Chronic kidney Disease Classification through Hybrid Feature Selection and Ensemble Deep Learning

N. Yogesh[1], Purohit Shrinivasacharya[1], Nagaraj Naik[2,*] and B.M. Vikranth[3]

[1]*Siddaganga Institute of Technology, Tumkur, Karanataka, and Visvesveraya Technological University, Belagavi, India*

[2]*Manipal Institute of Technology, MAHE, Manipal, 576104, Karnataka, India*

[3]*B.M.S. College of Engineering, Bangalore, 560 019, India*

Abstract: Diagnosing and treating at-risk patients for chronic kidney disease (CKD) relies heavily on accurately classifying the disease. The use of deep learning models in healthcare research is receiving much interest due to recent developments in the field. CKD has many features; however, only some features contribute weightage for the classification task. Therefore, it is required to eliminate the irrelevant feature before applying the classification task. This paper proposed a hybrid feature selection method by combining the two feature selection techniques: the Boruta and the Recursive Feature Elimination (RFE) method. The features are ranked according to their importance for CKD classification using the Boruta algorithm and refined feature set using the RFE, which recursively eliminates the least important features. The hybrid feature selection method removes the feature with a low recursive score. Later, selected features are given input to the proposed ensemble deep learning method for classification. The experimental ensemble deep learning model with feature selection is compared to Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF) models with and without feature selection. When feature selection is used, the ensemble model improves accuracy by 2%. Experimental results found that these features, age, pus cell clumps, bacteria, and coronary artery disease, do not contribute much to accurate classification tasks. Accuracy, precision, and recall are used to evaluate the ensemble deep learning model.

Keywords: Deep learning, Feature selection, Recursive Feature Elimination.

1. INTRODUCTION

Long-term chronic kidney disease (CKD) is significant in the recent medical field [1, 2]. Millions of individuals worldwide are affected, and it has been linked to higher illness, death, and healthcare expenditures [3, 4]. Early intervention and tailored therapies can slow disease development and minimize complications. Therefore, the timely classification of CKD is critical for appropriate therapy and management. This study aims to develop CKD classification model with high accuracy. Clinical variables such as age, gender, blood pressure, laboratory test results, and medical history are included in the CKD dataset used to evaluate the effectiveness of the proposed methodology [5, 6].

Moreover, there are 24 features in CKD, but analyzing the critical feature for classification is challenging [7, 8]. Selecting appropriate features is essential when building reliable and robust classification models. Selecting a subset of features from the feature vector improves model performance by reducing the feature space. Limited studies in the literature have considered feature selection methods that could be more robust for the classification task. Therefore, this work adopts a hybrid feature selection method combining Boruta and RFE [9, 10].

Boruta and RFE are popular methods for feature selection tasks [11, 12]. This is the first approach for CKD classification, considering a hybrid feature selection method by combining Boruta and Recursive Feature Elimination (RFE) methods. Boruta provides a complete rating of feature relevance by comparing each feature to randomly generated shadow features. On the other hand, RFE uses a ranking system to gradually remove unnecessary features from feature sets.

Boruta delivers a thorough evaluation of feature value. This aids in determining which characteristics are most helpful in classifying CKD. However, RFE works by gradually excluding less relevant elements until only the most informative ones remain. The combination of Boruta and RFE feature selection strategies to keep just the most essential characteristics in the model.

Applying machine learning methods to medical diagnosis and decision-making has demonstrated encouraging outcomes in recent years [13-15]. In particular, a subset of machine learning techniques known as the deep learning method has received much attention for their impressive capacity to automatically

*Address correspondence to this author at the Manipal Institute of Technology, MAHE, Manipal, 576104, Karnataka, India Email:nagaraj.naik@manipal.edu

E-ISSN: 1929-6029/25

learn and extract complicated patterns and characteristics from complex datasets. Image analysis, disease diagnosis, and prognosis prediction are just a few examples of the many areas of medicine where deep learning models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have shown extraordinary performance [16, 17].

Conventional machine learning methods may not fully capture the complexity of CKD. To address this, we propose a hybrid feature selection method to identify important features and an ensemble deep learning technique to extract relevant patterns from the data. The model's performance is evaluated using accuracy, precision, and recall.

The key contributions of this work are as follows:

- We propose a Hybrid Feature Selection technique to identify high-regression features in Chronic Kidney Disease (CKD).
- We develop an ensemble deep learning classifier to improve the classification accuracy of CKD.

The paper is structured as follows: Section 2 reviews literature on CKD classification. Section 3 outlines the methodology combining Boruta, RFE, and ensemble deep learning for CKD classification. Section 4 presents experimental findings and analyses. Section 5 concludes with a summary of contributions and significance.

2 RELATED WORK

Ghosh *et al*. [18] introduced four classification methods for chronic kidney disease (CKD), including support vector machines (SVM), AdaBoost, linear discriminant analysis (LDA), and gradient boosting. Their study concluded that AdaBoost outperforms the other models.

Aljaaf *et al*. [19] evaluated the performance of various machine learning techniques for early prediction of chronic kidney disease (CKD), starting with 24 features and the target class. Among the four classifiers compared in a supervised learning setup, the multilayer perceptron achieved the best performance.

Alassaf *et al*. [20] explored data mining and machine learning techniques to preemptively identify chronic renal disease. Their findings revealed that artificial neural networks (ANN) and support vector machines (SVM) achieved higher accuracy than k-nearest neighbors (KNN). Similarly, Qin *et al*. addressed missing values in the UCI dataset using KNN imputation and evaluated six models, with random forest (RF) demonstrating the best performance. Mezzatesta *et al*. [21] applied support vector machines (SVM) with radial basis function (RBF) kernels, optimizing hyperparameters with GridSearch. This study proposes a hybrid selection-based ensemble approach for CKD.

3. PROPOSED WORK

There are 24 features in CKD described in Table **1**. A correlation plot is a graphical representation of the correlations between variables in CKD data, which is evaluated using correlation coefficients. The correlation coefficient quantifies the direction and intensity of a linear relationship between two variables. Figure **1** shows the possible interpretations of the many kinds of correlation coefficients.

Table 1: CKD Feature Description

Feature No	Name	Descriptions
1	Age	Patient Age
2	BP	Blood Pressure
3	SG	Specific gravity
4	AL	Albumin
5	SU	Sugar
6	RBC	Red blood cells
7	PC	Pus cell
8	PCC	Pus cell clumps
9	BA	Bacteria
10	BGR	Blood glucose random
11	BU	Blood urea
12	SC	Serum creatinine
13	SOD	Sodium
14	POT	Potassium
15	HEMO	Hemoglobin
16	PCV	Packed cell volume
17	WC	White blood cell count
18	RC	Red blood cell count
19	HTN	Hypertension
20	DM	Diabetes mellitus
21	CAD	Coronary artery disease
22	APPET	Appetite
23	PE	Pedal edema
24	ANE	Anemia
25	CLASS	Diagnosis ckd, notckd

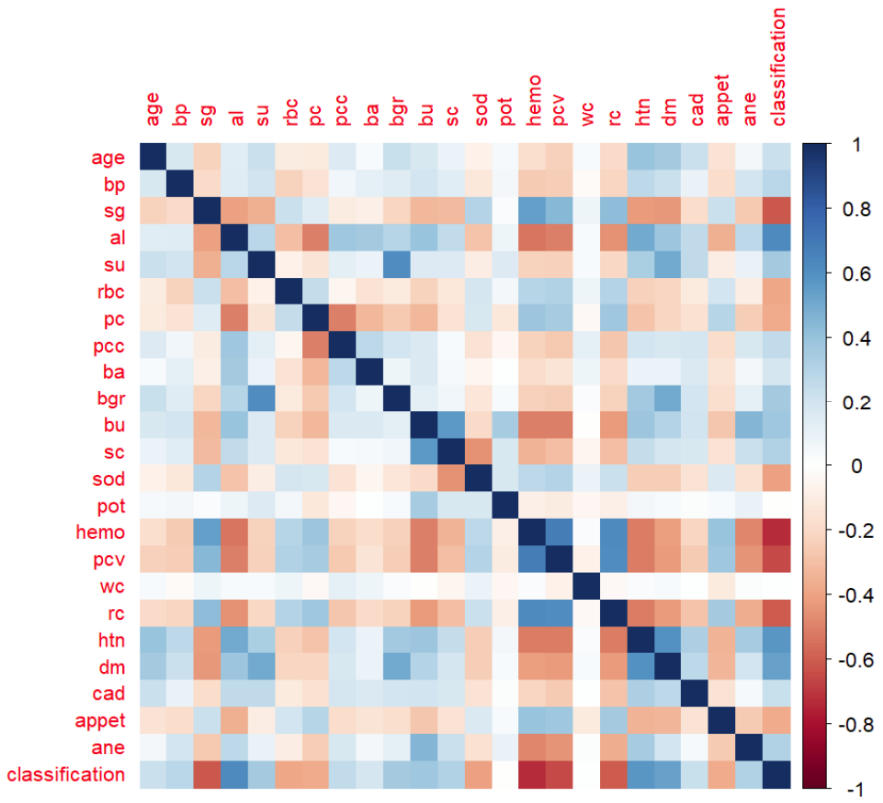

Figure 1: Correlation Coefficients of CKD Feature.

Feature assessment is crucial for improving CKD classification. Some features offer minimal contribution, necessitating the removal of irrelevant ones. Limited research exists on effective feature selection for CKD. Identifying unimportant features is essential, as they increase computational time and add noise. A hybrid feature selection approach is required to select CKD features. Redundant features can reduce accuracy and lead to overfitting, where the model performs well on training data but struggles with unseen data.

Algorithm 1: Hybrid Feature Selection

1. Read Data.
2. Handle missing values using the KNN method.
3. Split Data into Features and Target Variable. X_1 to X_{24} are input features, and Y_{25} is the Target Feature, i.e., class variable.
4. Apply the Boruta method for feature selection.
5. Compute the Z-Score:

$$\text{Z} - \text{Score} = \frac{\text{Feature Score} - \text{Shadow Feature Mean}}{\text{Shadow Feature Standard Deviation}}$$

6. Select features with a Z-Score above 0.90 using the Boruta method.
7. Select features with the lowest RMSE score using the RFE method.
8. Collect and combine the selected features from both Boruta and RFE methods.

Figure **2** presents the proposed methodology, which emphasizes evaluating each CKD feature. A hybrid feature selection approach is employed to identify the most significant features of CKD. Initially, there are 24 features in the datasets. To overcome the missing value in the datasets, we considered the K-nearest neighbors(KNN) method and preprocessed the missing value. The hybrid feature selection method combines the Boruta and RFE methods. Boruta is derived from the Random Forest, a decision tree-based ensemble learning technique. Each tree in a Random Forest model predicts the target variable using its unique combination of characteristics, and the models are constructed using bootstrapped samples of the dataset. In shadow, features are generated by the Boruta method. Later, it compares the relevance of each feature in the original dataset to the importance of its corresponding shadow features. The z-score, a

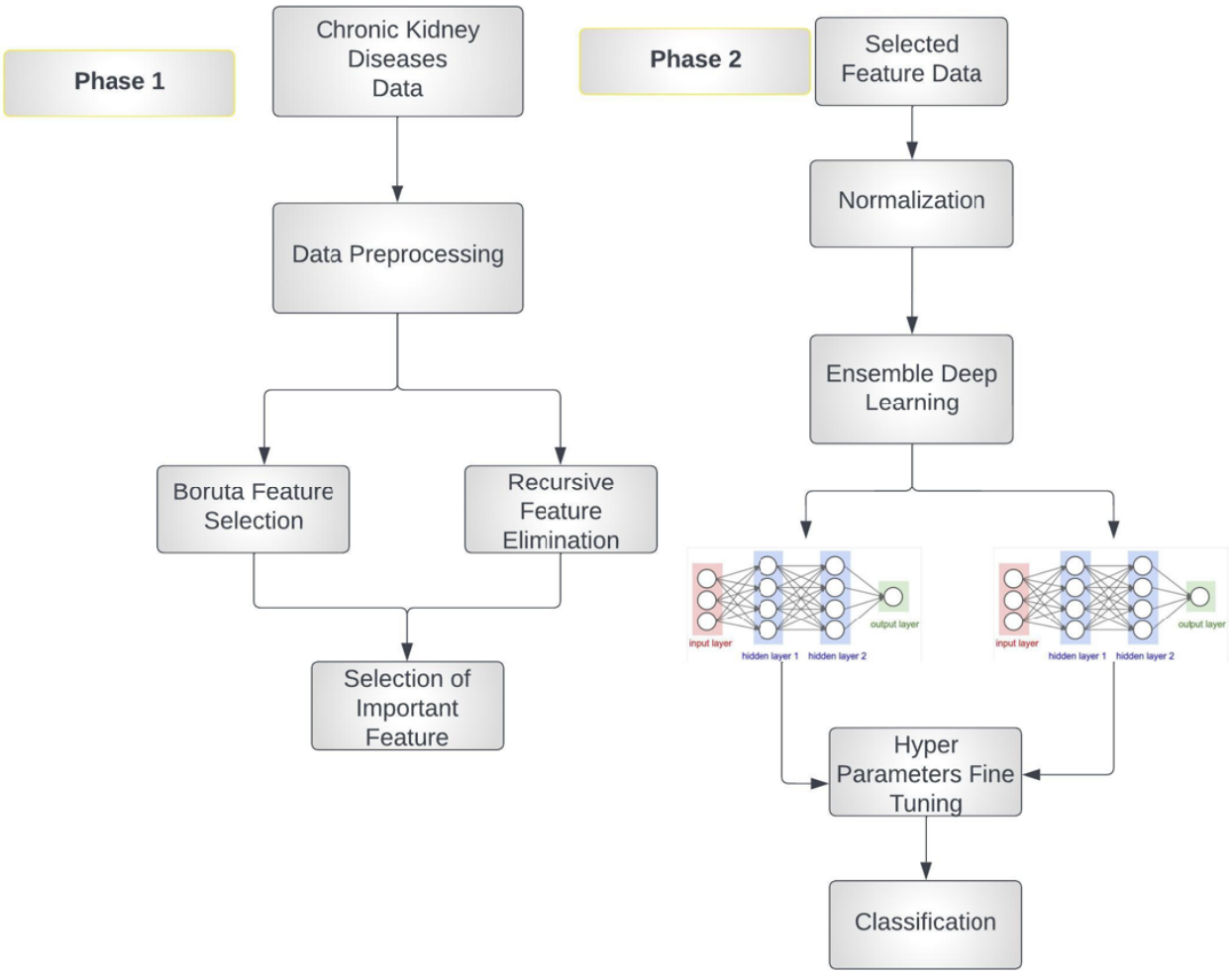

Figure 2: Overall Work.

statistical technique, evaluates the significance of a feature by comparing it to the most essential feature within its corresponding attributes.

RFE eliminates irrelevant features from a dataset depending on their relevance to the target variable. The base model is trained on all features, and each feature is given a relevance score depending on its performance. The dataset's lowest-ranked features are eliminated. The hybrid feature selection detail algorithm is described in 3. The selected features from hybrid feature selection are given input by the Ensemble-based deep learning model.

In phase 2, an ensemble-based deep learning approach is proposed for CKD classification. While most studies rely on individual classifiers, CKD datasets often exhibit nonlinearity, which can hinder accuracy. To address this, the proposed method leverages ensemble-based deep learning to overcome these challenges and improve performance. The proposed method is to classify whether the patient has chronic kidney disease or not. Therefore, we have constructed two, three, and four layers of deep neural networks in Ensemble-based deep learning. The input layer neurons represent the CKD features selected by the hybrid features selection method. The deep neural network equations are defined below. Where w represents the weight and A(x) represents the input data. The total number of input values is defined by variable n.

$$Y_i = (\sum_{k=1}^{n} w_k A(x) + Bias) \quad (1)$$

$$Hidden\ Layer\ 1 = Sigmoid\ Activation\ (\sum_{k=1}^{n} w_k h1(x) + Bias) \quad (2)$$

$$Hidden\ Layer\ 2 = Sigmoid\ Activation\ (\sum_{k=1}^{n} w_k h2(x) + Bias) \quad (3)$$

In the proposed model, the input layer neuron reads CKD data. Initially, random weights are multiplied with

input data and Bias is added to each neuron. After that, each layer neuron is activated using the activation function. The detail Equation are define in 1, 2 and 3. Sigmoid function was considered to activate the neurons in the models. In ensemble deep learning, training comprises fine-tuning the neural network's weights and biases to get the best possible accurate output. For that, we considered the backpropagation method to train the model. The loss function's gradient about the network's weights and biases is computed. Next, stochastic gradient descent (SGD) and its variants employ the gradients to update the weights and biases to minimize the loss function in the ensemble-based deep learning model. Hyperparameters are needed to perform during the settings of the deep learning model, such as the number of neurons, activation function, and learning algorithm. The model was trained with a learning rate of 0.001.

4. RESULTS

In the experimental work, the proposed framework considered CKD. The dataset initially consisted of 24 features, and preprocessing was performed to handle missing values using the KNN method.

The hybrid feature selection method, which combined the Boruta and RFE techniques, was implemented in the next step. Boruta, based on the random forest method, generated shadow features and compared the relevance of each original feature to its corresponding shadow features. The feature importance graph using the Boruta method is shown in Figure **3**, and the feature importance using the RFE method is shown in Figure **4**. Using the Boruta method, we found that these features, age, pus cell clumps (PCC), bacteria (BA), and coronary artery disease (CAD), have low scores for age feature 0.67, PCC 0.61, BA 0.56 and 0.30 and it is described in Table **2**. Using the RFE method, we found that 24 features are relevant. When we tested it on our CKD dataset, Boruta performed better than RFE in detecting relevant and irrelevant characteristics. Boruta was an excellent fit for our dataset because of its resilience in complicated feature interactions and its capacity to manage non-linear relationships. Furthermore, Boruta demonstrated efficacy in cases where the dataset had relevant and noisy characteristics.

In Figure **5**, each histogram represents one of the twenty features included in the dataset. The x-axis shows the feature values, while the y-axis shows how often they occur. It suggests the absence of outliers, and the skewness in the feature distributions indicates a relatively symmetric pattern.

$$Accuracy = \frac{True\ Pos + True\ Neg}{True\ Pos + True\ Neg + False\ Pos + False\ Neg} \quad (4)$$

$$Precision = \frac{True\ Pos}{True\ Pos + False\ Pos} \quad (5)$$

$$Recall = \frac{True\ Pos}{True\ Pos + False\ Neg} \quad (6)$$

$$FScore = 2X\frac{Precision\ X\ Recall}{Precision + Recall} \quad (7)$$

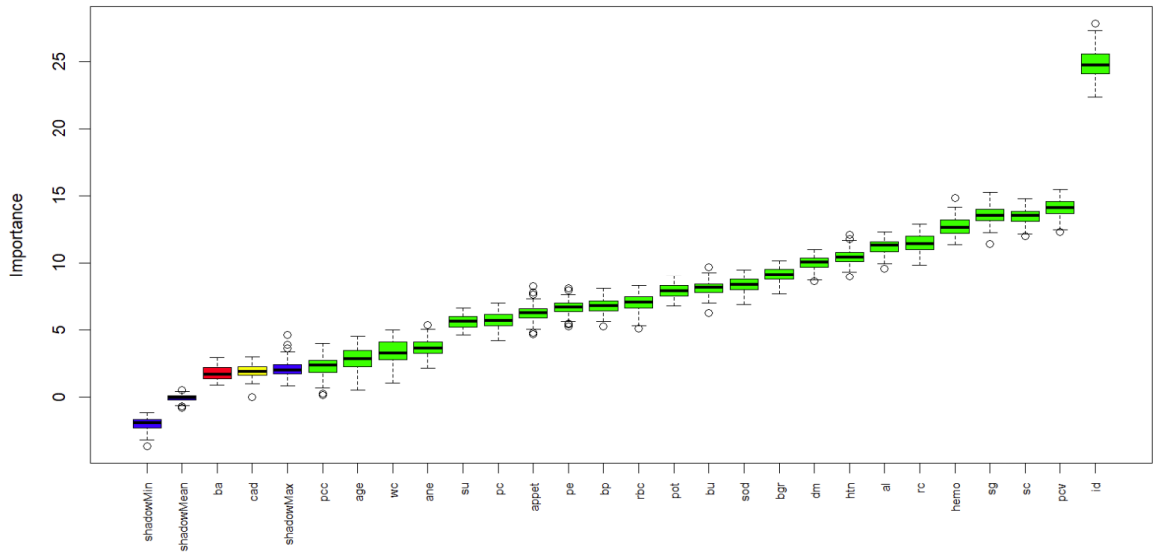

Figure 3: Feature Importance Graph using Boruta Method.

Variables	RMSE	Rsquared	MAE	RMSESD	RsquaredSD	MAESD	Selected
1	0.3142	0.5860	0.17341	0.03208	0.07873	0.02923	
2	0.2265	0.7823	0.10664	0.03373	0.06294	0.02737	
3	0.1748	0.8687	0.07924	0.03522	0.05386	0.02106	
4	0.1513	0.9001	0.06625	0.04253	0.05566	0.02012	
5	0.1369	0.9201	0.06406	0.03897	0.04567	0.01246	
6	0.1228	0.9347	0.04445	0.03323	0.03401	0.01119	
7	0.1195	0.9384	0.04364	0.03324	0.03252	0.01136	
8	0.1215	0.9371	0.04703	0.03255	0.03221	0.01042	
9	0.1284	0.9294	0.04935	0.03559	0.03831	0.01164	
10	0.1255	0.9329	0.04958	0.03352	0.03496	0.01116	
11	0.1212	0.9361	0.04857	0.03912	0.04018	0.01285	
12	0.1239	0.9346	0.04868	0.03287	0.03420	0.01042	
13	0.1214	0.9362	0.04837	0.03768	0.03873	0.01212	
14	0.1210	0.9372	0.04860	0.03511	0.03616	0.01019	
15	0.1216	0.9365	0.04914	0.03604	0.03707	0.01113	
16	0.1198	0.9385	0.04856	0.03649	0.03665	0.01158	
17	0.1203	0.9376	0.04917	0.03832	0.03909	0.01157	
18	0.1222	0.9364	0.04979	0.03489	0.03634	0.01059	
19	0.1211	0.9371	0.04993	0.03704	0.03801	0.01136	
20	0.1215	0.9366	0.05025	0.03728	0.03845	0.01105	
21	0.1228	0.9350	0.05052	0.03855	0.03959	0.01118	
22	0.1204	0.9384	0.05011	0.03545	0.03615	0.01035	
23	0.1186	0.9394	0.05022	0.03759	0.03764	0.01243	*
24	0.1240	0.9347	0.05183	0.03493	0.03615	0.01079	

Figure 4: RFE method Feature Score.

Table 2: Boruta Method Feature Score

Feature	meanImp	medianImp	minImp	maxImp	normHits	decision
1	2.648089	2.757746	0.4239344	4.618911	0.6767677	Rejected
2	6.794479393	6.776096505	5.264811762	8.052124859	1	Confirmed
3	13.51221699	13.52946847	11.43714431	15.24718145	1	Confirmed
4	11.22037475	11.3164117	9.57435126	12.33625622	1	Confirmed
5	5.591968494	5.606924547	4.610428416	6.624882448	1	Confirmed
6	6.974420729	7.019035362	5.093922316	8.309885643	1	Confirmed
7	5.66329417	5.67849841	4.179199421	6.993594095	1	Confirmed
8	2.337024	2.339759	0.1172789	3.389289	0.6161616	Rejected
9	2.209811	2.243064	1.0010015	3.194971	0.5656566	Rejected
10	9.122409835	9.10887171	7.639293553	10.1917889	1	Confirmed
11	8.098976056	8.140202018	6.244948989	9.714103895	1	Confirmed
12	13.5033133	13.53778147	12.00107755	14.80026741	1	Confirmed
13	8.352010786	8.336130196	6.881738209	9.504118438	1	Confirmed
14	7.917124394	7.867963667	6.788852362	9.016262071	1	Confirmed
15	12.71933976	12.63822502	11.38235922	14.82353995	1	Confirmed
16	14.0811054	14.1057367	12.29697066	15.44975835	1	Confirmed
17	3.420818316	3.249837242	1.067315725	5.001732766	0.909090909	Confirmed
18	11.48468289	11.42866842	9.88319948	12.88816424	1	Confirmed
19	10.46986209	10.45011664	8.982997282	12.13038177	1	Confirmed
20	10.01915779	10.07753625	8.585376088	11.01018682	1	Confirmed
21	1.676608	1.867675	1.0010015	2.186250	0.3000000	Rejected
22	1.915084553	1.910700479	0	3.000451001	0.904343434	Confirmed
23	6.21510273	6.231228097	4.673973975	8.216953805	1	Confirmed
24	6.635320355	6.664036864	5.234527214	8.092351127	1	Confirmed

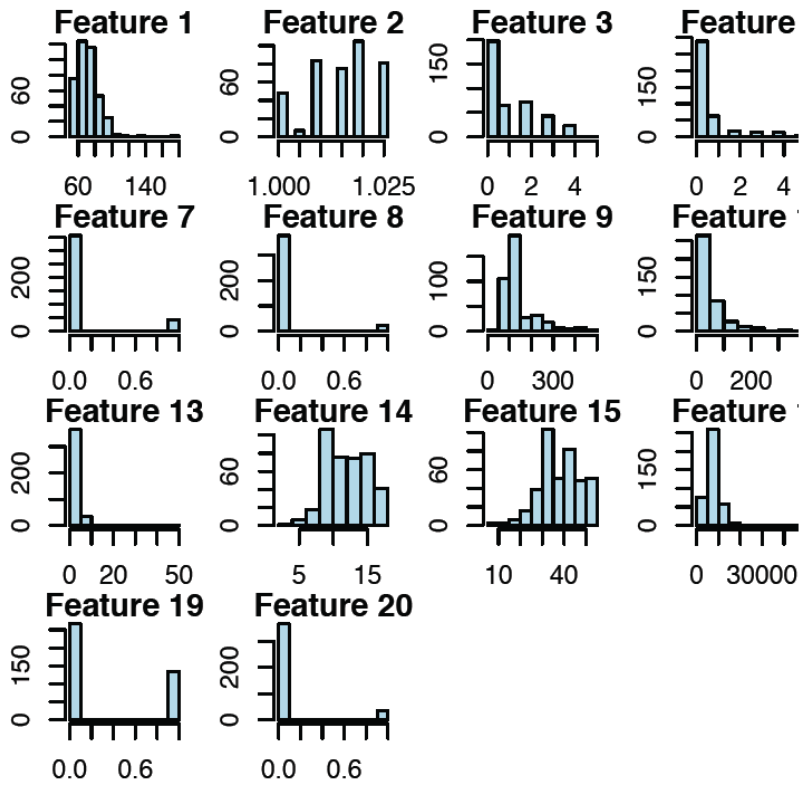

Figure 5: Feature Normal Distribution PLOT.

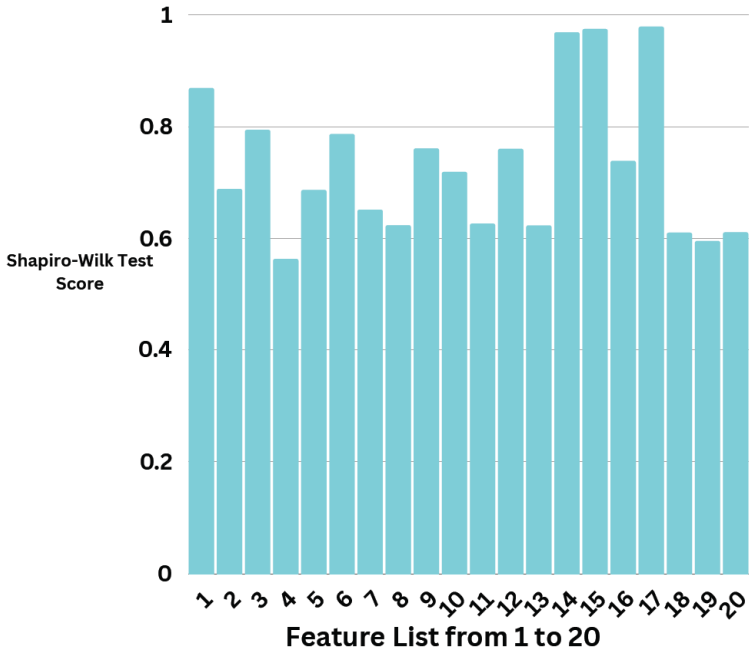

Figure 6: Shapiro-Wilk Test Score.

The next critical step is to validate the features selected by the hybrid feature selection method. To ensure that the selected features have a normal distribution, we considered the Shapiro-Wilk statistical test [22, 23]. The features found by the hybrid selection technique are confirmed to be reliable through this validation process. Using the Shapiro-Wilk statistical test, we found that the BP test score was 0.8691, SG was 0.6888, AL was 0.7944, SU was 0.5636, RBC was 0.6869, PC was 0.7869, BGR was 0.6516, BU was 0.6238, SC was 0.7612, SOD was 0.7192, POT was 0.6267, HEMO was 0.7605, PCV was 0.6235, WC was

Table 3: Result Comparison

Models	Feature-Selection S	Accuracy	Precisions	Recall	F-score
Support Vector Machine	No	93	0.8	0.9	0.9
Logistic Regression	No	92	0.8	0.9	0.8
Random Forest	No	94	0.9	0.8	0.9
Proposed Ensemble Deep learning	No	96	0.8	0.9	0.9
Support Vector Machine	Proposed Hybrid Feature selection	95	0.8	0.9	1
Logistic Regression	Proposed Hybrid Feature selection	94	0.8	0.8	1
Random Forest	Proposed Hybrid Feature selection	95.70	0.9	0.8	1
Proposed Ensemble Deep learning	**Proposed Hybrid Feature selection**	97.50	0.9	1	1

0.9687, RC was 0.9747, HTN was 0.7389, DM was 0.9789, APPET was 0.6105, PE was 0.5956 and ANE was 0.6112. The 20 features selected through the hybrid feature selection method were validated using the Shapiro-Wilk statistical test. Results confirm that all features are normally distributed, as shown in Figure **6**.

An ensemble deep learning approach was used to classify CKD, utilizing a deep neural network with varying layer configurations. Hyperparameters such as neuron count, activation function, and learning algorithm were optimized for accuracy. The model was evaluated using accuracy with an 80-20 train-test split. Results, shown in Table **3**, were compared to Support Vector Machine, Logistic Regression, and Random Forest, with and without feature selection.

5. CONCLUSION

This work proposesd a novel approach for categorizing Chronic Kidney Disease (CKD) patients by employing a fusion of Boruta and Recursive Feature Elimination (RFE) techniques alongside deep learning methodologies. The objective was to enhance the accuracy and interpretability of the classification model by meticulous feature selection and the implementation of deep learning methodologies. The proposed study employs a hybrid feature selection method to identify highly predictive features for Chronic Kidney Disease. The effectiveness of the selected features was validated using the Shapiro-Wilk test, confirming that all 20 features are normally distributed. Additionally, an ensemble-based deep learning method was proposed for CKD classification. The proposed ensemble-based deep learning method outperforms SVM LR and RF in classifying CKD. This enhanced the model's capacity for pattern extraction and led to greater classification accuracy. When applied to the chosen features, the deep learning model demonstrated its ability to autonomously discover intricate patterns and relationships within the CKD data. Based on experimental results, we conclude that these features, age, pus cell clumps, bacteria, and coronary artery disease do not contribute much to accurate classification tasks. In the medical field, the proposed method's interpretability was essential. The approach provided clarity and helped explain prediction outcomes by carefully selecting key variables for CKD classification.

ACKNOWLEDGEMENTS

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

[1] Shakhshir MH, Vanoh D, Hassan M, Zyoud SH. Mapping the global research landscape on malnutrition for patients with chronic kidney disease: a visualization analysis. Journal of Health Population and Nutrition 2023; 42(1): 101. https://doi.org/10.1186/s41043-023-00445-8

[2] Islam MA, Majumder MZH, Hussein MA. Chronic kidney disease prediction based on machine learning algorithms. Journal of Pathology Informatics 2023; 14: 100189. https://doi.org/10.1016/j.jpi.2023.100189

[3] Levey A, Atkins R, Coresh J, Cohen E, Collins A, Eckardt K-U, *et al*. Chronic kidney disease as a global public health problem: approaches and initiatives-a position statement from kidney disease improving global outcomes. Kidney International 2007; 72(3): 247-259. https://doi.org/10.1038/sj.ki.5002343

[4] Lv J-C, Zhang L-X. Prevalence and disease burden of chronic kidney disease. Renal fibrosis: mechanisms and therapies 2019; 3-15. https://doi.org/10.1007/978-981-13-8871-2_1

[5] Tannor EK, Sarfo FS, Mobula LM, Sarfo-Kantanka O, Adu-Gyamfi R, Plange-Rhule J. Prevalence and predictors of chronic kidney disease among ghanaian patients with hypertension and diabetes mellitus: A multicenter cross-sectional study. The Journal of Clinical Hypertension 2019; 21(10): 1542-1550. https://doi.org/10.1111/jch.13672

[6] Bai Q, Su C, Tang W, Li Y. Machine learning to predict end stage kidney disease in chronic kidney disease. Scientific Reports 2022; 12(1): 8377. https://doi.org/10.1038/s41598-022-12316-z

[7] Sawhney R, Malik A, Sharma S, Narayan V. A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. Decision Analytics Journal 2023; 6: 100169. https://doi.org/10.1016/j.dajour.2023.100169

[8] Harimoorthy K, Thangavelu M. Multi-disease prediction model using improved svm-radial bias technique in healthcare monitoring system. Journal of Ambient Intelligence and Humanized Computing 2021; 12: 3715-3723. https://doi.org/10.1007/s12652-019-01652-0

[9] Fahimifar S, Mousavi K, Mozaffari F, Ausloos M. Identification of the most important external features of highly cited scholarly papers through 3 (ie, ridge, lasso, and boruta) feature selection data mining methods: Identification of the most important external features of highly cited scholarly papers through 3 (ie, ridge, lasso, and boruta) feature selection data mining methods. Quality & Quantity 2023; 57(4): 3685-3712. https://doi.org/10.1007/s11135-022-01480-z

[10] Habibi A, Delavar MR, Sadeghian MS, Nazari B, Pirasteh S. A hybrid of ensemble machine learning models with rfe and boruta wrapper-based algorithms for flash flood susceptibility assessment. International Journal of Applied Earth Observation and Geoinformation 2023; 122: 103401. https://doi.org/10.1016/j.jag.2023.103401

[11] Tang R, Zhang X. Cart decision tree combined with boruta feature selection for medical data classification. in 2020 5th IEEE International Conference on Big Data Analytics (ICBDA) IEEE 2020; 80-84. https://doi.org/10.1109/ICBDA49040.2020.9101199

[12] Zhang B, Dong X, Hu Y, Jiang X, Li G. Classification and prediction of spinal disease based on the smote-rfe-xgboost model. Peer J Computer Science 2023; 9: e1280. https://doi.org/10.7717/peerj-cs.1280

[13] Ebiaredoh-Mienye SA, Swart TG, Esenogho E, Mienye ID. A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. Bioengineering 2022; 9(8): 350. https://doi.org/10.3390/bioengineering9080350

[14] Ferguson T, Ravani P, Sood MM, Clarke A, Komenda P, Rigatto C, Tangri N. Development and external validation of a machine learning model for progression of ckd. Kidney International Reports 2022; 7(8): 1772-1781. https://doi.org/10.1016/j.ekir.2022.05.004

[15] Lee J, Warner E, Shaikhouni S, Bitzer M, Kretzler M, Gipson D, *et al*. Unsupervised machine learning for identifying important visual features through bag-of-words using histopathology data from chronic kidney disease. Scientific Reports 2022; 12(1): 4832. https://doi.org/10.1038/s41598-022-08974-8

[16] Cui R, Liu M, Initiative ADN, *et al*. Rnn-based longitudinal analysis for diagnosis of alzheimer's disease. Computerized Medical Imaging and Graphics 2019; 73: 1-10. https://doi.org/10.1016/j.compmedimag.2019.01.005

[17] Shi X, Wang T, Wang L, Liu H, Yan N. Hybrid convolutional recurrent neural networks outperform cnn and rnn in task-state eeg detection for parkinson's disease. in 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE 2019; 939-944. https://doi.org/10.1109/APSIPAASC47483.2019.9023190

[18] Ghosh P, Shamrat FJM, Shultana S, Afrin S, Anjum AA, Khan AA. 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; 1-6. https://doi.org/10.1109/iSAI-NLP51646.2020.9376787

[19] Aljaaf AJ, Al-Jumeily D, Haglan HM, Alloghani M, Baker T, Hussain AJ, Mustafina J. Early prediction of chronic kidney disease using machine learning supported by predictive analytics. in 2018 IEEE congress on evolutionary computation (CEC), IEEE 2018; 1-9. https://doi.org/10.1109/CEC.2018.8477876

[20] Alassaf RA, Alsulaim KA, Alroomi NY, Alsharif NS, Aljubeir MF, Olatunji SO, *et al*. Preemptive diagnosis of chronic kidney disease using machine learning techniques. in 2018 international conference on innovations in information technology (IIT), IEEE 2018; 99-104. https://doi.org/10.1109/INNOVATIONS.2018.8606040

[21] Mezzatesta S, Torino C, De Meo P, Fiumara G, Vilasi A. A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. Computer Methods and Programs in Biomedicine 2019; 177: 9-15. https://doi.org/10.1016/j.cmpb.2019.05.005

[22] Schucany WR, Tony Ng H. Preliminary goodness-of-fit tests for normality do not validate the one-sample student t. Communications in Statistics-Theory and Methods 2006; 35(12): 2275-2286. https://doi.org/10.1080/03610920600853308

[23] Pacheco F, de Oliveira JV, Sanchez R-V, Cerrada M, Cabrera D, Li C, Zurita G, Artes M. A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions. Neurocomputing 2016; 194: 192-206. https://doi.org/10.1016/j.neucom.2016.02.028

Received on 06-01-2025 Accepted on 13-02-2025 Published on 03-03-2025

https://doi.org/10.6000/1929-6029.2025.14.11