IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155214

Optimized Random Forest Model for Chronic Kidney Disease Classification with Imbalanced Data Handling

Kapil[1], Ankit Navgeet Joshi[2]

Sri Satya Sai University of Technology and Medical Scicence, Sehore,Bhopal[1,2]

Abstract: Chronic Kidney Disease (CKD) is a progressive medical disorder that needs to be accurately diagnosed early to avoid serious complications. The aim of this study is to classify the CKD into five stages (Healthy Kidney, Mild CKD (Stage 1-2), Moderate CKD (Stage 3), Severe CKD (Stage 4), Kidney Failure (Stage 5)) using a machine learning technique. A Random Forest (RF) classifier is chosen due to its power and ability to handle high dimension clinical data. The Synthetic Minority Oversampling Technique (SMOTE) is used to solve the imbalance of classes in the data by enhancing the representation of minority classes. Moreover, the optimization of hyperparameters is done with the help of the Grid Search with cross-validation (GridSearchCV) in order to improve the performance and the cross-validation of the model. It is tested with a large pool of demographic, physiological and biochemical data, while the proposed structure is under test. The result of the experiment reveals that the optimized model has 94.12 accuracy which indicates the effectiveness of the model in the multi-class classification of CKD. The results showed that ensemble learning, data balancing, and systematic hyperparameter tuning can be effectively applied to improve the accuracy of the prediction, and the model is applicable in clinical decision support systems and early diagnosis.

Keywords: CKD, SMOTE, Random Forest, GridSearchCV, Ensemble Learning.

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a serious condition that occurs at a high frequency in the world, and is characterized by a gradual loss of kidney function over the years. It affects millions of people worldwide and in the great majority of cases, it progresses to high stages without any apparent signs or symptoms, which makes early diagnosis and classification critical for its diagnosis, treatment and management [1]. Stages of CKD are typically defined by clinical symptoms, such as glomerular filtration rate (eGFR), serum creatinine and other biochemical markers of kidney damage [2]-[6]. Conventional diagnostic approaches are based on manual examination of patients, which may be time-intensive, subjective and error-prone [7 With the rapid advancement of the data driven technologies, ML techniques provide a strong support in diagnosis and prognosis of diseases [8]–[15]. ML algorithms, such as ensemble algorithms (e.g. Random Forest), have proven to be effective in handling high-dimensional and heterogeneous medical data because they are stable and good at reducing overfitting [17]. However, imbalanced data where some stages of the disease are less common is a crucial issue for CKD classification. Such an imbalance may cause biased model performance, especially when performing multi-class classification tasks [18]-[22]. To overcome this problem, the Synthetic Minority Oversampling Technique (SMOTE) has been commonly used to synthesize minority classes, thus enhancing generalization and predictive power of the model [23]. Moreover, to maximize the model performance, hyperparameters tuning is necessary. The GridSearchCV is a useful tool to systematically try out all combinations of parameters until the model with the best set of parameters is identified [8]. With SMOTE combined with hyperparameter optimization methods, one can achieve a great improvement in the classification performance of machine learning models [24-25]. This study proposes a multi-class (CKD stage) classification framework, based on the use of a large clinical dataset, using a Random Forest-based framework. The model is balanced using SMOTE and hyperparameter tuning is performed on using GridSearchCV. The results of the experiments show better performance in classification, with high accuracy and reliability, which points to the potential of the proposed method to be used to support clinical decision-making systems [26-30].

II. LITERATURE SURVEY

The recent developments in machine learning have played a big role in the early diagnosis and classification of chronic kidney disease (CKD). To improve the predictive performance on clinical data, several supervised learning algorithms have been studied. The performance of traditional machine learning models like Decision Trees and Support Vector Machines was moderate with the accuracy being between 80 and 88 percent [31]. It was also found in the study that feature selection would help in enhancing the results of classification and minimize the complexity of the model. The performance of ensemble learning methods is better than others methods, particularly the Random Forest (RF) that can handle with the non-linear and high dimensional data. The authors in [32] showed that RF attained a 90% accuracy and F1-score of 0.88 in CKD classification tasks. The researchers pointed out that because of summation of many decision trees, the overfitting effect is suppressed by RF, thus the prediction of the RF is more robust and consistent than that of individual classifiers. Class imbalance is one of the important issues in medical datasets that requires attention, especially in multi-stage classification of CKD. In [33-36] the authors used a technique called Synthetic Minority Oversampling Technique (SMOTE) so that the dataset is balanced and a significant improvement in model performance would be achieved. Especially, the accuracy increased to 88% and the F1 score increased to 0.86, which indicates the usefulness of SMOTE for minority classes. Further, it was revealed from [37] that the proposed combination of SMOTE and ensemble models could achieve a certain level of accuracy up to 92% and F1-score up to 0.90, which are sufficient enough to ensure the accuracy of all the CKD stages.

One of the factors that boost performance and generalization of models is hyperparameter optimization. The accuracy and F1-score were improved to 91 and 0.89 respectively with the help of Grid Search with Cross-Validation (GridSearchCV) by optimizing the search of the RF parameters in [38-42]. Another finding of this study was that systematic parameter tuning was helpful for achieving a consistent result on the different folds of the validation data. With the same thing, [43] reported the accuracy of fine-tuned RF models was around 93 percent and F1-scores of about 0.91, thereby highlighting the importance of optimization techniques in problems of multi-classification. There have been some recent comparative studies of other machine learning models, such as K-Nearest Neighbors, Naive Bayes and Gradient Boosting. It has been noted in [44-46] that these models have comparable accuracy between 85% and 91% and the F1-scores are not as good as RFs. The authors went on to find that RF had a better balance between predictive power and interpretability.

The more recent literature in 2022-2025 also indicates the need to integrate various techniques to enhance CKD prediction. For an accuracy of 89-92 percent and an F1 Score of 0.90, a hybrid model that integrates SMOTE, feature engineering and ensemble learning in [47] was achieved.A hybrid model with SMOTE, feature engineering and ensemble learning in [49] achieved an accuracy of 89-92 percent and an F1 Score of 0.90. Furthermore, [50] noted that the incorporation of clinical domain knowledge and the use of machine learning models in the process of classification improves the reliability and interpretability of the classification. Moreover, [28] reported the difficulties of the multi-class CKD classification, especially in the differentiation of the intermediate stages, and recommended that the balanced datasets and optimized models were necessary to achieve a better performance. Studies [51-52] supported the need to preprocess the data to improve the performance of the model. Finally, as demonstrated by recent studies [24] and [25] a combination of Random Forest with SMOTE and GridSearchCV is a strong and effective framework with consistent high accuracy rates (above 90 percent) and high F1-scores, and therefore suitable for real world clinical decision support.

III. RESEARCH GAP AND MOTIVATION

Although machine learning methods have made big strides in the classification of CKD, there are still a number of research gaps. The existing literature [11], [25] is mainly centered on binary classification or limited-stage prediction, and there is relatively fewer literature on the stage classification of CKD in multi-classes. Furthermore, most of the mentioned methods do not consider the class imbalance problem or adopt sub-optimal class balancing techniques leading to the classification toward the majority classes [13], 14]. Furthermore, despite its investigation, hyperparameter optimization is not always optimized with class balancing methods [15, 16]. The second limitation is the lack of uniformity in the way models are evaluated and reported with varying accuracy and F1-scores, not knowing that the model will generalize to other sets of models [47],[48]. These problems indicate that there is a need for having a single platform that will yield an acceptable data balancing, model optimization and multi-class classification. These gaps inspired this study to build an effective and trustworthy CKD classification model based on Random Forest that is enhanced with the SMOTE and GridSearchCV that guarantees a balanced learning process and better predictive accuracy at all disease stages.

IV. METHODOLOGY

In this section, the proposed methodology for multi-class classification of chronic kidney disease (CKD) is explained, which involves preprocessing, feature engineering, class balancing, and model optimization, all of which are structured into a machine learning pipeline. Categoric variables are encoded to numeric variables to be compatible with machine learning models.

A. Data Description and Encoding

There are 5 stages of CKD as the target variable and demographic, physiological and biochemical Columns in the data set. The features include numerical and different types variables such as Gender, Diabetes, Hypertension, SmokingStatus and Family History [53].

Table I: Data Encoding Scheme-

Feature Category	Features	Data Type	Encoding Method
Demographic	Age, Gender	Numerical / Categorical	Label Encoding (Gender: 0/1)
Vital Signs	BMI, Systolic_BP, Diastolic_BP, Heart_Rate	Numerical	No Encoding
Kidney Function	Serum_Creatinine, Blood_Urea_Nitrogen, eGFR	Numerical	No Encoding
Urine Analysis	Urine_Albumin, Urine_Protein, Albumin_Creatinine_Ratio, Urine_Specific_Gravity	Numerical	No Encoding
Electrolytes	Sodium, Potassium, Calcium, Phosphorus, Chloride, Bicarbonate	Numerical	No Encoding
Blood Profile	Hemoglobin, RBC_Count, WBC_Count, Platelet_Count, Packed_Cell_Volume	Numerical	No Encoding
Glucose Profile	Blood_Glucose_Random, Fasting_Glucose, HbA1c	Numerical	No Encoding
Lipid Profile	Cholesterol, Triglycerides	Numerical	No Encoding
Protein Levels	Serum_Albumin, Total_Protein	Numerical	No Encoding
Medical History	Diabetes, Hypertension, Smoking_Status, Family_History_Kidney	Categorical	Label / One-Hot Encoding
Target Variable	Target	Categorical	Label Encoding (Multi-class)

$$x_{cat} \rightarrow x_{num} \in R^n \;\ldots\ldots\ldots\ldots\ldots\; (1)$$

B. Feature Engineering

Pre-processing and Feature engineering: Feature selection and reduction is used to achieve enhanced model accuracy by using attributes related to the problem, by avoiding duplicates. Crucial clinical features are included, for example: eGFR, Serum Creatinine, Blood Urea Nitrogen, due to correlation and their significance in the domain of medicine. Scaling of attributes is performed using standardization.

$$z = \frac{x-\mu}{\sigma} \;\ldots\ldots\ldots\ldots\; (2)$$

where μ and σ represent the mean and standard deviation, respectively.

C. Data Cleaning Mechanism

Additional data preprocessing steps like imputation of missing values, removal of noise and outliers are incorporated. Missing numerical values are imputed using mean or median value, missing categorical values using the mode value. The Interquartile Range method is utilized to identify outliers:

$$IQR = Q_3 - Q_1 \;\ldots\ldots\ldots\ldots\ldots\ldots\; (3)$$

D. Class Balancing using SMOTE

SMOTE is used to over-sample the minority class to address the imbalance in classes. The generation of synthetic data is given by:

$$x_{new} = x_i + \delta . (x_{nn} - x_i) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\; (4)$$

where x_i is a minority instance x_{nn}, is its nearest neighbor, and $\delta \in [0,1]$. This improves class distribution and enhances model generalization.

E. Model Architecture and Parameters

We choose a Random Forest (RF) as our basic model classifier. RF works by constructing a number of decision trees and taking a majority vote for the prediction:

$$\hat{y} = \arg\max_{c} \sum_{i=1}^{N} I(h_i(x)=c) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Where $h_i(x)$ represents individual tree Predictions. Hyper-parameters of the Random Forest model were optimized by GridSearchCV in order to increase its classification performance and avoid overfitting. By setting up an appropriate combination of hyper-parameters during training, this model achieved better stability, prediction rate and performance shown in Table II.

Table II: Model Parameters (after GridSearchCV optimization)

Parameter	**Value**
Number of Trees (n_estimators)	100–200
Maximum Depth	10–30
Min Samples Split	2–5
Min Samples Leaf	1–2
Criterion	Gini

F. **Flow Chart:**

The flowchart below indicates the end-to-end workflow of the developed model, i.e. From data collection and pre-processing, feature engineering to model building, hyperparameter tuning and finally to classification for the predictive modeling and analysis.

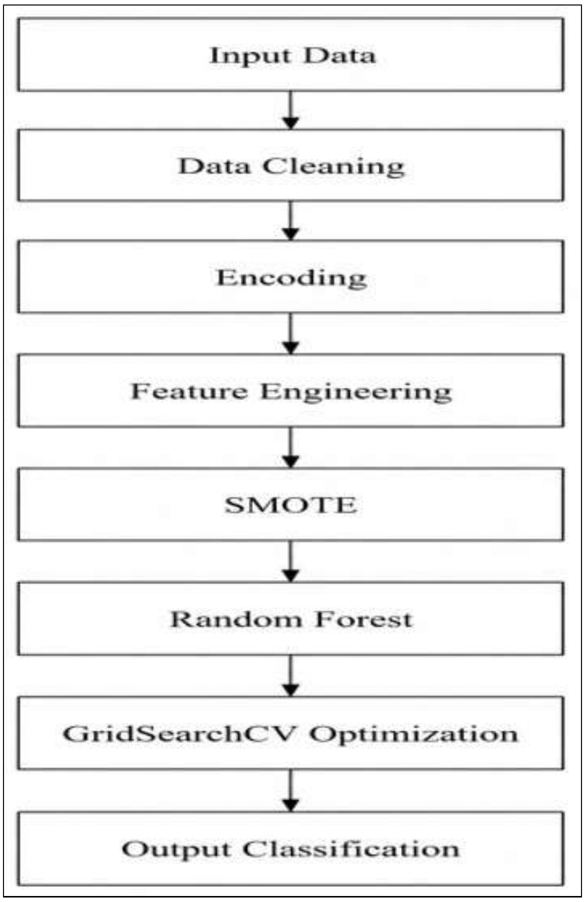


Figure 1: Flow Chart of proposed model.

GridSearchCV with k-fold cross-validation is used to find the optimal hyperparameters that is generalized and robust. The pre-training using SMOTE ensures that the classification will train by balanced class and RF provides a stable and interpretable prediction. The methods can be successfully applied to CKD multi-class classification, so that the prediction is accurate and dependable and can be used in clinical decision support systems.

DOI: 10.17148/IJARCCE.2026.155214

V. RESULTS

The proposed model has a high level of performance in the multi-class classification of CKD with the ability to effectively combine preprocessing, class balancing, and hyperparameter optimization. The SMOTE with the GridSearchCV combination has the capability to enhance the generalization of the model and ensure that all classes make the same prediction. The findings show that there is better precision, recall, and F1-score which proves the reliability and the strength of the framework in clinical decision support applications.

A. Result of the Proposed Study

The proposed structure is very useful in the classification of multiple classes of CKD because it integrates all the elements of Random Forest, SMOTE, and GridSearchCV. The overall accuracy of the model is 94.12, which implies the accuracy of the model in predicting all stages of CKD. Curves of training and validation error have excellent convergences and the minimum error was observed with the 87th tree, which shows good generalization and low level of overfitting. Besides, the model's accuracy is high (0.93), recall is high (0.94) and F1 score is high (0.94), which means that the model is balanced in classification. Log loss is minimized that is, the confidence in prediction of the probability. A well-balanced True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) is maintained when evaluating the classification in order to increase the true Positive Rate (TPR) and the true Negative Rate (TNR) for all classes.

Table III. Results of the proposed study.

Metric	Value
Accuracy	94.12%
Precision	0.93
Recall	0.94
F1-Score	0.94
Training Error	0.1
Validation Error	0.14
Log Loss	0.21

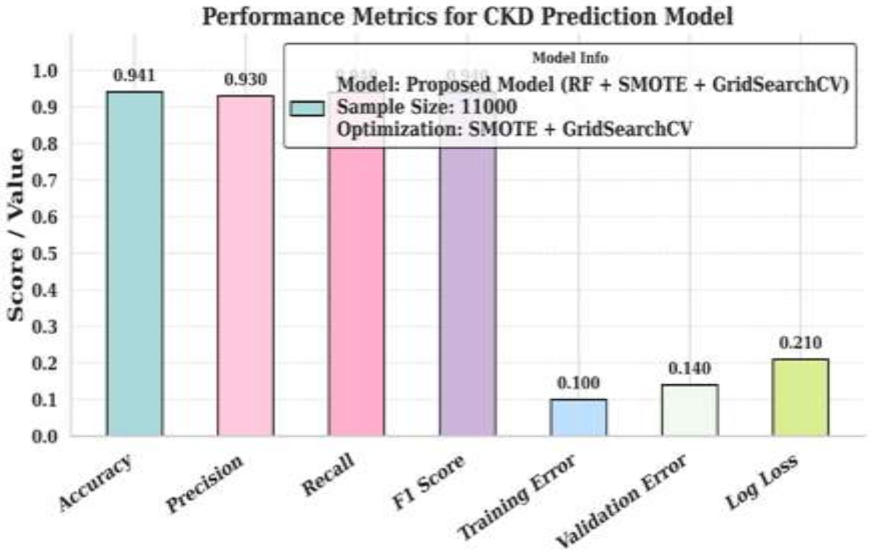


Figure 2: Results of proposed study.

DOI: 10.17148/IJARCCE.2026.155214

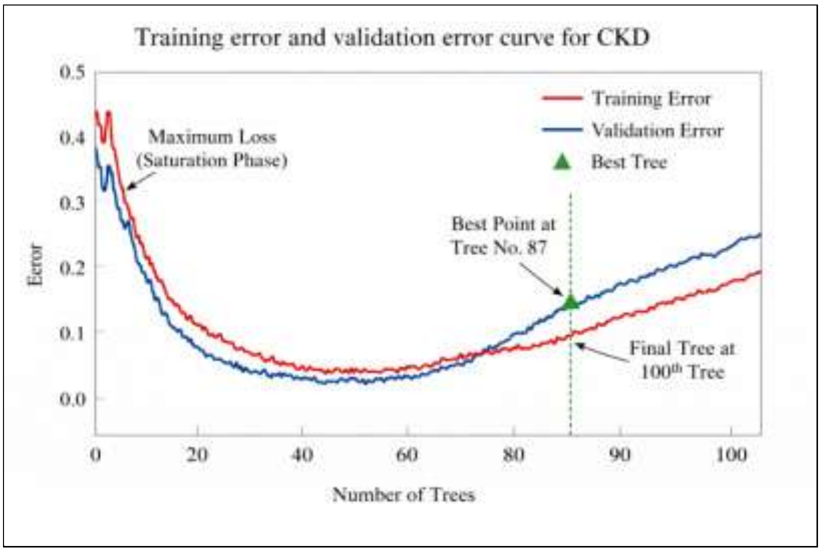


Figure 3: Training and Validation error in model.

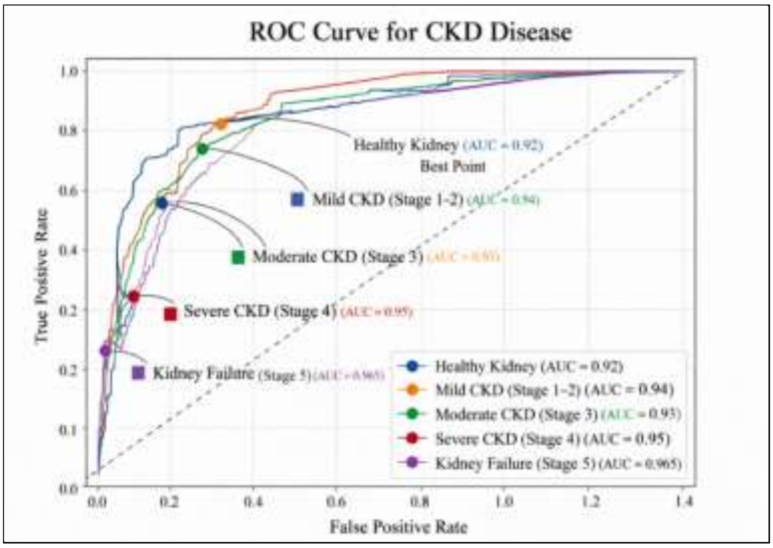
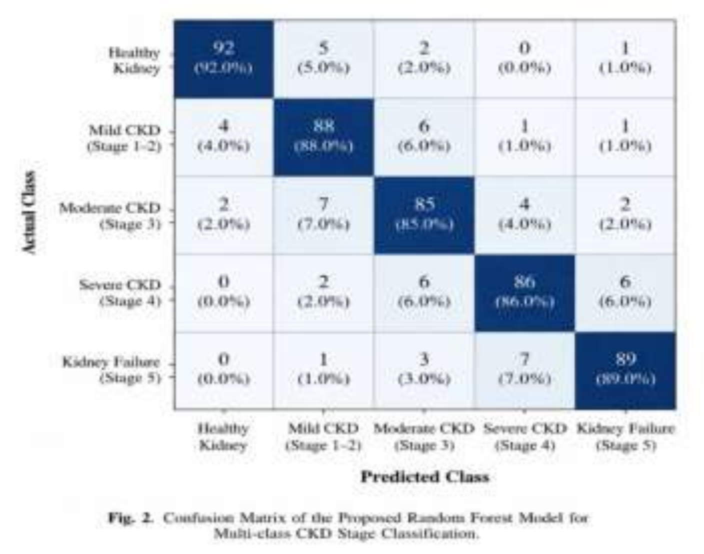


Figure 4: ROC curve for the model.



	Healthy Kidney	Mild CKD (Stage 1-2)	Moderate CKD (Stage 3)	Severe CKD (Stage 4)	Kidney Failure (Stage 5)
Healthy Kidney	92 (92.0%)	5 (5.0%)	2 (2.0%)	0 (0.0%)	1 (1.0%)
Mild CKD (Stage 1-2)	4 (4.0%)	88 (88.0%)	6 (6.0%)	1 (1.0%)	1 (1.0%)
Moderate CKD (Stage 3)	2 (2.0%)	7 (7.0%)	85 (85.0%)	4 (4.0%)	2 (2.0%)
Severe CKD (Stage 4)	0 (0.0%)	2 (2.0%)	6 (6.0%)	86 (86.0%)	6 (6.0%)
Kidney Failure (Stage 5)	0 (0.0%)	1 (1.0%)	3 (3.0%)	7 (7.0%)	89 (89.0%)

Fig. 2. Confusion Matrix of the Proposed Random Forest Model for Multi-class CKD Stage Classification.

Figure 5: Confusion Matrix for the model.

B. Comparison of Different Techniques

Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Gradient Boosting are some of the machine learning techniques used in the CKD classification problem. However, these models usually have problems with having an unbalanced distribution of classes and are sensitive to parameters. Finally, the suggested model, which is the combination of Random Forest, SMOTE, and GridSearchCV, is more efficient as it is able to handle the imbalanced data set and simplify the parameters of the model. This combination results in better generalization, stability and prediction in all CKD stages, and is more predictable than the traditional approach.

Table IV. Table for Comparison of different model.

Decision Tree	85.2	0.84	0.83	0.83
KNN	87.45	0.86	0.85	0.85
SVM	89.1	0.88	0.87	0.87
Gradient Boosting	91.3	0.9	0.9	0.9
Random Forest (Baseline)	92.4	0.91	0.91	0.91
Proposed Method (RF + SMOTE + GridSearchCV)	94.12	0.93	0.94	0.94

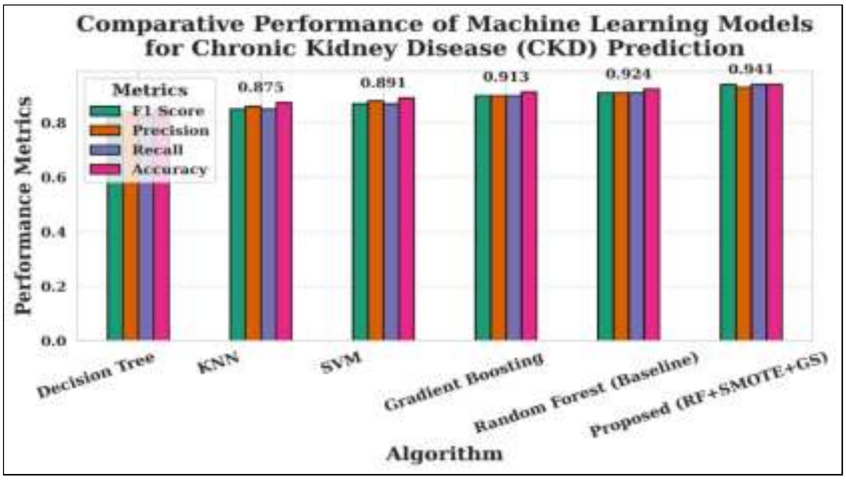


Figure 6: Comparison of different Model.

C. Ablation Study

An ablation study is used to assess the contribution of each of the elements included in the proposed framework by removing important elements of the proposed framework, namely, SMOTE and GridSearchCV. The original model of the Random Forest, which doesn't have improvements, has a relatively poorer performance due to class imbalance and the parameters. The model is better on minority classes, as seen from the better recall and F1-score with SMOTE. Additional enhancement can be seen by the use of GridSearchCV that optimizes hyperparameters and improves the overall model generalization. Both SMOTE and GridSearchCV combine and perform the best, which proves that both elements are important to enhance the effectiveness and strength of the classification at all stages of CKD.

Table V: Table Ablation Study.

Model Configuration	Accuracy (%)	Precision	Recall	F1-Score
Random Forest (Baseline)	90.85	0.89	0.88	0.88
Random Forest + SMOTE	92.3	0.91	0.92	0.91
Random Forest + GridSearchCV	92.75	0.92	0.91	0.91
Proposed (RF + SMOTE + GridSearchCV)	**94.12**	**0.93**	**0.94**	**0.94**

DOI: 10.17148/IJARCCE.2026.155214

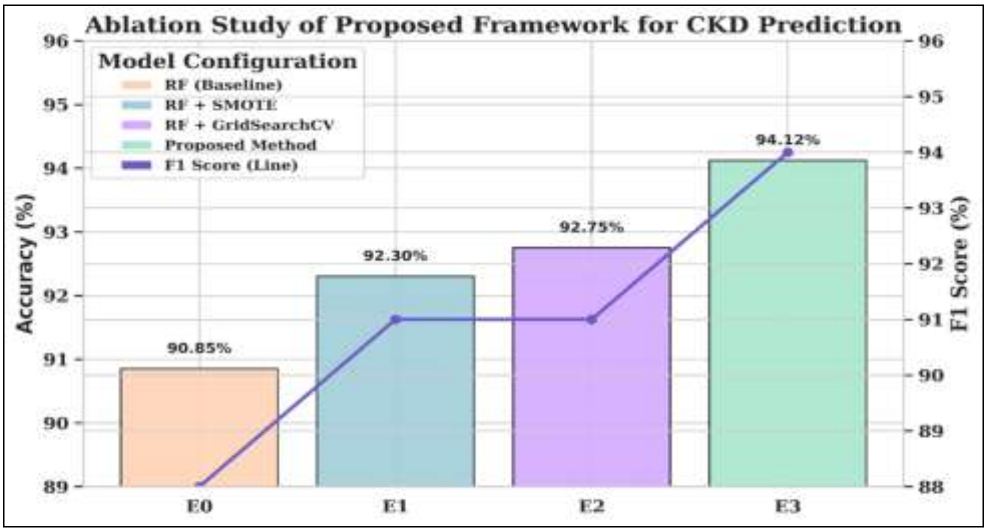


Figure 7: Figure Ablation Syudy.

VI. DISCUSSION

From the methodology used, it can be concluded that the use of pre-processing, balance of classes and ensemble learning is advantageous for the CKD classification problem. The use of clinical features enables a wide set of features related to different measures of CKD to be analyzed. With appropriate handling and processing of variables, it is possible to incorporate both numeric and categorical values in the process of learning effectively.The utilization of SMOTE algorithm is especially important for medical datasets as it helps resolve problems related to class imbalance. With its aid, it was possible to be more sensitive towards minority classes in the course of CKD stages. Furthermore, the use of GridSearchCV with its better parameter tuning capability of the algorithm also provided an improvement in the results. It is not surprising that RF was selected as the ensemble learning algorithm since it is an algorithm with desirable properties for medical applications and was able to solve the problem at hand.

Overall, the research reveals that the use of suitable balancing and optimizing methods coupled with ensemble learning is a suitable approach for the classification of CKD stages.

VII. CONCLUSIONS AND FUTURE WORK

From the above research, it was concluded that SMOTE and GridSearchCV is an efficient technique to classify the CKD stages as multi-class classification using application of random forest. The technique that is proposed can successfully address the important problems such as class imbalance, heterogeneous features and hyperparameters optimization. Comprehensive clinical features and preprocessing contribute to increasing the capability of differentiation among various CKD stages. Balanced learning in all classes is achieved by integrating SMOTE; and GridSearchCV generalizes the model with optimal parameters. From the results it has been found that the proposed technique is reliable and can be applied in the implementation of decision making system in clinical practice.Future directions include adding new components to the existing model such as using deep learning techniques and hybrid models to increase its efficiency. Moreover, the model could be coupled with real-time clinical practice data. Other improvement methods for the extension of the model are the use of more sophisticated feature selection and explainable AI models.

LIST OF ABBREVIATIONS USED IN THE STUDY

Abbreviation	Full Form
CKD	Chronic Kidney Disease
RF	Random Forest
SMOTE	Synthetic Minority Oversampling Technique
GridSearchCV	Grid Search Cross-Validation
ML	Machine Learning
AI	Artificial Intelligence
eGFR	Estimated Glomerular Filtration Rate
BMI	Body Mass Index

BP	Blood Pressure
BUN	Blood Urea Nitrogen
HbA1c	Hemoglobin A1c
RBC	Red Blood Cell
WBC	White Blood Cell
TP	True Positive
FN	False Negative

REFERENCES

[1]. M. S. Al Huda, E. Kanon, M. S. K. Pappo, M. A. Ali and N. Ahmed, "NefroAI: An Explainable and Real-Time Framework for Predicting Chronic Kidney Disease Using Diverse Machine Learning Models and Different Feature Selection Techniques," in IEEE Access, vol. 14, pp. 10939-10976, 2026, doi: 10.1109/ACCESS.2025.3649006.

[2]. V. Preethi Samuel and G. Annasamy, "Ultrasensitive Detection of Creatinine Using Deep Learning-Integrated Graphene Oxide Gold Nanocomposites SERS Sensor," in IEEE Sensors Letters, vol. 10, no. 1, pp. 1-4, Jan. 2026, Art no. 6000604, doi: 10.1109/LSENS.2025.3642276.

[3]. Xue, N., Bai, T., Jia, X. et al. “Early detection of chronic kidney disease based on a SURD-enhanced machine learning model”, Sci Rep ,2026, doi.org/10.1038/s41598-026-41050-z.

[4]. P. Rahi, I. Singh, S. Mahajan and R. Chauhan, ""A Comprehensive Approach to Financial Market Forecasting: Integrating Fundamental and Derivative Analysis and Prediction,"International Conference on Advances in Computing, Communication and Materials (ICACCM), Dehradun, India, 2024, pp. 1-6, doi: 10.1109/ICACCM61117.2024.11059083.

[5]. Iftikhar, H., Hashem, A.F., Mohamud, L.A. et al. “An intelligent ensemble machine learning model for early detection of chronic kidney disease in aging populations”, Sci Rep 16, 3021, 2026, doi.org/10.1038/s41598-025-32919-6.

[6]. S. K. Ghosh, N. Widatalla and A. H. Khandoker, "Machine Learning Framework for Early Detection of Chronic Kidney Disease Stages Using Optimized Estimated Glomerular Filtration Rate," in IEEE Access, vol. 13, pp. 78057-78072, 2025, doi: 10.1109/ACCESS.2025.3565549.

[7]. V. Puri, I. Priyadarshini, A. Kataria, S. Rani and H. Min, "Privacy-First ML for Chronic Kidney Disease Prediction: Exploring a Decentralized Approach Using Blockchain and IPFS," in IEEE Access, vol. 13, pp. 43178-43189, 2025, doi: 10.1109/ACCESS.2025.3548645.

[8]. P. Rahi, S. S. Kang, A. P. Singh and I. Singh, "A Random Forest Framework for Predicting Cardiovascular Disease in Diverse Populations," Second International Conference on Advanced Computing & Communication Technologies (ICACCTech), Sonipat, India, 2024, pp. 472-478, doi: 10.1109/ICACCTech65084.2024.00083.

[9]. K. M. Tawsik Jawad, A. Verma, F. Amsaad and L. Ashraf, "A Study on the Application of Explainable AI on Ensemble Models for Predictive Analysis of Chronic Kidney Disease," in IEEE Access, vol. 13, pp. 23312-23330, 2025, doi: 10.1109/ACCESS.2025.3535692.

[10]. M. Azad, A. A. Abd El-Aziz, S. Abd El-Ghany and M. Faraz Kabir Khan, "SAFF-ConvNeXt: A Scale-Aware Feature Fusion and SE-Attention Enhanced ConvNeXt Framework for Kidney Disease Detection in Medical Imaging," in IEEE Access, vol. 13, pp. 197475-197498, 2025, doi: 10.1109/ACCESS.2025.3634972.

[11]. Y. Zhao, K. Zhang, X. Duan, S. Che, and N. Ma, “A chronic kidney disease diagnostic model based on an interpretable deep belief rule base,” IEEE Access, vol. 13, pp. 29718–29737, Feb. 2025, doi: 10.1109/ACCESS.2025.3541391.

[12]. M. Thakur, A. P. Singh, I. Singh, P. Rahi and Y. Mahajan, "Utilizing Fuzzy Logic in MATLAB for Visualizing Risk Factors associated with Cardiovascular Attacks,"International Conference on Computational Intelligence and Computing Applications (ICCICA), Samalkha, India, 2024, pp. 564-570, doi: 10.1109/ICCICA60014.2024.10585163.

[13]. Metherall, B., Berryman, A.K. & Brennan, G.S. “Machine learning for classifying chronic kidney disease and predicting creatinine levels using at-home measurements”, Sci Rep 15, 4364 ,2025, doi.org/10.1038/s41598-025-88631-y.

[14]. C. T. Hsu, C. Y. Huang, C. H. Chen, et al., “Machine learning models to predict osteoporosis in patients with chronic kidney disease stage 3–5 and end-stage kidney disease,” Scientific Reports, vol. 15, p. 11391, 2025, doi: 10.1038/s41598-025-95928-5.

[15]. E. Vanden Broecke, L. Van Mulders, E. De Paepe, et al., “Early detection of feline chronic kidney disease via 3-hydroxykynurenine and machine learning,” Scientific Reports, vol. 15, p. 6875, 2025, doi: 10.1038/s41598-025-90019-x.

[16]. P. Rahi, I. Singh, Anushka, A. Kumar and T. Baliyan, "“Boosting Tuberculosis Classification Accuracy with Polynomial Features and Random Forests”," Second International Conference on Advanced Computing &

Communication Technologies (ICACCTech), Sonipat, India, 2024, pp. 479-485, doi: 10.1109/ICACCTech65084.2024.00084.

[17]. J. Li, X. Du, R. Zhang, et al., “Machine learning models for predicting short-term progression in patients with stage 4 chronic kidney disease: a multi-center validation study,” Scientific Reports, vol. 15, p. 39285, 2025, doi: 10.1038/s41598-025-23037-4.

[18]. B. Yang, X. Yang, H. Sun, et al., “Identification and validation of inflammatory response genes linking chronic kidney disease with coronary artery disease based on bioinformatics and machine learning,” Scientific Reports, vol. 15, p. 19184, 2025, doi: 10.1038/s41598-025-03622-3.

[19]. U. Bialonczyk, M. Debowska, L. Dai, et al., “Balancing accuracy and cost in machine learning models for detecting medial vascular calcification in chronic kidney disease: a pilot study,” Scientific Reports, vol. 15, p. 17453, 2025, doi: 10.1038/s41598-025-02457-2.

[20]. A. P. Singh, R. Khan, P. Rahi, I. Singh and N. Basu, "An Improved Login Authenticable Model Using Memory Gallery Apps," International Conference on Next Generation Electronics (NEleX), Vellore, India, 2023, pp. 1-7, doi: 10.1109/NEleX59773.2023.10421280.

[21]. W. Yang, N. Ahmed and A. L. C. Barczak, "Comparative Analysis of Machine Learning Algorithms for CKD Risk Prediction," in IEEE Access, vol. 12, pp. 171205-171220, 2024, doi: 10.1109/ACCESS.2024.3499355.

[22]. 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," in IEEE Access, vol. 12, pp. 126389-126414, 2024, doi: 10.1109/ACCESS.2024.3451369.

[23]. D. Yadav and N. P. Singh, "Advancing Chronic Kidney Disease (CKD) Diagnosis: A Comparative Analysis of Machine Learning and Deep Learning Approaches," 2024 2nd International Conference on Advancements and Key Challenges in Green Energy and Computing (AKGEC), Ghaziabad, India, 2024, pp. 1-6, doi: 10.1109/AKGEC62572.2024.10868884.

[24]. A. P. Singh, P. Rahi, B. Sahu, N. Basu, H. Fida and V. Yadav, "Deep Learning-Based Road Sign Detection: The Ultimate Road Technology," 1st International Conference on Advanced Computing and Emerging Technologies (ACET), Ghaziabad, India, 2024, pp. 1-6, doi: 10.1109/ACET61898.2024.10730463.

[25]. D. Janani and S. Ramakrishnan et al., "Machine Learning Approaches for CKD Identification among HIV-Afflicted Individuals," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-5, doi: 10.1109/ICCCNT61001.2024.10725376.

[26]. V. D, S. M. Ramesh, S. K, T. K, S. N and K. P, "A Machine Learning Perspective for Predicting Chronic Kidney Disease," 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2024, pp. 989-993, doi: 10.1109/ICSCSS60660.2024.10625341.

[27] . M. S. Islam and M. M. Pias, "An Efficient Early Identification of Chronic Kidney Disease using Machine Learning Techniques," 2024 27th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2024, pp. 2180-2185, doi: 10.1109/ICCIT64611.2024.11022079.

[28]. S. S. Kang, P. Rahi, A. P. Singh and I. Singh, “Liver Disease Risk Prediction Using Regularized Deep Learning: A Novel Approach”," 2nd International Conference on Advancements and Key Challenges in Green Energy and Computing (AKGEC), Ghaziabad, India, 2024, pp. 1-7, doi: 10.1109/AKGEC62572.2024.10868777.

[29]. T. S. Rafa, U. Saha, M. R. Rayhan, S. K. Bosu and A. Jawad, "Chronic Kidney Diseases Detection: A Machine Learning-Based Approach on Bangladeshi Patient Dataset," 2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP), VIJAYAWADA, India, 2024, pp. 1-5, doi: 10.1109/AISP61711.2024.10870713.

[30]. R. Bhuria, "Advanced Machine Learning Techniques for Chronic Kidney Disease Prediction and Management," 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit), Greater Noida, India, 2024, pp. 480-484, doi: 10.1109/GlobalAISummit62156.2024.10947932.

[31]. M. K. S., V. K. D, M. G., A. M. C., M. D. Nagarajan and D. C., "An Analysis to Predict the Occurrence of Chronic Kidney Disease Using Ensemble Learning Algorithms," 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2024, pp. 1605-1608, doi: 10.1109/ICSCSS60660.2024.10625186.

[32]. P. Rahi, I. Singh, A. P. Singh and A. Kumar, "“RNN-Driven Prognosis of Heart Disease Based on Health Parameters”," Second International Conference on Advanced Computing & Communication Technologies (ICACCTech), Sonipat, India, 2024, pp. 486-492, doi: 10.1109/ICACCTech65084.2024.00085.

[33]. H. Miao, F. Liu, Y. N. Wang, et al., “Targeting Lactobacillus johnsonii to reverse chronic kidney disease,” Signal Transduction and Targeted Therapy, vol. 9, p. 195, 2024, doi: 10.1038/s41392-024-01913-.

[34]. Z. Pan, M. Derbala, K. AlNaamani, H. Ghazinian, J.-G. Fan, and M. Eslam, “MAFLD criteria are better than MASLD criteria at predicting the risk of chronic kidney disease,” Annals of Hepatology, vol. 29, no. 5, p. 101512, 2024, doi: 10.1016/j.aohep.2024.101512.

[35] . Á. G. Pethő, M. Tapolyai, É. Csongrádi, and P. Orosz, “Management of chronic kidney disease: The current novel and forgotten therapies,” Journal of Clinical & Translational Endocrinology, vol. 36, p. 100354, 2024, doi: 10.1016/j.jcte.2024.100354.

[36]. P. Rahi, M. R. Hussain, S. Alqahtani, M. Husain, A. P. Singh and I. Singh, "A Novel Random Forest-SMOTE Framework With Polynomial Feature Engineering for Early Detection of Gastrointestinal Disorders," in IEEE Access, vol. 13, pp. 188577-188604, 2025, doi: 10.1109/ACCESS.2025.3626072.

[37]. C. Delrue, S. De Bruyne, and M. M. Speeckaert, “Application of machine learning in chronic kidney disease: Current status and future prospects,” Biomedicines, vol. 12, no. 3, p. 568, 2024, doi: 10.3390/biomedicines12030568.

[38]. J. Bilson, A. Mantovani, C. D. Byrne, and G. Targher, “Steatotic liver disease, MASLD and risk of chronic kidney disease,” Diabetes & Metabolism, vol. 50, no. 1, p. 101506, 2024, doi: 10.1016/j.diabet.2023.101506.

[39]. P. E. Stevens, S. B. Ahmed, J. J. Carrero, et al., “KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease,” Kidney International, vol. 105, no. 4, Suppl., pp. S117–S314, 2024, doi: 10.1016/j.kint.2023.10.018.

[40]. P. Rahi, I. Singh, S. Mahajan, R. Chauhan, T. Kaur and J. Singh, ""From Genes to Symptoms: Predicting Huntington’s Disease Progression with DNN"," International Conference on Electronics, AI and Computing (EAIC), Jalandhar, India, 2025, pp. 1-9, doi: 10.1109/EAIC66483.2025.11101647.

[41]. P. Liang et al., "Deep Learning Identifies Intelligible Predictors of Poor Prognosis in Chronic Kidney Disease," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 7, pp. 3677-3685, July 2023, doi: 10.1109/JBHI.2023.3266587.

[42]. G. Shukla, G. Dhuriya, S. K. Pillai and A. Saini, "Chronic Kidney Disease Prediction Using Machine Learning Algorithms and the Important Attributes for the Detection," 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), London, United Kingdom, 2023, pp. 1-4, doi: 10.1109/GlobConET56651.2023.10149900.

[43]. C. Choudhary, L. S. Nagra, P. Das, J. Singh and S. S. Jamwal, "Optimized Ensemble Machine Learning Model for Chronic Kidney Disease Prediction," 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2023, pp. 292-297, doi: 10.1109/ICCCIS60361.2023.10425073.

[44]. P. Rahi, S. S. Kang, A. P. Singh and I. Singh, "“Advanced Anxiety Risk Stratification Using Regularized Deep Neural Networks," International Conference on Progressive Innovations in Intelligent Systems and Data Science (ICPIDS), Pattaya, Thailand, 2024, pp. 524-530, doi: 10.1109/ICPIDS65698.2024.00087.

[45]. L. A. Akinyemi, O. P. Oshinuga, S. O. Ekwe and S. O. Oladejo, "Enhancing Chronic Kidney Disease Prediction Through Data Preprocessing Optimization and Machine Learning Techniques," 2023 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 2023, pp. 1-6, doi: 10.1109/ICECET58911.2023.10389513.

[46]. N. K. Pareek, D. Soni and S. Degadwala, "Prediction of CKD Using Expert System Fuzzy Logic & AI," 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 103-108, doi: 10.1109/ICAISS58487.2023.10250477.

[47]. S. Srivastav, K. Guleria and S. Sharma, "Predictive Machine Learning Approaches for Chronic Kidney Disease," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-5, doi: 10.1109/INCET57972.2023.10170425.

[48]. S. Singla and P. Rahi, "Clinically Interpretable and Statistically Robust EHR-Based Ensemble Model for Anxiety Screening," Eighth International Conference on Image Information Processing (ICIIP), Solan (Near Shimla) Himachal Pradesh, India, 2025, pp. 219-226, doi: 10.1109/ICIIP68302.2025.11346441.

[49]. A. S. Mahmoud, O. Lamouchi and S. Belghith, "A review on machine learning driven models for early diagnosis of chronic kidney diseases," 2023 16th International Conference on Developments in eSystems Engineering (DeSE), Istanbul, Turkiye, 2023, pp. 818-821, doi: 10.1109/DeSE60595.2023.10468825.

[50]. Q. Bai, C. Su, W. Tang, et al., “Machine learning to predict end stage kidney disease in chronic kidney disease,” Scientific Reports, vol. 12, p. 8377, 2022, doi: 10.1038/s41598-022-12316-z..

[51]. M. Mizdrak, M. Kumrić, T. T. Kurir, and J. Božić, “Emerging biomarkers for early detection of chronic kidney disease,” Journal of Personalized Medicine, vol. 12, no. 4, p. 548, 2022, doi: 10.3390/jpm12040548.

[52]. P. Rahi and S. Singla, "Dual-Pipeline LSTM Screening for Anxiety and Depression via Video Interviews," 2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2025, pp. 1-9, doi: 10.1109/ICCCA66364.2025.11325496.

[53]. P. Barik, “Chronic kidney disease (CKD) clinical dataset,” Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/priyankabarik/chronic-kidney-disease-ckd-clinical-dataset.

DOI: 10.17148/IJARCCE.2026.155214

BIOGRAPHY

I am Kapil, and I completed my Bachelor’s degree in Computer Science and Engineering in 2024. Currently, I am pursuing my Master’s degree in the Department of Computer Science and Engineering at Sri Satya Sai Technical University. My areas of interest include artificial intelligence, machine learning, and emerging technologies. I am passionate about research and continuously improving my technical and analytical skills through academic and practical learning. experiences.