Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 4
April 2024

Integrative Feature Selection and Enhanced Classification of Chronic Kidney Disease through Novel Weight Convolutional Neural Network Fusion

[1]Mr.P.Sivaraman , [2]Dr. Ritu Sindhu
Research Scholar, Department of CSE.
Lingaya's Vidyapeeth
Faridabad, Haryana, 121002, India.
Professor, Department of CSE.
Lingaya's Vidyapeeth
Faridabad, Haryana, 121002, India.

Abstract
Introduction: Chronic Kidney Disease (CKD) is a progressive condition marked by a progressive decline of renal function over time, resulting in complications like electrolyte imbalances and cardiovascular issue.
Objectives: Elucidating the complex ethology and global impact of CKD emphasizes the necessity for early detection and holistic management strategies to alleviate its far-reaching public health burden.
Methods: This study explores three primary feature selection techniques: Embedded, Wrapper, and Filter techniques in machine learning. Specifically, Recursive Feature Elimination with Cross-Validation (RFECV) is applied as the Wrapper method, aimed at developing a robust machine learning model by iteratively discarding irrelevant features. The research focuses on identifying and utilizing the most relevant features for training the model, enhancing its predictive performance by eliminating redundant or less informative attributes. This approach aims to optimize model accuracy and efficiency by iteratively selecting the most impactful features from the dataset. This research introduces the NWCNN as a novel classification method. The NWCNN approach is enhanced by integrating the LGWO technique, aiming to optimize both computation and classifier performance.
Results: The suggested model for the CKD dataset is planned for implementation in the Python tool. When compared to existing methods, the proposed model demonstrates a significantly higher accuracy. The proposed model for chronic kidney disease (CKD) classification achieved exceptional performance metrics, including an accuracy of 99.86%, precision of 98.70%, recall of 100%, F1 score of 99%, specificity of 100%, Matthews Correlation Coefficient (MCC) of 97.80%, and negative predictive value (NPV) of 99.50%.
Conclusions: The research explores the feature selection methods in machine learning, such as Filter, Wrapper, and Embedded methods, with emphasis on RFECV as the Wrapper method generally aims to improve the ability of ML models to predict CKD has been improved by selecting the most appropriate quality. The classification technique is further advanced with the introduction of the NWCNN and the incorporation of LGWO. By optimizing the computation and classification using LGWO, the NWCNN method seeks to increase efficiency and accuracy.

Keywords: Chronic Kidney Disease, Recursive Feature Elimination with Cross-Validation, Novel Weight Convolutional Neural Network, Lion Grey Wolf Optimization

1. Introduction

Demand for end-stage renal disease therapy is growing rapidly due to the rising global incidence of CKD and related risk factors (ESKD) [1]. CKD is the progressive loss of kidney function in bloodstream filtering and waste product removal. This illness affects more than 10% of people worldwide and 15% of people in South Africa [2]. Both inpatient and outpatient pneumonia are more common in patients with CKD. Furthermore, CKD patients appear to have a mortality rate from pneumonia that is many times greater than that of the general population [3]. Low-grade persistent inflammation is increasingly recognized as a key feature of CKD. Even with the significant advancements in the field of healthcare, CKD remains a serious health issue that affects 10-15% of the population and is becoming more and more common. Due to its delicate nature, early detection of CKD is uncommon [4]. Patients with CKD have immune dysfunction shortly after the course of renal insufficiency, regardless of the underlying ailment [5].

The term CKD describes a gradual and irreversible loss of the kidneys' structure and functions, particularly the glomerular filtration rate, which deteriorates over a while [6]. Approximately 10% of people globally have CKD and millions of them—more often the elderly—die each year from the lack of access to inexpensive treatment. The International Society of Nephrology's 2010 research, "Global Burden Disease," states that CKD has become a major global cause of death, with 82.3% more fatalities from the condition in the previous 20 years [7]. CKD is likewise referred to as chronic kidney failure or chronic renal failure. It is an illness that affects many people but is often overlooked until it reaches a severe stage. In the human body, the kidney plays a crucial role in the egesting and elimination process, which removes all harmful and unnecessary substances-typically wastes-from the body [8]. In 2016, 417 million women and 336 million men worldwide were afflicted with CKD. Every year, more than a million individuals in 112 developing nations pass away from renal insufficiency due to their inability to pay for the extremely high cost of kidney replacement surgery or routine dialysis [9] [10].

Chronic renal disease is mostly caused by two factors: excessive blood pressure and diabetes. High blood sugar levels are a hallmark of diabetes, which affects the kidneys, heart, blood vessels, and eyes. Furthermore, heart attack, stroke, and chronic renal disease can all be directly linked to inadequate management of high blood pressure [11]. Typically, CKD remains silent until significant renal damage occurs. According to research, the prevalence of CKD is rising quickly, with an annual rise in hospitalization cases of 6.23%, while the worldwide death rate stays constant. This issue affects a lot of individuals in Bangladesh. Since limited senses and a lack of broad awareness are the primary causes of CKD, most individuals in rural regions are unaware of it [12]. Patients with CKD exhibit severe immunological and neurological system dysfunctions that significantly impair their quality of life and interfere with several everyday activities [13]. Because CKD raises the possibility of unfavorable clinical consequences, it is a significant global public health issue. The prognosis for individual individuals is currently little understood, although it is widely known that CKD is independently linked to elevated risks for end-stage renal disease, cardiovascular events, and all-cause mortality [14].

2. Objectives

In order to increase the critical need for better diagnosis accuracy in CKD, this work explores sophisticated feature selection approaches and presents a unique classification strategy. RFECV is a machine learning approach that repeatedly eliminates unnecessary features, improving model resilience by analysing Embedded, Wrapper, and Filter strategies. The goal of the research is to find and use the most significant aspects to improve prediction performance and get rid of redundant information. In order to maximize both computational efficiency and classifier precision, the paper also presents the NWCNN approach, enhanced with LGWO. By optimizing neural network weights through the use of LGWO, the NWCNN technique performs better.

3. Methods

The proposed methodology shows comprehensive feature selection techniques with an innovative classification approach to enhance CKD prediction. Initially, the study employs Filter, Wrapper, and Embedded feature selection methods in machine learning to identify pertinent features from the dataset, focusing on RFECV to iteratively eliminate irrelevant attributes and optimize model robustness. Subsequently, the research introduces the NWCNN as a novel classification method. This methodological innovation combines the NWCNN architecture with LGWO to optimize neural network weights and improve computational efficiency and classifier performance. The workflow entails iterative feature selection to craft a refined dataset, followed by NWCNN-LGWO integration for accurate and efficient CKD prediction, offering a comprehensive approach for enhanced disease detection and management. Figure 1 shows the flow diagram of CKD.

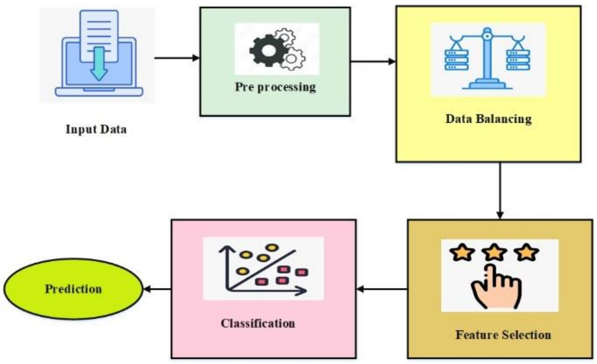


Figure 1: Flow Diagram of Chronic Kidney Disease

4.1 Dataset Description

The CKD data set utilized in this investigation came from the machine learning library on Kaggle. Each data set comprises 400 observations. With the CKD data set mentioned earlier, each model has 24 predicted characteristics traits. Consisting of eleven numerical variables and thirteen category (nominal) variables in addition to a categorical response variable (class). For every class, there are 2 numbers: CKD (sample with CKD) and notched (sample without CKD). the 400 samples, 250 are classified as CKD samples and 150 as notched samples. It's crucial to recognize there are a lot of misplaced numbers in the data. Here 20% for testing data and 80% for training data.

4.2. Pre-processing

All of the category (nominal) variables were coded to make computer processing easier. Present and absent values for PCC and ba were recorded as 1 and 0, respectively. Representing the given numbers of htn, dm, cad, pe, and ane, the corresponding numbers for yes and no were 1 and 0, correspondingly. Regarding the applet value, good and bad were recorded as 1 and 0, correspondingly. Three variables-sg, al, and su- were regarded as numeric variables even though their values are still dependent on numbers, although the original data description classifies them as categorical kinds. Each category variable was converted into a factor. A separate number between one and four hundred was assigned to each sample. There are only 158 complete instances in the data collection, meaning that there are a lot of missing variables. Overall, patients may choose to forego some tests for some reasons before receiving a diagnosis. Consequently, missing values will appear in the data when the diagnostic categories of the specimens are unknown, requiring the adoption of an appropriate imputation strategy. After the categorical categories were encoded, missing values in the initial CKD data set were addressed and filled upwards. The values that are missing for each of the numerical components are filled by using the average of the variable in question in K full examples, whereas the category variable with the highest frequency in the corresponding variable in K complete samples is used to fill in the values that are not present. Individuals sharing similar physical conditions are expected to exhibit comparable physiological data. This rationale underlies the utilization of K-nearest neighbors (KNN) techniques to address missing values in metabolic measurements. For example, physiological measurements in healthy persons should remain constant within a given range. When comparing the physiological data of individuals with similar degrees of a certain illness, identical results should be obtained. Specifically, for individuals in comparable circumstances, there shouldn't be significant variations between the physiological measurement results. As this approach has been used in the field of hyperuricemia, it should also be modified for the diagnostic data of other disorders. It is better to choose K as a number that is even when determining the median of associated factors K complete models since, in this instance, the middle number is automatically median when outcomes numerical variables in K full numerical values are used to sort the samples. K shouldn't be chosen in an excessively huge or tiny manner. A too-big K value can overlook the potentially significant unobtrusive mode. Conversely, noise is produced by an excessively tiny K value, which also has a major effect on filling in the missing data [20].

4.3. Under sampling for Data Balancing

There are very few examples that deviate from the median, with the majority of the data being concentrated close to it. An ML model will perform worse in areas with fewer data and fit better in the region that has additional samples. Put another way, in situations when there are fewer samples, the forecast accuracy confidence interval will be larger. Imbalanced data is not a concern for ordinary regression tasks. Classify samples as outliers if there are fewer samples in an interval. For this work, though, the reverse is true. Our goal is to accurately forecast individuals with high creatinine values, even if the data for that range is not very extensive. When developing a model for machine learning using the entire dataset, examples with high creatinine values where there is less data will be disregarded since the algorithm used for training will attempt to reduce the sum error for the entire dataset.

To demonstrate the effect of the imbalance issue on regression more clearly, a straightforward experiment was conducted utilizing XGBoost and the Minimum Square Error (MSE) as the loss function. Found that when there were fewer training examples, the model was unable to make predictions. Generally speaking, the issue arises from the fact that the machine learning model cannot learn from a sufficient number of samples that have a high creating value. Both data element and model level approaches may often be used to handle the data imbalance issue. The most used data-level approach is oversampling. Nevertheless, oversampling is not appropriate for these data. Oversampling is the process of creating new data from a small amount of existing data by resampling or creating new data. Elevated creatinine levels are quite low in this situation. Noise in infrequent data will be increased several times again due to oversampling. To address this issue, suggested a cost-sensitive MSE loss function and the under-sampling approach [21].

4.4. Feature Selection

Each machine learning classifier that has been trained requires the outcome might be impacted if the superfluous feature is left in the dataset. Higher efficiency and a reduction in the model's time of execution are provided by the feature selection classifier algorithm. In this study, three distinct feature selection techniques were applied.

4.4.1. Filter Method

One way to choose the right feature is to use the filter. Despite using a learning classifier method, it chooses the features based on their integral characteristics. When compared to the wrapper technique, this approach yields speedier results. Every characteristic is given a score by the approach based on the statistical correlation between those attributes. Although there are other filtering techniques, the Correlation-based Feature Selection (CFS) approach is applied. The feature selection mechanism used in CFS is based on attribute rankings. Based on the correlation heuristic assessment function, the rank of the attribute subset is assigned. The function operates using an approach that generates 2 class labels: a low-correlated class and a correlated class. It then only chooses characteristics from the correlated label class.

4.4.2. Wrapper Method

The wrapper technique uses an exact machine learning algorithm to choose the subset of characteristics. It looked for a potential subset of characteristics using the greedy search technique. One can employ various algorithms such as forward selection, backward elimination, and recursive elimination for feature selection -which can be used to carry out the procedure. We employed the forward feature selection strategy in the study. The feature is chosen repeatedly by the forward feature selection method. This process begins with the null model and adds the attribute iteratively at each stage. Up until the characteristic fails to enhance model performance, it is continuously added to the model.

4.4.3. Embedded Method

The choice tree algorithm is the integrated approach for choosing features. When the tree grows, it picks a characteristic in each step and divides the sample set into smaller subsets repeatedly. The ID3, C4.5, and CART decision tree algorithms are the most often used ones. Other approaches to the creation of linear models are accessible. The most often used techniques are Ridge with an L2 penalty and LASSO [30] with an L1 penalty. In this work, the least absolute shrinkage and selection operator, or LASSO, approach has been used Regularization and feature selection are its two primary functions. Certain feature coefficients are reduced to zero during regularization, indicating that the characteristics are not crucial to the prediction model [22].

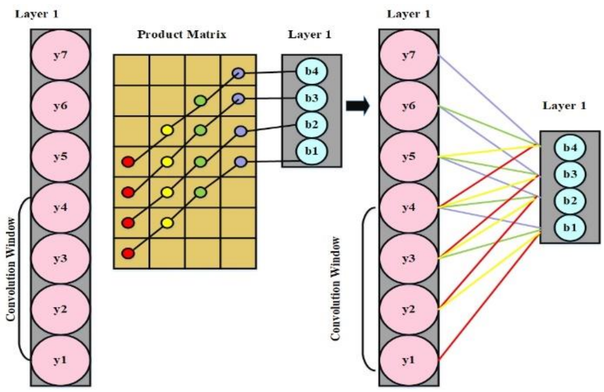
Due to its speed, efficiency, and scalability, the univariate feature selection approach from filter methods has been chosen for this investigation. The wrapper feature selection approach has been replaced by Recursive Feature Elimination with Cross-Validation (RFECV).

4.4.4. Recursive Feature Elimination with Cross-Validation (RFECV):

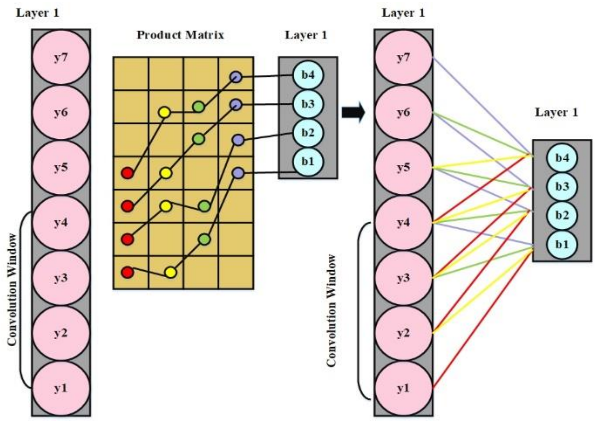
An optimization approach that constantly removes unimportant information to create a trained machine-learning model with relevant and chosen features. To choose the best subset of features, it repeatedly constructs the model, sets aside the feature that performs the poorest on each iteration, and then produces the next model using the remaining features. To enhance the model's generalization achievement, it removes the redundant and weak features whose removal has the least impact on training and retains the independent and strong features. To determine which traits have been deemed most significant, this approach ranks them using an iterative process. This method ranks each feature based on its relevance and constructs the model on the complete collection of features since it interacts with a machine learning model [23].

4.5. Innovative Weight Convolution Neural Network (NWCNN) Classification Model

When diagnosing illnesses, diagnostic specimens are distributed throughout multidimensional contexts, and data (CKD or notched) is classified using predictors. These places' data samples are organized into categories, which causes two categories to be divided and samples within categories to converge. The recommended method makes use of dynamic systems to determine the optimal convolution layer weight alignments in CNN. Overall classifier's weights for relevant qualities are then computed as positive guidance, and therefore, the IG of capabilities, based on IG (Information Gain). Classifier weights are determined by the value range of IG; higher values lead to higher weights; and smaller values result in lower weights. To reduce differences between comparable qualities of weights and input weights incoming window standards inside a convolution filter are continuously integrated, as opposed to being linear inner outputs of convolutions. Figure 2 shows a comparison between traditional linear convolution methods.



A. Convolution with Linear Weight Alignment



B. Convolution with Dynamic Weight Alignment

Figure 2: Comparison between Traditional Linear Convolution Methods [24].

Layer p is the feature map from the complexity that comes after layer p − 1, which has elements b1,….., b4, and y1,….., y7, as its elements. The blue circle shows the total of items, and each dot on it symbolizes a product with the appropriate weight and contribution. CNNs have more convolutional layers and are ANNs. The main structures of convolution layers are their use of parameter sharing and sparse connectivity. More specifically, each connected output item's local receptive field is shared by all of the compression layer's weights. This makes forward-looking calculation of the convolution layer similar to the filtering convolution method using joint weights to provide feature maps. The feature map b j((p)) of a layer is formalized as,

$$b_j^{(p)} = \sum_{i=0}^{p-1} c_i^{(p)} y_{r+(Q-1)}^{(p-1)} + z^{(p)} \quad (1)$$

For element j individually, p denotes the convolutional layer, p− 1 denotes the previous layer, q denotes the filter index, and I is the window size. Write$c_i^{(p)}$,$y_{q+j}^{(p-1)}$ and $b^{(p)}$ to represent the split weights, bias, and activations from the preceding layer, respectively. To clarify, $b_j^{(p)}$ represents the internal factor of the weights that are shared $c^{(p)}$ and every window of the layer that came before it, namely$y_j^{(p-1)}$,…, $y_{j+(Q-1)}^{(p-1)}$. The weights and inputs inside the window are linearly matched by this inner product. For instance, certain weights might need to harmonize with more ideal efforts when handling noisy elements, version of features, or scale variation inside the filter. A similarity function and the standard innermost product of a convolution function are somewhat alike. The primary objective is to align the weights such that windows that are similar but slightly out of agreement are triggered more intensely.

To optimize the weight alignment, Information Gain (IG) is used in this study. The quantity of knowledge that a certain variable or trait communicates about the outcome is measured as information gain. It may be calculated by subtracting the entropy of a feature derived from the total dataset's entropy.

$$Gain\ (Y,G) = T(G) - \sum_{j=1}^{f} \frac{|G_p|}{|G|} . T(G_J) = T(G) - T(Y,G) \quad (2)$$

$T(G)$ is the total dataset's entropy. G, $|G_J|$-the number of instances of an attribute Y with j value, $|G\ \ |$-the total number of examples in a dataset, the collection of distinct values of an attribute Y, $T(Y,G)$ is the entropy of the attribute Y subdivision of cases. $T(Y,G)$is the property Y's unpredictability. To compute information improvement, one must first compute entropy. Let's review the equation for entropy.

$$T(G) = -\sum_{i=1}^{N} L_i L_i \quad (3)$$

Here, the number of distinct class standards is N. Furthermore, the final answer will be either yes or no. There are thereafter two different class values. L_i stands for the probability of an occurrence.

The methodology is facilitated by Lion with Grey Wolf Optimization (LGWO), an innovative approach enhancing both computational efficiency and classifier effectiveness. LGWO, inspired by the social behavior of lions and grey wolves, amalgamates their characteristics to optimize classifier performance. Leveraging this optimization technique augments the learning process by efficiently fine-tuning model parameters, promoting faster convergence, and enabling classifiers to attain improved accuracy. The collaborative nature of LGWO harnesses the prowess of both lion and grey wolf behaviors, fostering a synergy that refines the classifiers' decision boundaries and strengthens their predictive capabilities. Through this approach, LGWO optimizes the classifiers' efficiency, ultimately elevating their performance to deliver more robust and accurate predictions.

4.5.1. Lion Grey Wolf optimization (LGWO)

The technique known as LGWO is utilized to minimize faults in training data and identify any harmful segments. Here, the alpha leader either male or female for hunting, sleeping, waking times, and location. In the grey wolves' pecking order, Alpha makes the decisions, and Beta takes care of second grade. Little leftovers were allowed to be magically arranged after the leader wolf' feast. Grey wolves reside in packs[25]. A wolf is classified as secondary or delta, and it appears when the assembly of an alpha (α), beta (β), or omega (ω) is insufficient. This is why the betas (β) and alphas (α) inspire adoration in the delta (δ) wolves. They grow more proportionately than omegas. In this procedure, a candidate determines that the alpha is the best choice for logically reproducing the pack hierarchy by visualizing the different kinds of wolves. Therefore, beta (β) and delta (δ) are the second and third most suitable provisions, respectively. One notices that omega (ω) is the remaining provision. The weight values (w), learning rate (α), number of layers (L), kernel sizes (k), and dropout rate (γ) are among the parameters used in Lion optimization [26].

Initialization Process

The output data is initially initialized with the coefficient vectors y, z, and w.

Fitness evaluation

assesses the fitness utility and forecasts the outcome

$$Fit_i = \max accuracy \quad (4)$$

Let starting fitness be x_α, next finest fitness be x_β, final finest fitness be x_δ

a) Roaming

Every male lion in the procedure visits the region, with a total location of the number of this location of %G. During its traveling, the lion modifies its location when it discovers the ideal spot. The improved location of the Lion is indicated below [27].

$$\gamma \sim E(0,2) \times s \quad (5)$$

s – The position of the male lion right now and the spot he chose inside the territory

E – Even Distribution

$$L_{si} = 0.1 + min\left(0.5, \frac{(A_i - A_{best})}{A}\right), \ i=1,\ 2,\ 3\ \dots \text{ number of nomad lions} \quad (6)$$

$S_{Gi} =$ the possibility that all nomadic lions will split apart

$A_i\ = i_{th}$ fitness value of Lion's

$A_{best} =$ the best nomad lion fit

b) Encircling prey

In addition to these three competitors, α, β, δ, and ω also track each other. The gang uses a prey's position to follow them in addiction. The following formulas are used to calculate the grey wolf's encircling or trapping behavior during hunting [28].

$$x(h+1) = x_{lg}(h) = \vec{Z}.\vec{K} \quad (7)$$

$$\vec{K} = \left|\vec{W}.x_{lg}(h+1) - x_{lg}(h)\right| \quad (8)$$

$$\vec{Z} = 2\vec{w}j_1 - \vec{w} \ \text{ And } \ \vec{W} = 2j_2 \quad (9)$$

$d_{lg}\ (h)$ - The posture of the prey

Z and W – The vector of coefficients

$\vec{w}$ - reduced linearly from 2 to 0

$j_1 and j_2$- Vector at random [0,1]

h - The number of iterations, n

c) Hunting

Lion Optimization is the basis for hunting operations. To feed their pride, the female in each pride focuses on a group of prey. It uses certain techniques to envelop the victim and maintain pride. Every lion adjusts its surroundings to depend on a certain area and the group associates areas. Because of this, the attackers encircle the victim and use their specialized knowledge to commit local conflict attacks. The searchers are divided into three categories as a result. The two wings refer to the other groups, and the centroid, or peak fitness of associates with the crowd, is the member of the other group. One prey is involved in the centre of seekers. When hunting, the hunters are selected in order, and the group that selects the lion as its member then attacks a fictitious victim in turn. A fake victim may escape from a searcher if it increases its fitness, and a new location of the prey may be found [29]

$$x_{lg}' = l + ran(0,1) \times LQ \times (l - t) \quad (10)$$

The victim's present location. Hunter is a symbol for the current location seeking who has struck a victim, whereas LQ is a symbol for the growing number of objective searchers. To replicate the chosen seeker crowds around the victim, the new locations of the hunters' fit in both the left and right wings are provided.

$$t' = \begin{cases} rand\big((2 \times l - t), l\big) \; if \; (2 \times l - t) < lg \\ rand\big(l, (2 \times l - t)\big) \; if \; (2 \times l - t) > lg \end{cases} \quad (11)$$

The updated locations for centroid seekers are provided.

$$t' = \begin{cases} rand(t, l) \; if \; (2 \times l - t) < lg \\ rand(l, h) \; if \; (2 \times l - t) > lg \end{cases}$$

(12)

An arbitrary value between z and w, or the upper and lower bounds, is obtained using the aforementioned equation.

4. Results

To evaluate the results and forecast renal illness, the dataset was first pre-processed by normalizing and removing outliers. A range of network optimization techniques were attempted to improve the model's effect and stability. Compute the accuracy, recall, precision, and F1-score indicators to assess these results.

$$\text{Accuracy}=\frac{TP+TN}{TP+FN+FP+TN} \quad (13)$$

$$\text{Recall=sentivity}=\frac{TP}{TP+FN} \quad (14)$$

$$\text{Specificity}=\frac{TN}{FP+TN} \quad (15)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad (16)$$

$$\text{F1 score}=2 \times \frac{precision \times recall}{precision+recall} \quad (17)$$

TP signifies the true positive, FP the false positive, TN the true negative, and FN the false negative.

Evaluate the proposed early-stage kidney disease prediction technique using the Python software in a simulated environment utilizing CKD datasets. An Intel(R) Core(TM) i5-3470-equipped computer is used to conduct the test. Additionally, the OS maker micro software 10 pro has inserted physical memory (RAM) 12.7 GM. Table 1 below shows the simulation parameters of the real-time dataset.

Table 1: Simulation parameters of CKD dataset.

Simulation parameters for CKD	Values
Batch_size	**20**
Epoch	**110**
Activation function	**sigmoid**

5. Discussion

Metrics for evaluations are essential to measure the accomplishment of classification. The most popular approach for achieving this goal is an accuracy evaluation. The percentage of test datasets that a classifier correctly categorizes may be used to calculate the classifier's accuracy for a particular dataset.

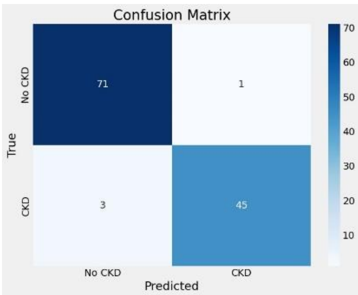


Figure 3: Confusion Matrix

The confusion matrix in Figure 3 presents how well our classification model discriminates between samples with and without CKD. Particularly noteworthy ratios of true positives to true negatives highlight how well the model detects cases of both CKD and non-CKD. Elevated true positives indicate CKD case sensitivity, indicating the model's capacity to identify real positive cases. In the meanwhile, the model's performance in properly detecting instances without CKD is shown by its specificity, which is shown by a significant count of true negatives. However, the occurrence of false positives and false negatives warrants careful examination and suggests areas for improvement to increase the overall accuracy and reliability of our CKD classification system.

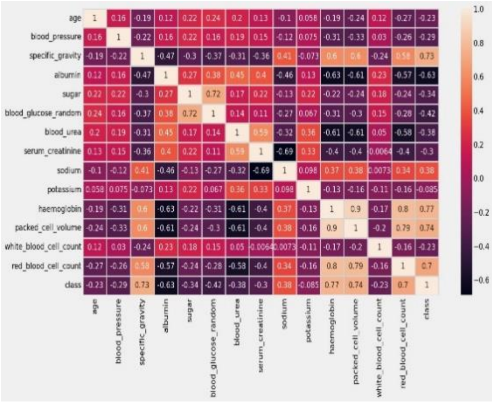


Figure 4: Correlation Matrix for CKD Dataset

In Figure 4, the Correlation Matrix for the CKD dataset illustrates interrelationships among variables. The high correlation between specific factors provides insight into possible factors involved in chronic kidney disease. This visualization facilitates the identification of important variables, and guides focused analysis to better understand the data.

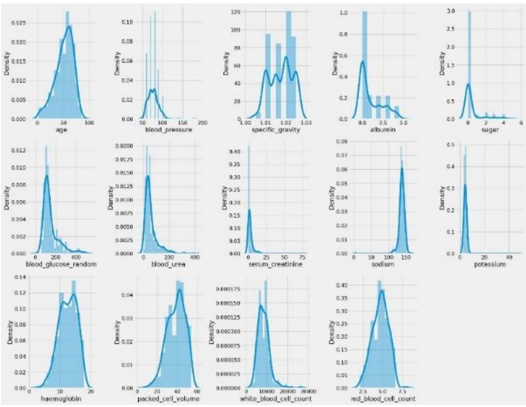
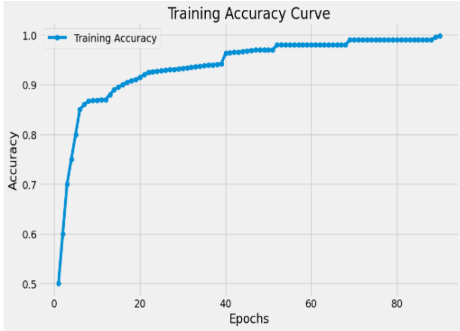
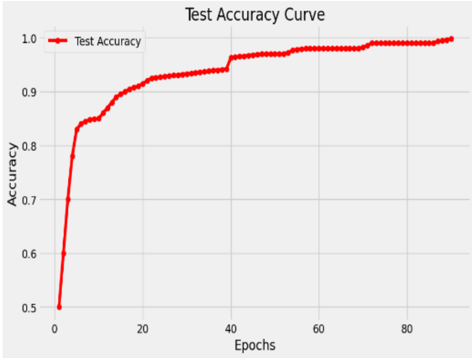


Figure 5: Density of CKD Dataset

A visual depiction of the distribution of data points across different characteristics may be seen in Figure 5, which shows the density of the CKD dataset. The density map provides information on the distribution and prevalence of chronic kidney disease in the dataset. This diagram is a useful tool for analyzing the classification data and helps to identify any patterns or abnormalities that may affect the course of CKD.



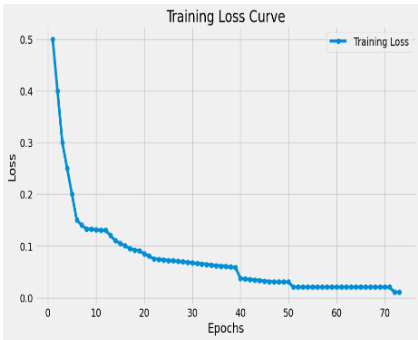
(a)



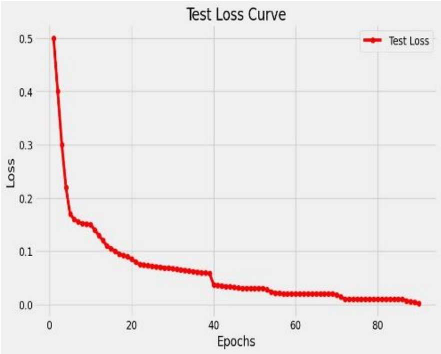
(b)

Figure 6: (a) Training Accuracy Curve (b) Testing Accuracy Curve

Subplots (a) Training Accuracy and (b) Testing Accuracy in Figure 6 provide important information about our model's performance. The accuracy attained during the training phase is displayed in Subplot (a), demonstrating the model's capacity to learn from the training data. In the meanwhile, subplot (b) offers a quick look at how well the model generalizes on test data that hasn't been observed before. This accuracy pattern enhances the learning outcome and predictive power of the model on the training and testing datasets, making it an important indicator for model evaluation.



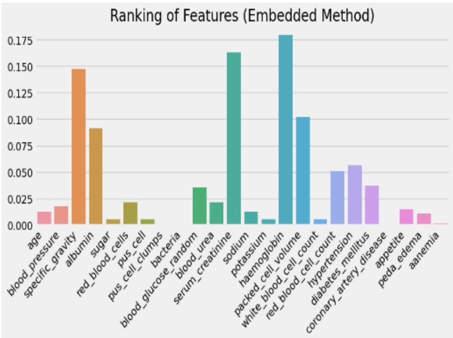
(a)



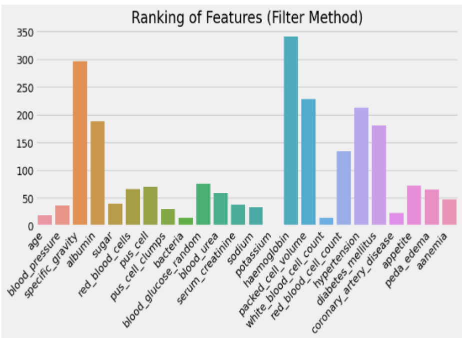
(b)

Figure 7: (a) Training Loss Curve (b) Testing Loss Curve

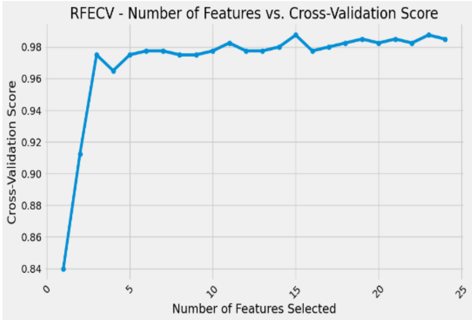
The related subplots in Figure 7, Training Loss Curve and Testing Loss Curve offer important insights into our model's learning dynamics. Subplot (a) illustrates the optimization process during model training by showing the decrease in training loss throughout epochs. Simultaneously, subplot (b) presents the testing loss, providing an evaluation of the model's generalization effectiveness on unknown data. These loss coefficients contribute to a more accurate knowledge of the overall performance of the model by acting as important indicators to evaluate the consistency of the model to reduce errors in the training and testing datasets.



(a)



(b)



(c)

Figure 8: Ranking of Features (a) Embedded Method (b) Filter Method (c)RFECV Vs Cross-Validation Score

In Figure 8, the subplots (a) Embedded Method, (b) Filter Method, and (c) RFECV vs Cross-Validation Score collectively illustrate the ranking of features using different techniques. Subplot (a) showcases the feature ranking through embedded methods, highlighting features deemed most relevant by the model during training. In subplot (b), the filter method ranks features based on statistical measures, providing insights into their contributions. Subplot (c) demonstrates the recursive feature elimination with cross-validation, offering a dynamic view of feature importance across different subsets. This multidimensional visualization helps to understand the concepts of priorities and helps to make informed decisions for the next modeling step.

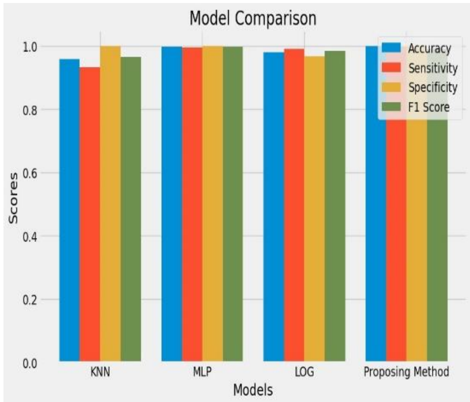


Figure 9: Comparison Graph

The Comparison Graph in Figure 10 compares the performance of the three currently available methods k-nearest neighbor (KNN) [30], MLP [31], and Logistic regression (LOG) [31] with our suggested technique, which achieves an impressive 99.86% accuracy. The illustration highlights the remarkable improvement provided by our proposed method, highlighting its superiority over traditional methods.

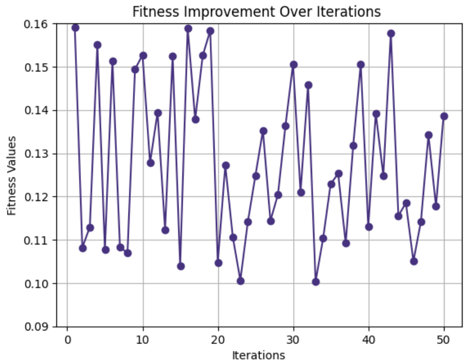


Figure 10: Fitness Improvement over Iteration Graph

The Fitness Improvement over Iteration is shown in Figure 10, which offers a dynamic perspective of the optimization process' development. The graph shows how the fitness measure improves over time due to the re-improvement of the model or method. This graphical representation provides insightful information about convergence behavior, which helps to monitor the effectiveness of the optimization process over time.

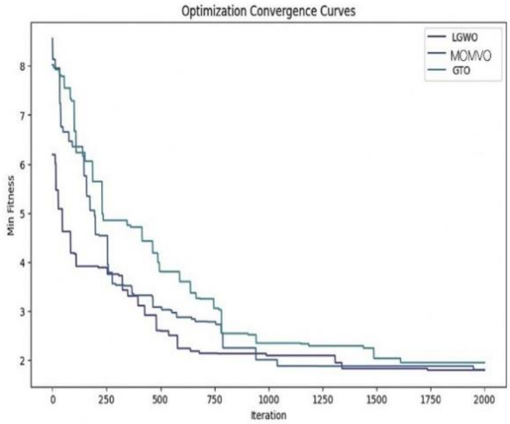


Figure 11: Convergence Curve for Optimization

The Convergence Curve for Optimization, shown in Figure 11, compares our suggested optimization strategy to other approaches already in use, including Artificial Gorilla Troops Optimizer (GTO) [32] and Multi-Objective Multi-Verse Optimizer (MOMVO) [33]. The illustration highlights the effectiveness of our scheme by revealing the fast convergence of optimal responses. This comparative study highlights the advantages of our proposed algorithm in terms of speed of assembly speed and effectiveness and provides insightful information about the performance of different optimization methods. Table 2 shows the performance parameters of the proposed method.

Table 2: Performance Metrics Table for Proposed Method

Performance	Values (%)
Accuracy	99.86
Precision	98.70
Recall	100
F1 Score	99
Specificity	100
MCC	97.80
NPV	99.50

Table 1 provides the detailed performance parameters of the proposed method, showing its unique efficiency in information classification. The model has an amazing accuracy of 99.86%, great precision (98.70%), and recall (100%), resulting in an amazing F1 Score of 99%. The suggested method's robustness in producing a balanced classification conclusion is highlighted by the Specificity and Matthews Correlation Coefficient (MCC), which reaches 100% and 97.80%, respectively. Furthermore, the model's accuracy in recognizing instances of interest is further shown by the Negative Predictive Value (NPV), which is 99.50%.

Refrences

[1] C.-W. Yang *et al.*, "Global case studies for chronic kidney disease/end-stage kidney disease care," *Kidney Int. Suppl.*, vol. 10, no. 1, pp. e24–e48, Mar. 2020, doi: 10.1016/j.kisu.2019.11.010.

[2] A. Ogunleye and Q.-G. Wang, "XGBoost Model for Chronic Kidney Disease Diagnosis," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 17, no. 6, pp. 2131–2140, Nov. 2020, doi: 10.1109/TCBB.2019.2911071.

[3] B. M. Henry and G. Lippi, "Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection," *Int. Urol. Nephrol.*, vol. 52, no. 6, pp. 1193–1194, Jun. 2020, doi: 10.1007/s11255-020-02451-9.

[4] M. Elhoseny, K. Shankar, and J. Uthayakumar, "Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease," *Sci. Rep.*, vol. 9, no. 1, p. 9583, Jul. 2019, doi: 10.1038/s41598-019-46074-2.

[5] M. Syed-Ahmed and M. Narayanan, "Immune Dysfunction and Risk of Infection in Chronic Kidney Disease," *Adv. Chronic Kidney Dis.*, vol. 26, no. 1, pp. 8–15, Jan. 2019, doi: 10.1053/j.ackd.2019.01.004.

[6] M. Nishat *et al.*, "A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms," *EAI*

Endorsed Trans. Pervasive Health Technol., p. 170671, Jul. 2018, doi: 10.4108/eai.13-8-2021.170671.

[7] M. Almasoud and T. E, "Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 8, 2019, doi: 10.14569/IJACSA.2019.0100813.

[8] B. Gudeti, S. Mishra, S. Malik, T. F. Fernandez, A. K. Tyagi, and S. Kumari, "A Novel Approach to Predict Chronic Kidney Disease using Machine Learning Algorithms," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India: IEEE, Nov. 2020, pp. 1630–1635. doi: 10.1109/ICECA49313.2020.9297392.

[9] W. Wang, G. Chakraborty, and B. Chakraborty, "Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm," *Appl. Sci.*, vol. 11, no. 1, p. 202, Dec. 2020, doi: 10.3390/app11010202.

[10] B. Bikbov, N. Perico, G. Remuzzi, and on behalf of the GBD Genitourinary Diseases Expert Group, "Disparities in Chronic Kidney Disease Prevalence among Males and Females in 195 Countries: Analysis of the Global Burden of Disease 2016 Study," *Nephron*, vol. 139, no. 4, pp. 313–318, 2018, doi: 10.1159/000489897.

[11] E. Dritsas and M. Trigka, "Machine Learning Techniques for Chronic Kidney Disease Risk Prediction," *Big Data Cogn. Comput.*, vol. 6, no. 3, p. 98, Sep. 2022, doi: 10.3390/bdcc6030098.

[12] M. Islam, S. Akter, M. Hossen, S. Keya, S. Afrin, and S. Hossain, *Risk Factor Prediction of Chronic Kidney Disease Based on Machine Learning Algorithms*. 2021. doi: 10.1109/ICISS49785.2020.9315878.

[13] Md. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. W. Quinn, and M. A. Moni, "Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening," *IEEE J. Transl. Eng. Health Med.*, vol. 9, pp. 1–11, 2021, doi: 10.1109/JTEHM.2021.3073629.

[14] J. Xiao *et al.*, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression," *J. Transl. Med.*, vol. 17, no. 1, p. 119, Dec. 2019, doi: 10.1186/s12967-019-1860-0.

[15] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," *IEEE Access*, vol. 8, pp. 20991–21002, 2020, doi: 10.1109/ACCESS.2019.2963053.

[16] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021, doi: 10.1109/ACCESS.2021.3053763.

[17] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy," *IEEE Access*, vol. 8, pp. 55012–55022, 2020, doi: 10.1109/ACCESS.2020.2981689.

[18] G. Chen *et al.*, "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," *IEEE Access*, vol. 8, pp. 100497–100508, 2020, doi: 10.1109/ACCESS.2020.2995310.

[19] S. M. M. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Early Prediction of Chronic Kidney Disease Using Deep Belief Network," *IEEE Access*, vol. 9, pp. 135542–135549, 2021, doi: 10.1109/ACCESS.2021.3114306.

[20] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," *IEEE Access*, vol. 8, pp. 20991–21002, 2020, doi: 10.1109/ACCESS.2019.2963053.

[21] W. Wang, G. Chakraborty, and B. Chakraborty, "Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm," *Appl. Sci.*, vol. 11, no. 1, p. 202, Dec. 2020, doi: 10.3390/app11010202.

[22] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021, doi: 10.1109/ACCESS.2021.3053763.

[23] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," *J. Big Data*, vol. 9, no. 1, p. 109, Nov. 2022, doi: 10.1186/s40537-022-00657-5.

[24] T. Saroja and Y. Kalpana, "Hybrid missing data imputation and novel weight convolution neural network classifier for chronic kidney disease diagnosis," *Meas. Sens.*, vol. 27, p. 100715, Jun. 2023, doi: 10.1016/j.measen.2023.100715.

[25] Q. M. Alzubi, M. Anbar, Z. N. M. Alqattan, M. A. Al-Betar, and R. Abdullah, “Intrusion detection system based on a modified binary grey wolf optimisation,” *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6125–6137, May 2020, doi: 10.1007/s00521-019-04103-1.

[26] V. A. Chinnasamy and D. R. Shashikumar, “Breast cancer detection in mammogram image with segmentation of tumour region”.

[27] R. Yazdani, M. Mirmozaffari, E. Shadkam, and M. Taleghani, “Minimizing total absolute deviation of job completion times on a single machine with maintenance activities using a Lion Optimization Algorithm,” *Sustain. Oper. Comput.*, vol. 3, pp. 10–16, Jan. 2022, doi: 10.1016/j.susoc.2021.08.003.

[28] V. A. Chinnasamy and D. R. Shashikumar, “Breast cancer detection in mammogram image with segmentation of tumour region,” *Int. J. Med. Eng. Inform.*, vol. 12, no. 1, pp. 77–94, 2020.

[29] P. H. N. N. Tajunisha, “OPTIMAL PARAMETER SELECTION-BASED DEEP SEMI-SUPERVISED GENERATIVE LEARNING AND CNN FOR OVARIAN CANCER CLASSIFICATION”.

[30] A. Subasi, E. Alickovic, and J. Kevric, “Diagnosis of Chronic Kidney Disease by Using Random Forest,” in *CMBEBIH 2017*, vol. 62, A. Badnjevic, Ed., in IFMBE Proceedings, vol. 62. , Singapore: Springer Singapore, 2017, pp. 589–594. doi: 10.1007/978-981-10-4166-2_89.

[31] “Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics | IEEE Conference Publication | IEEE Xplore.” Accessed: Jan. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8477876

[32] B. Abdollahzadeh, F. Soleimanian Gharehchopogh, and S. Mirjalili, “Artificial gorilla troops optimizer: A new nature-inspired metaheuristic algorithm for global optimization problems,” *Int. J. Intell. Syst.*, vol. 36, Jul. 2021, doi: 10.1002/int.22535.

[33] S. Mirjalili, P. Jangir, S. Z. Mirjalili, S. Saremi, and I. N. Trivedi, “Optimization of problems with multiple objectives using the multi-verse optimization algorithm,” *Knowl.-Based Syst.*, vol. 134, pp. 50–71, Oct. 2017, doi: 10.1016/j.knosys.2017.07.018.