab test results, specialist advice with physicians, and outcomes from general medical training. Currently, disease identification tasks from medical data have numerous essential challenges, including missing values [30], attribute contributions [31], and imbalanced data [32].

K-nearest neighbour (KNN) is a straightforward method that relies on its nearest neighbours and is easy to implement. The KNN algorithm is used most broadly among the vast reservoir of ML algorithms. KNN is a supervised ML algorithm that can classify the data according to the attributes and labels of the training data. The KNN algorithm classifies data based on the model trained using the K nearest training data points. Lastly, the algorithm applies a majority voting rule to determine the final classification [33].

Random forest (RF) is a robust disease prediction model and ensemble learning algorithm that constructs numerous decision trees (DT). RF is expected to be very accurate, is immune to overfitting, and can effectively cope with many input attributes. RF's performance relies on data quality, attribute choice, and hyperparameter optimisation [34].

One of the most straightforward of the earlier ML algorithms is the DT, a flowchart-like model that reveals the paths of decision as a function of input features. DT structure is an imperative tree structure that provides a hierarchy of dominant features revealing transparency. DT categorises the original data into subsets until the training data is categorised perfectly and searches for the internal, leaf, and root nodes using a recursive algorithm. DTs apply probability when deciding whether the event occurs or not in categorising the target variables [35].

Gradient boosting (GB) is an effective and commonly used technique, as it improves the model's accuracy. The GB technique develops a robust ensemble technique by combining the output of various weak models. GB can be applied to most ML tasks, can be used for both regression and classification, and can handle complex data patterns. GB mainly handles misclassification, which assigns them greater weightage and is less sensitive to noisy training data [36] [37].

NB is a straightforward yet very popular and well-adapted technique for ML classification. The NB classifier is probabilistic and supports classification by summing up the most likely value of training data [38]. NB works best if there is a controllable level of dependency between the features. NB transforms the test data into the correct output set [39].

One of the most potent applications of neural networks in deep learning (DL) is the convolutional neural network (CNN). CNN has been incredibly successful in numerous areas, including healthcare, computer vision, natural language processing, face detection, and autonomous vehicles, among others [40].

The Logistic Regression (LR) model is employed to determine the probability of the disease. LR, or the logit model, operates by setting the relationship between one or more independent variables and a binary dependent variable. The model generated translates the output onto the predictor variable. The probability of production is estimated by translating the input data into the logistic curve and expressed as regression coefficients [41].

1.1 Contributions and novelty of this study

1. Implements a comparative evaluation of CKD prediction models, including support vector machine (SVM), RF, GB, NB, AdaBoost, and multilayer perceptron (MLP), and proposes a stacking ensemble combining RF and GB.
2. Evaluates model performance using comprehensive metrics such as area under the curve (AUC), classification accuracy (CA), F1-score, precision, recall, and Matthews Correlation Coefficient (MCC) to ensure reliable and unbiased assessment.
3. Experimental results demonstrate that RF and the proposed stacking ensemble achieve the most balanced and robust performance across evaluation metrics, highlighting their suitability for CKD prediction.
4. Presents a data-driven AI-based diagnostic framework that supports early detection and proactive management of CKD.

2 LITERATURE SURVEY

To enhance the life expectancy of CKD patients, there is a trial involving the application of ML and DL algorithms in early disease diagnosis. The research utilised the readily available clinical characteristics of the CKD patients. Eight ML algorithms were used to predict the characteristics to predict the probability of CKD patients reaching the end stage within a duration of three years. The experimental assessment confirmed that the DL model AUC-ROC was 89.9% [42].

A new hybrid DL model, Deep Separable CNN, was proposed for the prediction and early detection of CKD. The Capsule Network is utilised to select the appropriate features for kidney problems. Chosen features are specified using the Aquila Optimisation Algorithm to enhance the processing rate. The model was validated

on the CKD dataset selected from the UCI ML database, achieving 95.08% accuracy without using a feature selection method [43].

The advent of AI has addressed problems in the medical field, including healthcare, disease control, and disease diagnosis. Better observation of the kidney and diabetes is suggested with the adoption of AI techniques. Six ML and four DL models are formulated to serve the purpose. The experimental results revealed that the RF technique achieved an accuracy level of 97.5%, while the LSTM-CNN technique achieved an accuracy level of 98.9% [44].

An exhaustive belief network with Softmax activation and cross-entropy loss function was utilised for CKD prediction in the first stage. The model demonstrated a performance improvement of 98.5% compared to other methods [45].

CKD is a disease of the kidney that can cause numerous health issues in the late stage if it is not diagnosed in childhood. Ultrasonography, being a side-effect-free technique, can be utilised for the examination of the kidney. A new work utilising CNN from an ultrasound image is proposed, which can automatically identify the deep and texture features for classification. Transfer learning is employed to train the model. According to experimental verification, the proposed algorithm achieves a precision of 96.01% [46].

CKD is among the top 20 causes of death. ML algorithms are implemented using data from 400 patients with 24 features to classify CKD in its initial stage. The missing values are replaced with the mean approach. The recursive feature elimination method is utilised to identify the most relevant features. Four ML classifiers, SVM, KNN, DT, and RF, are used for classification. The experimental results showed that RF achieved an accuracy of 90% [47].

Detection of CKD progression in all stages is highly beneficial for enhancing the patient's health. An RF-based model was proposed to predict improvement based on population analysis and clinical evidence. Approximately 77,196 participants with over 80 features were used for analysis. The experimental outcome demonstrated that the end model led to an AUC-ROC of 88% at year and 84% at 5 years. Additionally, there was a 40% reduction in kidney failure [48].

A model derived from an ANN for the forecasting of the terminal phase of CKD is constructed based on IgAN. For the research, 948 patients with IgAN were included in the data analysis. The model is divided into two stages: the first stage forecasts the CKD, and the second stage utilises a regression model to predict the CKD. Experimental evaluation indicated that the model was 82% accurate [49].

An attempt was made to categorise CKD as severe, mild, or moderate by comparing the performance of various methods, including NN, KNN, SVM, LR, lasso and ridge regression, Elastic Net, XGBoost, and RF methods. The study included 551 patients with 18 attributes. The experimental outcome indicated that the LR method performed better than the other method with 87.3% accuracy [50].

Data analysis can help predict CKD before it occurs and also plays a vital role in health conditions. An emerging study focusing on the incidence of CKD within 6–12 months is proposed by developing an ensemble method combining DL techniques like LSTM, LSTM-BLSTM, and CNN with a special emphasis on feature selection, optimisation, and data imbalance. The combination strategy had 98% accuracy at 6 months and 97% accuracy at 12 months and was closest to the individual methods [51].

From the literature review, we can see that DL and ML methods are used to forecast CKD outcomes, as evident from Table 1.

Table 1. Summarisations of the existing different ML and DL models result

Reference No.	Model Used	Result Achieved
[42]	• DL Model	• AUC-ROC of 89.9%
[43]	• Hybrid DL model (Deep Separable CNN with Capsule Network and Aquila Optimisation Algorithm)	• Accuracy of 95.08% (without feature selection)
[44]	• ML (RF), DL (LSTM-CNN)	• RF Accuracy: 97.5%, LSTM-CNN Accuracy: 98.9%
[45]	• Deep Belief Network (DBN) with Softmax and Cross-Entropy Loss	• Accuracy: 98.5%
[46]	• CNN (Ultrasound Image-Based) with Transfer Learning	• Accuracy: 96.01%
[47]	• ML (SVM, KNN, DT, RF)	• RF Accuracy: 90%
[48]	• RF (RF) for Population Analysis	• AUC-ROC: 88% (2 years), 84% (5 years); 40% decline in kidney failure
[49]	• Artificial Neural Network (ANN)	• Accuracy: 82%
[50]	• ML (NN, KNN, SVM, LR, Lasso & Ridge Regression, Elastic Net, XGBoost, RF)	• LR Accuracy: 87.3%
[51]	• Ensemble DL (LSTM, LSTM-BLSTM, CNN)	• Accuracy: 98% (6 months), 97% (12 months)

3 METHODOLOGY

3.1 Dataset

400 examples and 25 attributes, with 4.7% missing data, are processed using preprocessing methods, including imputation by the average value and the majority attribute value. The Kaggle CKD Dataset, with various features, is presented in Table 2.

Table 2. Summary of the dataset

Attribute Name	Type	Description
age	Numeric	• Age of the patient in years
bp	Numeric	• Blood pressure in mmHg
sg	Numeric	• Urine specific gravity (typically ranges from 1.005 to 1.030)
al	Numeric	• Albumin levels in urine (0–5 scale, higher indicates severity)
su	Numeric	• Sugar levels in urine (0–5 scale, higher may indicate diabetes)
rbc	Categorical	• Red blood cell counts in urine: normal or abnormal
pc	Categorical	• Pus cell counts in urine: normal or abnormal
pcc	Categorical	• Presence of pus cell clumps in urine (Yes/No)
ba	Categorical	• Presence of bacteria in urine sample (Yes/No)
bgr	Numeric	• Random blood glucose level measured in mg/dL

(Continued)

Table 2. Summary of the dataset *(Continued)*

Attribute Name	Type	Description
bu	Numeric	• Blood urea level in mg/dL, high values indicate renal dysfunction
sc	Numeric	• Serum creatinine level in mg/dL, a key kidney function marker
sod	Numeric	• Sodium concentration in the blood (mEq/L)
pot	Numeric	• Potassium concentration in the blood (mEq/L)
hemo	Numeric	• Haemoglobin level in g/dL, low levels suggest anaemia
pcv	Numeric	• Volume of red blood cells as a percentage of total blood
wc	Numeric	• Count of white blood cells (cells per cubic mm)
rc	Numeric	• Count of red blood cells (millions per cubic mm)
htn	Categorical	• Indicates presence of hypertension (Yes/No)
dm	Categorical	• Indicates presence of diabetes mellitus (Yes/No)
cad	Categorical	• Indicates history of coronary artery disease (Yes/No)
appet	Categorical	• Patient's appetite: good or poor
pe	Categorical	• Swelling in feet or ankles (Yes/No)
ane	Categorical	• Indicates if the patient has anaemia (Yes/No)
classification	Categorical	• Target variable indicating CKD/Not CKD

3.2 Methods

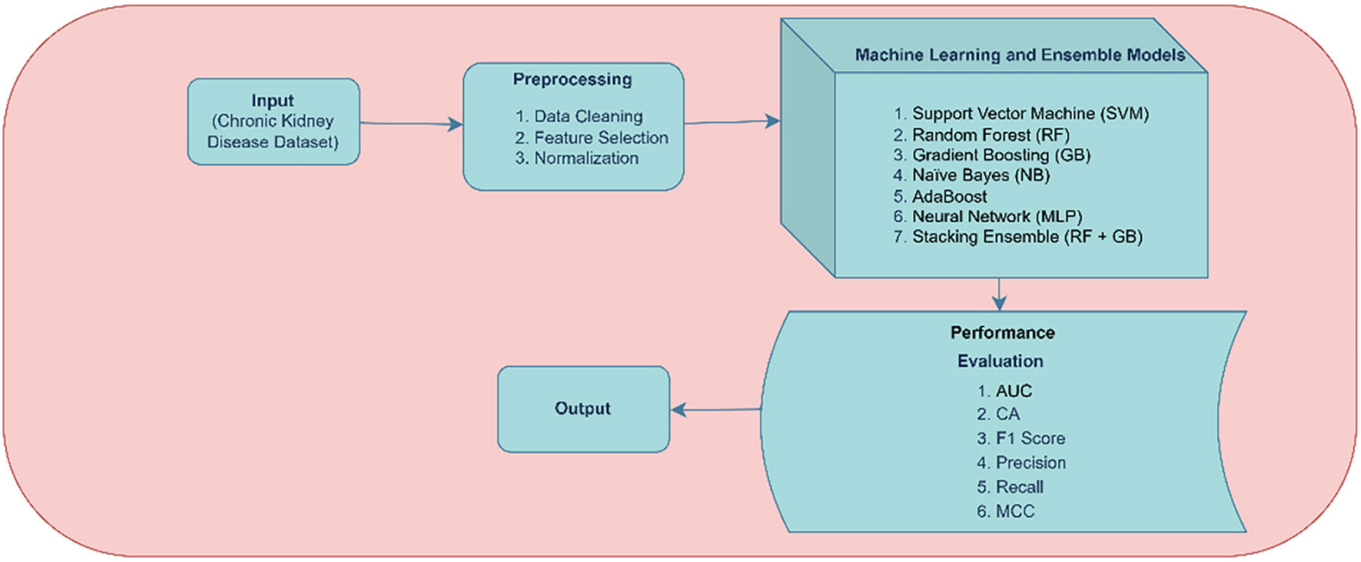


Fig. 1. Block diagram for CKD prediction

The process begins with the CKD dataset, which is cleaned using data cleaning, feature selection, and normalisation to obtain high-quality input for training models, as shown in Figure 1. The pre-processed data is then used to train and evaluate multiple classifiers, including SVM, RF, GB, NB, AdaBoost, and MLP. In addition, a stacking ensemble combining RF and GB is evaluated. These models are evaluated

based on performance metrics such as AUC, CA, F1 Score, Precision, Recall, and MCC to assess their performance. Lastly, the best-performing model is chosen based on these metrics for CKD prediction, demonstrating accuracy and potential application in real-life scenarios.

Algorithm: CKD Prediction Using ML

Step 1: Input Data Collection
- Loaded the CKD dataset from Kaggle.

Step 2: Data Preprocessing
1. Data Cleaning
   - Handled missing values using imputation techniques (mean, median, mode).
   - Removed the duplicate or inconsistent records.
2. Feature Selection
   - Identified relevant features using correlation analysis and statistical tests.
   - Removed irrelevant or redundant features to improve model performance.
3. Data Normalisation
   - Normalised numerical features using min-max scaling or standardisation.
   - Converted categorical variables into numerical form using one-hot encoding or label encoding.

Step 3: Model Implementation
- Trained multiple ML models on the processed dataset:
  1. SVM
  2. RF
  3. GB
  4. NB
  5. AdaBoost
  6. MLP
  7. Stacking Ensemble (RF + GB)

Step 4: Model Training & Validation
- The dataset was split into training and testing sets using an 80:20 stratified split.
- Models were trained on the training set and evaluated on the held-out test set using AUC, CA, F1-score, Precision, Recall, and MCC.

Step 5: Performance Evaluation
- Evaluated models using key classification metrics:
  1. AUC – Measures classification ability.
  2. CA – Measures correct predictions.
  3. F1 Score – Balances precision and recall.
  4. Precision – Measures positive predictive value.
  5. Recall – Measures sensitivity (true positive rate).
  6. MCC – Measures overall classification quality.

Step 6: Best Model Selected and Predicted
1. Compared all models based on AUC, CA, F1 Score, Precision, Recall, and MCC.
2. Identified the best-performing model with the highest metrics.
3. Utilise the selected model to predict CKD risk for new patients.

3.3 Stacking model: RF combined with GB

Eq. (1) represents the Hybrid Stacking Model, which mathematically defines the ensemble mechanism used in our stacking model, combining the predictions of *RF* and *GB*. By tuning weights α and β, we achieved an optimal CA of 99.4%, as shown in Table 4.

$$Stacked_{Prediction} = \alpha \cdot RF(x) + \beta \cdot GB(x) \quad (1)$$

Where $\alpha + \beta = 1$

4 RESULT AND DISCUSSION

Hyperparameters are fixed configuration parameters that control the training of an ML algorithm, such as the number of trees in an RF or the learning rate in GB. They are set before training and play a fundamental role in specifying the model's performance. They need to be tuned for accuracy and generalisability, as shown in Table 3.

Table 3. Different techniques for utilisation of the hyperparameters

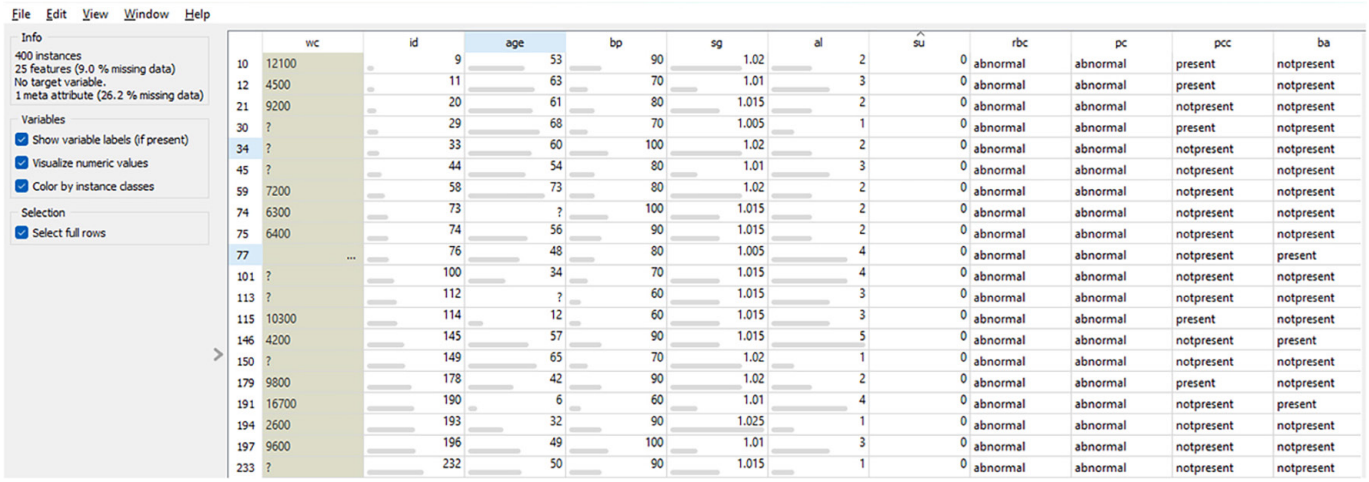
Algorithm	Hyperparameters
SVM	• Regression loss epsilon: 0.1 • Optimisation iterations: 100 • Kernel: RBF
RF	• Max depth of tree: 5 • Number of trees: 10
Neural Network	• Hidden layer neurones: 100 • Activation function: ReLU • Solver: Adam
GB	• Number of trees: 100 • Learning rate: 0.100 • Tree depth: 3
Adaboost	• Number of estimators: 50 • Learning rate: 1.00000 • Classification: SAMM.R • Regression: Linear
Stacking Model	• RF combined with GB

4.1 Before pre-processing data

Before preprocessing, the dataset consisted of 400 samples and 25 features, with missing values represented by question marks ("?"). The numerical features, such as age, blood pressure (bp), and serum glucose (sg), are raw and have different ranges and non-uniform scales. The categorical features, such as RBC, PC, and PCC, have text values like "abnormal" and "not present," which cannot be handled by the majority of ML algorithms without preprocessing. Missing data can lead to biased or incorrect model training if not appropriately handled. Data must be transformed, cleaned, and encoded before analysis or forecasting, as shown in Figure 2.

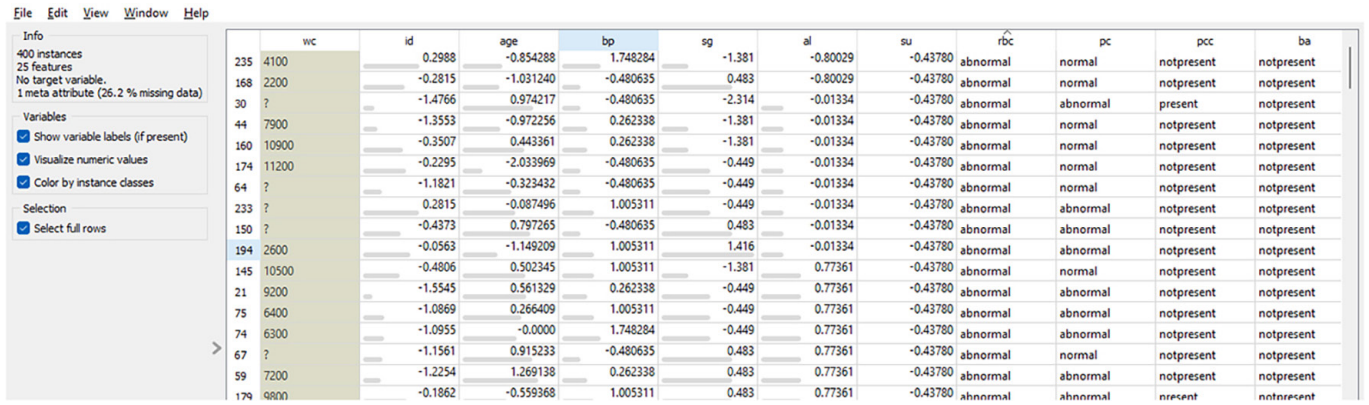
4.2 After pre-processing data

After preprocessing, the dataset comprises 400 cases and 25 clinical values and laboratory test features. Missing values accounted for approximately 26.2% and were addressed as needed for model preparation. Categorical values were one-hot encoded, while numerical features were normalised to enhance the model's performance, achieve feature homogeneity, and are represented in Figure 3. The dataset initially contained approximately 26.2% missing values at the cell level, while 4.7% of records had at least one missing attribute. All missing values were handled during pre-processing, resulting in a complete dataset for model training.



	wc	id	age	bp	sg	al	su	rbc	pc	pcc	ba
10	12100	9	53	90	1.02	2	0	abnormal	abnormal	present	notpresent
12	4500	11	63	70	1.01	3	0	abnormal	abnormal	present	notpresent
21	9200	20	61	80	1.015	2	0	abnormal	abnormal	notpresent	notpresent
30	?	29	68	70	1.005	1	0	abnormal	abnormal	present	notpresent
34	?	33	60	100	1.02	2	0	abnormal	abnormal	notpresent	notpresent
45	?	44	54	80	1.01	3	0	abnormal	abnormal	notpresent	notpresent
59	7200	58	73	80	1.02	2	0	abnormal	abnormal	notpresent	notpresent
74	6300	73	?	100	1.015	2	0	abnormal	abnormal	notpresent	notpresent
75	6400	74	56	90	1.015	2	0	abnormal	abnormal	notpresent	notpresent
77	...	76	48	80	1.005	4	0	abnormal	abnormal	notpresent	present
101	?	100	34	70	1.015	4	0	abnormal	abnormal	notpresent	notpresent
113	?	112	?	60	1.015	3	0	abnormal	abnormal	notpresent	notpresent
115	10300	114	12	60	1.015	3	0	abnormal	abnormal	present	notpresent
146	4200	145	57	90	1.015	5	0	abnormal	abnormal	notpresent	present
150	?	149	65	70	1.02	1	0	abnormal	abnormal	notpresent	notpresent
179	9800	178	42	90	1.02	2	0	abnormal	abnormal	present	notpresent
191	16700	190	6	60	1.01	4	0	abnormal	abnormal	notpresent	present
194	2600	193	32	90	1.025	1	0	abnormal	abnormal	notpresent	notpresent
197	9600	196	49	100	1.01	3	0	abnormal	abnormal	notpresent	notpresent
233	?	232	50	90	1.015	1	0	abnormal	abnormal	notpresent	notpresent

Fig. 2. Before pre-processing data with instances and features



	wc	id	age	bp	sg	al	su	rbc	pc	pcc	ba
235	4100	0.2988	-0.854288	1.748284	-1.381	-0.80029	-0.43780	abnormal	normal	notpresent	notpresent
168	2200	-0.2815	-1.031240	-0.480635	0.483	-0.80029	-0.43780	abnormal	normal	notpresent	notpresent
30	?	-1.4766	0.974217	-0.480635	-2.314	-0.01334	-0.43780	abnormal	abnormal	present	notpresent
44	7900	-1.3553	-0.972256	0.262338	-1.381	-0.01334	-0.43780	abnormal	normal	notpresent	notpresent
160	10900	-0.3507	0.443361	0.262338	-1.381	-0.01334	-0.43780	abnormal	normal	notpresent	notpresent
174	11200	-0.2295	-2.033969	-0.480635	-0.449	-0.01334	-0.43780	abnormal	normal	notpresent	notpresent
64	?	-1.1821	-0.323432	-0.480635	-0.449	-0.01334	-0.43780	abnormal	normal	notpresent	notpresent
233	?	0.2815	-0.087496	1.005311	-0.449	-0.01334	-0.43780	abnormal	abnormal	notpresent	notpresent
150	?	-0.4373	0.797265	-0.480635	0.483	-0.01334	-0.43780	abnormal	abnormal	notpresent	notpresent
194	2600	-0.0563	-1.149209	1.005311	1.416	-0.01334	-0.43780	abnormal	abnormal	notpresent	notpresent
145	10500	-0.4806	0.502345	1.005311	-1.381	0.77361	-0.43780	abnormal	normal	notpresent	notpresent
21	9200	-1.5545	0.561329	0.262338	-0.449	0.77361	-0.43780	abnormal	abnormal	notpresent	notpresent
75	6400	-1.0869	0.266409	1.005311	-0.449	0.77361	-0.43780	abnormal	abnormal	notpresent	notpresent
74	6300	-1.0955	-0.0000	1.748284	-0.449	0.77361	-0.43780	abnormal	abnormal	notpresent	notpresent
67	?	-1.1561	0.915233	-0.480635	0.483	0.77361	-0.43780	abnormal	normal	notpresent	notpresent
59	7200	-1.2254	1.269138	0.262338	0.483	0.77361	-0.43780	abnormal	abnormal	notpresent	notpresent
179	9800	-0.1862	-0.559368	1.005311	0.483	0.77361	-0.43780	abnormal	abnormal	present	notpresent

Fig. 3. After preprocessing data with instances and features

4.3 Data quality score (DQS)

DQS is given in Eq. (2) and quantifies the effectiveness of preprocessing by reflecting the percentage of missing data handled. Our processed dataset achieved a DQS of 100%, ensuring high-quality input for model training.

$$DQS = \left(1 - \frac{Missing_{after}}{Total}\right) \cdot 100 \tag{2}$$

The performance outcomes of the models after training using an 80:20 data sampling method indicate that RF produced the highest CA (CA = 0.994) and (AUC = 0.987) values, verifying its superiority in classifying CKD. GB and SVM also produced high-performance outcomes, with CA values of 0.991 and AUC values of 0.979 and 0.983, respectively. AdaBoost and Neural Networks provided competitive performance, both with high precision and recall scores. NB considerably underperformed (CA = 0.330, AUC = 0.984), consistent with the expectation based on its inappropriateness for this dataset. The stacking model, using ensemble learning, tied RF in accuracy (CA = 0.994), again illustrating the value added by combining models (refer to Table 4). Although NB achieved a relatively high AUC, its low CA indicates poor probability calibration and threshold sensitivity, making it unsuitable for reliable CKD prediction on this dataset. RF and stacking were the best overall strategies,

with GB and SVM offering plausible alternatives to predictive modelling for CKD classification, as shown in Figure 4.

Table 4. Results – performance metrics

Model	AUC	CA	F1	Precision	Recall	MCC
SVM	0.983	0.991	0.988	0.985	0.991	0.98
RF	0.987	0.994	0.991	0.988	0.994	0.987
Naive Bayes	0.984	0.33	0.343	0.364	0.33	0.458
Neural Network	0.987	0.988	0.986	0.985	0.988	0.974
GB	0.979	0.991	0.988	0.985	0.991	0.98
AdaBoost	0.984	0.988	0.986	0.985	0.988	0.974
Stack	0.984	0.994	0.991	0.988	0.994	0.987

4.4 Composite evaluation score (CES)

The CES, computed using Eq. (3), provides a unified metric combining six evaluation scores for fair comparison across models. It highlights that RF and stacking models maintain balanced superiority across all evaluation parameters.

$$CES = \frac{AUC + CA + F1 + Precision + Recall + MCC}{6} \tag{3}$$

5 CONCLUSIONS

The results of this study demonstrate the potential of ML and ensemble-based approaches for accurate prediction of CKD using clinical data. Careful data preprocessing and consistent evaluation ensured the model's reliable performance. Among the evaluated models, RF and the proposed stacking ensemble achieved the most robust and balanced predictive outcomes across multiple performance metrics. These findings highlight the effectiveness of ensemble learning for CKD screening and its applicability as an AI-assisted decision-support tool in healthcare. Future work may focus on validation with larger datasets and real-world clinical integration.

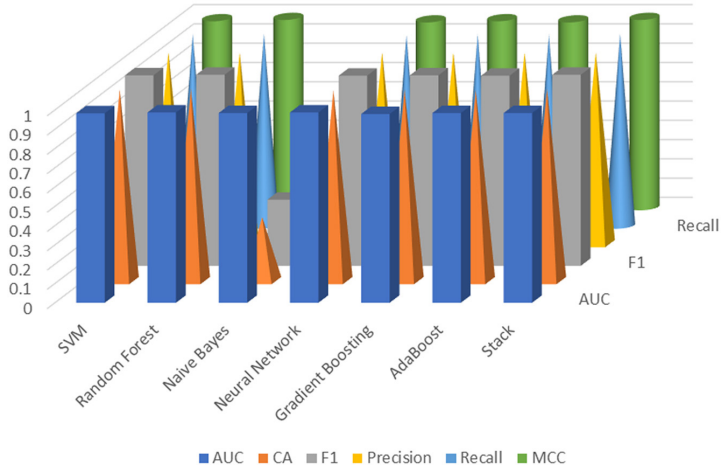


Fig. 4. Model evaluation results

6 REFERENCES

[1] M. Jadoul, M. Aoun, and M. M. Imani, "The major global burden of chronic kidney disease," *The Lancet Global Health*, vol. 12, no. 3, pp. e342–e343, 2024. https://doi.org/10.1016/S2214-109X(24)00050-0

[2] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. W. Quinn, and M. A. Moni, "Clinically applicable ML approaches to identify attributes of CKD (CKD) for use in low-cost diagnostic screening," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1–11, 2021. https://doi.org/10.1109/JTEHM.2021.3073629

[3] M. C. Thomas, M. E. Cooper, and P. Zimmet, "Changing epidemiology of type 2 diabetes mellitus and associated CKD," *Nature Rev. Nephrol.*, vol. 12, no. 2, pp. 73–81, 2016. https://doi.org/10.1038/nrneph.2015.173

[4] M. W. Taal and B. M. Brenner, "Predicting initiation and progression of CKD: Developing renal risk scores," *Kidney International*, vol. 70, no. 10, pp. 1694–1705, 2006. https://doi.org/10.1038/sj.ki.5001794

[5] M. Evans *et al.*, "A narrative review of CKD in clinical practice: Current challenges and future perspectives," *Advances in Therapy*, vol. 39, pp. 33–43, 2022. https://doi.org/10.1007/s12325-021-01927-z

[6] P. Kushner, K. Khunti, A. Cebrián, and G. Deed, "Early identification and management of CKD: A narrative review of the crucial role of primary care practitioners," *Advances in Therapy*, vol. 41, pp. 3757–3770, 2024. https://doi.org/10.1007/s12325-024-02957-z

[7] S. V. Mahadevkar, B. Khemani, S. Patil, K. Kotecha, D. R. Vora, and A. Abraham, "A review on ML styles in computer vision—Techniques and future directions," *IEEE Access*, vol. 10, pp. 107293–107329, 2022. https://doi.org/10.1109/ACCESS.2022.3209825

[8] M. Zhao, Y. Cong, and L. Carin, "On leveraging pretrained GANs for generation with limited data," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 11340–11351.

[9] H. Ali, C. Grönlund, and Z. Shah, "Leveraging GANs for data scarcity of COVID-19: Beyond the hype," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, vol. 12892, 2023, pp. 659–667. https://doi.org/10.1109/CVPRW59228.2023.00073

[10] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "DL on image denoising: An overview," *Neural Netw.*, vol. 131, pp. 251–275, 2020. https://doi.org/10.1016/j.neunet.2020.07.025

[11] M. Ribeiro *et al.*, "Denoising and decoding spontaneous vagus nerve recordings with ML," in *Proc. 45th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2023, pp. 1–4. https://doi.org/10.1109/EMBC40787.2023.10340443

[12] S. Dara and P. Tumma, "Feature extraction by using DL: A survey," in *Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2018, pp. 1795–1801. https://doi.org/10.1109/ICECA.2018.8474912

[13] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, 2016. https://doi.org/10.1016/j.neucom.2015.08.104

[14] T. Jawad, R. G. L. Koh, and J. Zariffa, "Selective peripheral nerve recording using simulated human median nerve activity and CNNs," *Biomed. Eng. OnLine*, vol. 22, no. 1, pp. 1–19, 2023. https://doi.org/10.1186/s12938-023-01181-0

[15] A. Zermane, M. Z. M. Tohir, H. Zermane, M. R. Baharudin, and H. M. Yusoff, "Predicting fatal fall from heights accidents using RF classification ML model," *Saf. Sci.*, vol. 159, p. 106023, 2023. https://doi.org/10.1016/j.ssci.2022.106023

[16] J. R. de Miras, A. J. Ibáñez-Molina, M. F. Soriano, and S. Iglesias-Parro, "Schizophrenia classification using ML on resting state EEG signal," *Biomed. Signal Process. Control*, vol. 79, p. 104233, 2023. https://doi.org/10.1016/j.bspc.2022.104233

[17] A. Siemuri, K. Selvan, H. Kuusniemi, P. Valisuo, and M. S. Elmusrati, "A systematic review of ML techniques for GNSS use cases," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 6, pp. 5043–5077, 2022. https://doi.org/10.1109/TAES.2022.3219366

[18] R. R. Kadhim and M. Y. Kamil, "Comparison of ML models for breast cancer diagnosis," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 12, no. 1, pp. 415–421, 2023. https://doi.org/10.11591/ijai.v12.i1.pp415-421

[19] G. Kumawat, S. K. Vishwakarma, P. Chakrabarti, P. Chittora, T. Chakrabarti, and J. C.-W. Lin, "Prognosis of cervical cancer disease by applying ML techniques," *J. Circuits, Syst. Comput.*, vol. 32, no. 1, 2023. https://doi.org/10.1142/S0218126623500196

[20] S. Revathy, "CKD prediction using ML models," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 6364–6367, 2019. https://doi.org/10.35940/ijeat.A2213.109119

[21] K. R. A. Padmanaban and G. Parthiban, "Applying ML techniques for predicting the risk of CKD," *Indian J. Sci. Technol.*, vol. 9, no. 29, pp. 1–6, 2016.

[22] V. Stepanyan, "Comparative analysis of ML methods for prediction of heart disease," *J. Mach. Manuf. Reliab.*, vol. 51, no. 8, pp. 789–799, 2022. https://doi.org/10.3103/S1052618822080210

[23] M. Alfadli and A. O. Almagrabi, "Feature-limited prediction on the UCI heart disease dataset," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 5871–5883, 2023. https://doi.org/10.32604/cmc.2023.033603

[24] R. Huang *et al.*, "Stroke mortality prediction based on ensemble learning and the combination of structured and textual data," *Comput. Biol. Med.*, vol. 155, p. 106176, 2023. https://doi.org/10.1016/j.compbiomed.2022.106176

[25] P. B. Dash, "Efficient ensemble learning based CatBoost approach for early-stage stroke risk prediction," in *Ambient Intelligence in Health Care: Proceedings of ICAIHC 2022*, Singapore: Springer, 2022, pp. 475–483. https://doi.org/10.1007/978-981-19-6068-0_46

[26] B. S. Ahamed, M. S. Arya, and A. O. V. Nancy, "Diabetes mellitus disease prediction using ML classifiers with oversampling and feature augmentation," *Adv. Hum.-Comput. Interact.*, vol. 2022, pp. 1–14, 2022. https://doi.org/10.1155/2022/9220560

[27] P. Theerthagiri, A. U. Ruby, and J. Vidya, "Diagnosis and classification of the diabetes using ML algorithms," *Social Netw. Comput. Sci.*, vol. 4, no. 1, p. 72, 2022. https://doi.org/10.1007/s42979-022-01485-3

[28] W. Chang *et al.*, "A machine-learning-based prediction method for hypertension outcomes based on medical data," *Diagnostics*, vol. 9, no. 4, p. 178, 2019. https://doi.org/10.3390/diagnostics9040178

[29] M. A. J. Tengnah, R. Sooklall, and S. D. Nagowah, "A predictive model for hypertension diagnosis using ML techniques," in *Telemedicine Technologies,* Mauritius: Academic, 2019, pp. 139–152. https://doi.org/10.1016/B978-0-12-816948-3.00009-X

[30] X. Lu, L. Yuan, R. Li, Z. Xing, N. Yao, and Y. Yu, "An improved Bi-LSTM-based missing value imputation approach for pregnancy examination data," *Algorithms*, vol. 16, no. 1, p. 12, 2022. https://doi.org/10.3390/a16010012

[31] P. Muthulakshmi and M. Parveen, "Z-score normalized feature selection and iterative African buffalo optimization for effective heart disease prediction," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 1, pp. 25–37, 2022. https://doi.org/10.22266/ijies2023.0228.03

[32] F. Yang, K. Wang, L. Sun, M. Zhai, J. Song, and H. Wang, "A hybrid sampling algorithm combining synthetic minority oversampling technique and edited nearest neighbor for missed abortion diagnosis," *BMC Med. Informat. Decis. Making*, vol. 22, no. 1, p. 344, 2022. https://doi.org/10.1186/s12911-022-02075-2

[33] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Scientific Reports*, vol. 12, p. 6256, 2022. https://doi.org/10.1038/s41598-022-10358-x

[34] C. B. Obinnaya, C. O. Omankwu, E. C. Osodoeke, and V. I. Ubah, "Disease prediction using RF ML algorithm," *NIPES – Journal of Science and Technology Research*, vol. 6, no. 1, pp. 164–173, 2024. https://doi.org/10.5281/zenodo.11005308

[35] A. Balakrishnan, J. Medikonda, K. Pramod, and M. Natarajan, "Information set-based DT for Parkinson's disease severity assessment using multidimensional gait dataset," *IEEE Access*, vol. 12, pp. 129187–129201, 2024. https://doi.org/10.1109/ACCESS.2024.3456438

[36] I. M. Hassoon, "Boosting learning algorithms for chronic diseases prediction: A review," *Iraqi Journal for Computers and Informatics*, vol. 50, no. 2, pp. 22–30, 2024. https://doi.org/10.25195/ijci.v50i2.506

[37] A. J. Sánchez-García, C. López-Martín, and A. Abran, "GB optimized through differential evolution for predicting the testing effort of software projects," *IEEE Access*, vol. 11, pp. 135235–135254, 2023. https://doi.org/10.1109/ACCESS.2023.3337809

[38] A. Abraham, B. K. Gupta, A. S. Maurya, S. B. Verma, M. Husain, and A. Ali, "NB approach for word sense disambiguation system with a focus on Parts-of-Speech ambiguity resolution," *IEEE Access*, vol. 12, pp. 126668–126678, 2024. https://doi.org/10.1109/ACCESS.2024.3453912

[39] Z. Ahmed, B. Issac, and S. Das, "Ok-NB: An enhanced OPTICS and k-Naive bayes classifier for imbalance classification with overlapping," *IEEE Access*, vol. 12, pp. 57458–57477, 2024. https://doi.org/10.1109/ACCESS.2024.3391749

[40] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of CNNs: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2022. https://doi.org/10.1109/TNNLS.2021.3084827

[41] D. M. A. S. Elkahwagy, C. J. Kiriacos, and M. Mansour, "LR and other statistical tools in diagnostic biomarker studies," *Clinical and Translational Oncology*, vol. 26, pp. 2172–2180, 2024. https://doi.org/10.1007/s12094-024-03413-8

[42] P. Liang, J. Yang, W. Wang, G. Yuan, M. Han, and Q. Zhang, "DL identifies intelligible predictors of poor prognosis in CKD," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3677–3685, 2023. https://doi.org/10.1109/JBHI.2023.3266587

[43] K. Venkatrao and S. Kareemulla, "HDLNET: A hybrid DL network model with intelligent IOT for detection and classification of CKD," *IEEE Access*, vol. 11, pp. 99638–99652, 2023. https://doi.org/10.1109/ACCESS.2023.3312183

[44] M. Aljaafari, S. E. El-Deep, A. A. Abohany, and S. E. Sorour, "Integrating innovation in healthcare: The evolution of "CURA's" AI-Driven virtual wards for enhanced diabetes and kidney disease monitoring," *IEEE Access*, vol. 12, pp. 126389–126414, 2024. https://doi.org/10.1109/ACCESS.2024.3451369

[45] S. M. M. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Early prediction of CKD using deep belief network," *IEEE Access*, vol. 9, pp. 135542–135549, 2021. https://doi.org/10.1109/ACCESS.2021.3114306

[46] P.-Y. Hao *et al.*, "Texture branch network for CKD screening based on ultrasound images," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, pp. 1161–1170, 2020. https://doi.org/10.1631/FITEE.1900210

[47] E. M. Senan *et al.*, "Diagnosis of CKD using effective classification algorithms and recursive feature elimination techniques," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–10, 2021. https://doi.org/10.1155/2021/1004767

[48] T. Ferguson *et al.*, "Development and external validation of a ML model for progression of CKD," *Kidney International Reports*, vol. 7, no. 8, pp. 1772–1781, 2022. https://doi.org/10.1016/j.ekir.2022.05.004

[49] F. P. Schena *et al.*, "Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy," *Kidney International*, vol. 99, no. 5, pp. 1179–1188, 2021. https://doi.org/10.1016/j.kint.2020.07.046

[50] J. Xiao *et al.*, "Comparison and development of ML tools in the prediction of CKD progression," *Journal of Translational Medicine*, vol. 17, p. 119, 2019. https://doi.org/10.1186/s12967-019-1860-0

[51] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Early prediction of CKD based on ensemble of DL models and optimizers," *Journal of Electrical Systems and Information Technology*, vol. 11, p. 17, 2024. https://doi.org/10.1186/s43067-024-00142-4

7 AUTHORS

Raghavendra Srinivasaiah is currently working as an Associate Professor in the Department of Artificial Intelligence and Data Science Engineering at CHRIST University, Bangalore. He completed his Ph.D. degree in Computer Science and Engineering from VTU, Belgaum, India, in 2017 and has more than 21+ years of teaching experience. His interests include data mining, artificial intelligence, and big data (E-mails: s.raghavendra@christuniversity.in, raghav.trg@gmail.com).

Santosh Kumar Jankatti is currently working as an Associate Professor in the Department of Computer Science and Technology at Dayananda Sagar University, Bangalore South. He completed his Ph.D. degree in Computer Science and Engineering from VTU, Belgaum, India in 2022 and has more than 14 years of teaching experience and 3 years of IT Industry experience. His interests include data mining, artificial intelligence, and big data (E-mail: sjankatti@gmail.com).

Niranjana Shravanabelagola Jinachandra completed his Ph.D. from VTU, Belagavi, in 2022. He has done his masters in machine design from VTU, Belagavi. His areas of interest are image processing, machine learning, and fluid dynamics. He is currently working as an Associate professor in the Department of Mechanical Engineering at CHRIST University (E-mail: sjniranjan86@gmail.com).

Manjunath Ramanna Lamani holds a Ph.D. in Computer Science and Engineering from CHRIST University. He is currently working as Associate Professor and HoD in the Department of CSE, Moodlakatte Institute of Technology, Kundapura, Udupi, Karnataka, India. His academic interests span deep learning, AI, IoT, and programming (E-mail: manjunathlamani01@gmail.com).

Ravikumar Hodikehosahally Channegowda completed his Ph.D. from VTU, Belagavi, in 2021. He has done his masters in VLSI design and embedded systems from VTU Extension Centre, PESCE, Mandya. His areas of interest are image processing, machine learning, pattern recognition, and multimedia concepts. He is currently working as an Assistant Professor at Department of Electronics and Communication Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru (E-mail: raviec40@gmail.com).