IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Classification and Diagnosis of Renal Failure Disease Using Machines Learning

Ammar Ahmed Ali[1], Omar Q. Alshebly[2] and Rizgar Maghdid Ahmed[3]

[1]Department of Educational Sciences and Educational Psychology, University of Mosul, Mosul, Iraq.
[2] Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.
[3]Department of Statistics, Salahaddin University, Erbil, Iraq

ARTICLE INFO

Article history:
Received 14/11/2024
Revised 15/11/2024
Accepted 14/1/2025
Available online 15/5/2025

Keywords:
Renal failure disease (RFD)
K-nearest neighbors (KNN)
Classification
Machine learning
Decision tree (DT)
Random forest (RF)

ABSTRACT

During the past few years, the number of classification techniques has increased with the rapid growth of technology, which depends on machine learning. Recently, medical experts and doctors have widely utilized machine learning, a branch of artificial intelligence, to aid in predicting and diagnosing various diseases. Regarding health research, machine learning techniques are used extensively for data processing. In this study, we applied three different machine learning algorithms to a medical diagnosis problem and analyzed their efficiency in predicting the results. The study focuses on the diagnosis and factors influencing renal failure disease, using a serum test for both presence and absence patients. The dataset used for the study consists of 165 cases and 13 attributes of RFD patients. The goal of this study is to find out how well k-nearest neighbors (KNN), decision tree (DT), and random forest (RF) classifiers work by looking at things like accuracy, geometric mean, kappa coefficient and area under the curve for RFD prediction. The experimental results show that the RF classifier performs better than the other classifiers. Additionally, based on the final fitted models, it was found that urea, albumin, and magnesium are the most significant factors that clearly impact patients with renal failure disease.

1. Introduction

In recent years, the rising incidence of renal failure demanded a focused investigation of the illness and its influencing variables via the use of artificial intelligence methods, which have garnered significant attention recently.

Renal failure disease (RFD), also known as chronic renal disease, refers to the progressive deterioration of kidney function. The kidneys filter waste and surplus fluids from the circulation, which are subsequently expelled as urine. In severe stages of chronic renal disease, perilous accumulations of fluid, electrolytes, and waste products may occur in the body[1]. Between 1999 and 2010, the national health and nutrition examination survey indicated a prevalence of RFD affecting 28 million individuals out of an estimated 200 million in the United States. Of these, 65.3% were diagnosed with RFD, with individuals suffering from diabetes and hypertension exhibiting significantly higher prevalence rates of 37% and 26%, respectively, in contrast to approximately 11% among those without these conditions [2].

Machine learning is a domain of computer sciences focused on the autonomous acquisition of knowledge from inputs. Classification is the primary issue of supervised machine learning. Classification models forecast classified labels for various objects [3]. The performance assessment of classification models often relies on the outcomes of training and testing datasets. A

* Corresponding author. E-**mail** address: ammar.ahmed@uomosul.edu.iq
https://doi.org/10.62933/qgj6hr87
This work is licensed under https://creativecommons.org/licenses/by-nc-sa/4.0/

training set comprises data used for developing a classifier, while a testing set is employed to evaluate the classifier's efficacy[4]. Chen et al. [5] used three-dimensional power Doppler imaging to compare logistic regression, support vector machines, and artificial neural networks for telling the difference between solid breast cancers that are benign and those that are malignant. The diagnostic performances of the three models (LRA, SVM, and NN) are indistinguishable, as shown by ROC curve analysis. In 2012, Koh and Jyoti [6] demonstrated cardiac disease prediction with three data mining techniques: neural networks, decision trees, and naive bayes. They found that neural networks with 15 characteristics outperformed two other methods, leading to their designation as the prediction model. Abid Sarwar et al. [7] evaluated the accuracy of the Naïve Bayes, artificial neural network, and KNN algorithms for type II diabetes. Type II diabetes is a disorder characterized by the pancreas's inability to create sufficient insulin or the cells' incapacity to use the produced insulin, resulting in abnormal blood glucose levels. The findings indicated that the neural network, with a prediction accuracy of 96%, outperformed Naïve Bayes at 95% and KNN at 91%. George et al. [8] introduced a diagnostic method for breast cancer using several machine learning techniques, including support vector machines and neural networks, reporting accuracy rates between 76% and 94% on a dataset of 92 pictures.

The primary objective of this study is to use three supervised machine learning algorithms to classify and diagnose renal failure illness among two patient groups (presence and absence) in order to determine the most effective classifier based on various performance assessment criteria. The research also investigated the main variables influencing renal failure disease.

2. Methodology

2.1 K-Nearest Neighbors (KNN) Classifier

The k-Nearest Neighbor (k-NN) algorithm operates based on the principle of similarity between a new data point and a set of stored data points, referred to as training points. The primary objective of this algorithm is to classify the new test point into the class that is most similar among the available classes.

The k-NN algorithm is non-parametric, which means it does not make any assumptions about the underlying data distribution. It is often referred to as a "lazy learning" algorithm because it does not undergo a traditional training process; instead, it retains the training dataset in its entirety for future use [9].

When it comes to classifying a new dataset (test data), the k-NN algorithm evaluates the distance between the new point and the stored training points. The parameter $k$denotes the number of nearest neighbors to consider in the classification process. The Euclidean distance is typically employed to quantify the distance between the new point and each of the training points[10].

Once the distances are calculated, the algorithm identifies the k nearest neighbors of the new point. The classification of the new point is then determined by the class that has the highest frequency among these neighbors. This process leverages the Euclidean distance function D_i to effectively locate the closest neighbors within the feature vector space.

$$D_i = \sqrt{(x_1 + x_2) + (y_1 + y_2)^2} \qquad (1)$$

where x_1, x_2, y_1, and y_2 are variables for input data [9].

2.2 Decision Tree (DT) Classifier

Decision trees are a popular machine learning algorithm used for both classification and regression tasks. They mimic human decision-making by creating a model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. In the context of machine learning, decision trees learn from data, building a model that can make predictions about new data [11].

2.2.1 How Decision Trees Work

A decision tree is a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. Each node in the tree represents a test on an attribute (e.g., Is age > 25?), and each branch represents the outcome of the test (e.g., yes or no). The leaf nodes of the tree represent the class labels (in classification) or the predicted value (in regression)[11].

2.2.2 The process of building a decision tree typically involves[12]:

1. Selecting the root node: This is the first attribute that is tested.
2. Splitting the data: The data is split into subsets based on the outcome of the test at the current node.
3. Creating child nodes**:** A child node is created for each branch of the tree.
4. Repeating steps 2 and 3: The process is repeated for each child node until a stopping criterion is met, such as a maximum depth of the tree or a minimum number of samples in a leaf node.

2.2.3 Key Concepts

- Information Gain: This measures the decrease in entropy (uncertainty) after a dataset is split on an attribute. It's commonly used to select the best attribute for splitting at each node.
- Gini Impurity: Another measure of impurity used in decision tree algorithms. It calculates the probability of incorrectly classifying a randomly chosen element if it were randomly [12]. labeled according to the distribution of the class labels in the subset.
- Pruning**:** This process removes branches from a tree in order to reduce overfitting [14].

2.2.4 Mathematical Formulation

While decision trees are often visualized as trees, the underlying mathematical formulation involves concepts from information theory and probability. The core idea is to find the best split at each node that maximizes information gain or minimizes impurity[14].

2.2.4.1Information Gain[13]:

Information Gain (S, A) = Entropy(S) - ∑ |SV| / |S| * Entropy (SV) (2)

where:

- S is the set of examples
- A is an attribute
- SV is the subset of S for which attribute A has value v
- Entropy(S) is the entropy of the set S

2.2.4.2 Gini Impurity:

$Gini(S) = \sum p(i)(1-p(i))$

where p(i) is the probability of class i in the dataset S [11] .

2.3 Random Forest (RF)

Random Forests (RF) exemplify the power of ensemble learning, a statistical technique that combines multiple models to improve predictive performance. In the context of RF, numerous decision trees are aggregated to form a robust and accurate predictive model [13].

Mathematically, the prediction of an RF model, denoted as RF(x), for a given input x can be expressed as:

$RF(x) = (1/B) * \Sigma (Tb (x; \Theta b))$ (3)

from b=1 to B

- RF(x): The predicted output of the RF model for input x.

- Tb (x; Θb): The prediction of the bth decision tree for input x, where Θb represents the random parameters associated with that tree.
- B: The total number of trees in the forest.

The random nature of Θb ensures diversity among the trees, preventing overfitting and improving generalization. This concept aligns with the Law of Large Numbers, a fundamental statistical principle stating that the average of a large number of independent and identically distributed random variables converges to the expected value. In the case of RF, as the number of trees increases, the average prediction tends to become more accurate and stable [14].

By leveraging the collective wisdom of multiple decision trees, RF models offer several advantages [10]:

1.Reduced Overfitting**:** The ensemble nature of RF mitigates the risk of overfitting, a common issue in machine learning models.

2. Improved Accuracy: By combining multiple models, RF often achieves higher predictive accuracy compared to individual decision trees.

3.Robustness to Noise: RF is relatively insensitive to noisy data, as the ensemble nature helps to average out the impact of outliers.

4.Feature Importance: RF can provide insights into the relative importance of different features in the data.

2.4 Performance Evaluation

This section of the manuscript presents a discussion on various methods for assessing the performance of KNN, DT, and RF.
1. Accuracy
Accuracy quantifies the efficacy of our model. The expectation is that it will approximate 1, if our model is functioning well.

$$\text{Accuracy} = (TP+TN)/N \quad (5)$$

Whereas: - TN: The quantity of samples identified as negative (lacking the characteristic) is indeed negative.
TP: The quantity of samples identified as positive (exhibiting the feature) is indeed positive.
N: Aggregate quantity of samples. [15]
2.Geometric Mean

The joint performance was evaluated by utilizing the Geometric Mean of sensitivity and specificity [16].

$$G-Mean = \sqrt{\text{Sensitivity} \times \text{Specificity}}$$

3.Cohen's Kappa Coefficient

Cohen's Kappa, sometimes referred to as Kappa, serves as a metric for assessing the agreement between two persons when they use two binary variables to evaluate the same phenomenon. Kappa quantifies the proportion of data values located in the principal diagonal of the table and then calibrates these values for the level of concordance that might be anticipated only by chance.
To calculate Kappa, one must first determine the observed degree of agreement.

$$P_0 = \frac{TP+TN}{N} \quad (6)$$

This value needs to be compared to the value that you would expect if the two raters were totally independent:

$$P_e = \left(\frac{(TP+FP)}{N} * \frac{(TP+FN)}{N}\right) + \left(\frac{(TN+FN)}{N} * \frac{(TN+FP)}{N}\right) \quad (7)$$

Then, the value of Kappa is defined as:

$$K = \frac{P_0-P_e}{1-P_e} \quad (8)$$

The highest value for kappa is achieved when the observed level of agreement is 1, resulting in the numerator being equal to the denominator. As the observed probability of agreement declines, the numerator declines. Kappa can indeed be negative, although such occurrences are relatively rare.
The following explanations demonstrate Kappa's values along with the corresponding estimates for each explanation:

- Poor agreement = Less than 0.20
- Fair agreement = 0.20 to 0.40

- Moderate agreement = 0.40 to 0.60
- Good agreement = 0.60 to 0.80
- Very good agreement = 0.80 to 1.00

[17][18].

4. The Area under curve (ROC curve)

AUC is defined as an indicator of the classifier's overall performance over all potential threshold levels. When the probability distributions for both detection and mistaken alarms are established, one may generate a ROC curve by graphing the cumulative distribution. The area under the probability curve from (-∞ to +∞), often represented as the area under the ROC curve, serves as a metric for assessing the quality of probability classification. The area under the curve using the following formula: [19][20].

$$A_{ROC} = \int_0^1 \frac{TP}{P} d\frac{FP}{N} = \frac{1}{PN}\int_0^N TP * dFP \qquad (9)$$

3. Results and discussion

The dataset comprises 165 individuals and includes 14 variables, of which 13 are independent variables and one is a dependent variable indicating the presence (1) or absence (0) of renal failure disease (RFD) based on serum blood tests. The real data was obtained from the center of artificial kidney at Abn-Sina Teaching Hospital in Mosul city. There are no missing values in this dataset. The analyzed samples include 75 people without RFD and 90 patients with RFD, with ages ranging from 15 to 85 years.

Table 1 presents a description of the data analyzed in this study.

Table (1): Description of the study variables

No.	Variables(Name , and Type)
1.	Class(y) 1 presence of RFD. 0 absence of RFD. Categorical
2.	Sex(x_1) 1 male. , 2 female. categorical
3.	Age(x_2) numerical.
4.	Smoking(x_3) 1 smoked, 2 non smoked. Categorical
5.	Urea(x_4) numerical.
6.	Protein(x_5) numerical.
7.	Albumin(x_6) numerical.
8.	Clubin(x_7) numerical.
9.	Sodium NA(x_8) numerical.
10.	Potassium K(x_9) numerical.
11.	Calcium CA(x_{10}) numerical.
12.	Zinc ZN(x_{11}) numerical.
13.	Cupper CU(x_{12}) numerical.
14.	Magnesium MG(x_{13}) numerical.

The dataset is divided into two groups: the training dataset, comprising 70% (116 patients) of the samples, and the testing dataset, consisting of the remaining 30% (49 patients). To compare adjusters and mitigate the impact of data fragmentation, all employed classification methods are assessed based on their classification performance metrics through 10-fold cross-validation, averaged over 10 partitions. All implementations of the study on real data applications are conducted using R.

3.1 Performance Evaluation of Models Applied and Comparison

After partitioning the data into two categories (training and testing), we implemented the first model, which is the KNN model.

Table (2) presents the assessment criteria for the K-nearest neighbors model applied to the training and testing datasets.

Table (2): Performance of KNN Model

Criteria	Training Dataset	Testing Dataset
Accuracy	**96.4%**	**95.9%**
G-mean	**95.1%**	**94.86%**
Kappa Coefficient	**93%**	**91.68%**

Table 2 indicates that the KNN model exhibits an accuracy of 96.4% on the training dataset; however, this value declines to 95.9% on the testing dataset. Additionally, the G-mean values exceed 90%, recorded at 95.1% and 94.86% for the training and testing datasets, respectively, concerning RFD patients. Additionally, the model's coherence is commendable, as evidenced by the Kappa coefficient, which shows values of 93% and 91.68%, respectively.

The Receiver Operator Characteristic (ROC) is an additional tool utilized for measuring model performance. The model's accuracy is assessed through the area under the curve metric. The value of the Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC)

of the testing dataset for the KNN model is 0.971, as illustrated in Figure 1.

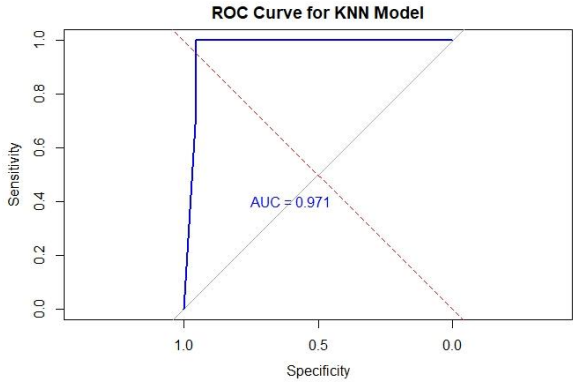

Figure (1) (ROC) curve of Testing dataset for KNN

The ROC curve is presented with sensitivity on the y-axis and specificity on the x-axis.Figure 1 indicates that the area under the curve value is 97.1%. The ROC is a statistic used to assess the efficacy of classifiers.

Table (3) presents the assessment criteria for the decision tree model applied to the training and testing datasets.

Table (3): Performance of DT Model

Criteria	Training Dataset	Testing Dataset
Accuracy	**96.5%**	**93.88%**
G-mean	**95.97%**	**93.48%**
Kappa Coefficient	**93.03%**	**87.68%**

Table 3 demonstrates that the DT model achieves an accuracy of 96.5% on the training dataset, although this figure decreases to 93.88% on the testing dataset. The G-mean values surpass 90%, measured at 95.97% for the training dataset and 93.48% for the testing dataset, pertaining to RFD patients. The model's coherence is notable, shown by the Kappa coefficient values of 93.03% and 87.68%, respectively.

The area under the curve (AUC) for the ROC of the testing dataset is 0.9411, as illustrated in Figure 2.

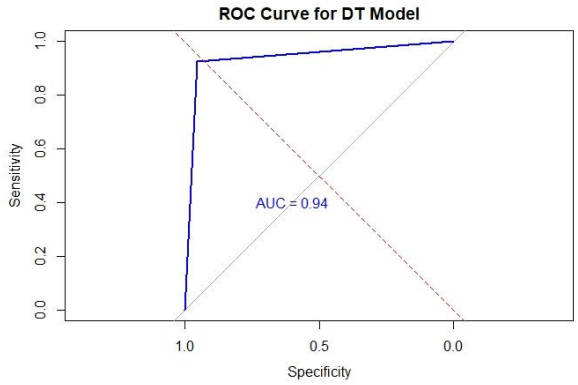

Figure (2) (ROC) curve of Testing dataset for DT

Figure (2) illustrates the ROC curve, with an area under the curve of 94%.

Table (4) delineates the evaluation criteria for the random forest model used on the training and testing datasets.

Table (4): Performance of RF Model

Criteria	Training Dataset	Testing Dataset
Accuracy	**100%**	**97%**
G-mean	**100%**	**94.91%**
Kappa Coefficient	**100%**	**91.8%**

Table 4 highlights that the random forest model successfully classified all 116 cases without any mistakes. The model exhibits an Accuracy of 100%, indicating that all other metrics, including G-mean and kappa coefficient, will also equal 100%. The prior findings indicate that the RF models have attained exceptional performance on the training dataset. Conversely, we observed that all metrics (Accuracy, G-mean, and kappa coefficient) have declined in the testing dataset compared to the training dataset, with values of 97%, 94.91%, and 91.8%, respectively. The Area under the Curve (AUC) of the ROC for the testing dataset is 1, as seen in Figure 3.

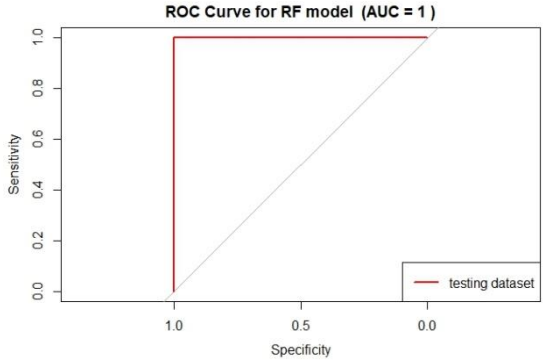

Figure (3) (ROC) curve of Testing dataset for RF

Figure (3) illustrates the ROC curve, with an area under the curve of 100%.

Finally, we examined the comparison of three classification models (KNN, DT, and RF) based on the metrics of model accuracy, G-mean, Kappa coefficient, and area under the ROC curve.

Table 5 presents a comparison just for the testing dataset. The optimal classifier is determined by the testing dataset.

Table (5): Performance Evaluation Criteria between Models

Model	K-nearest neighbors	Decisions tree	Random forest
Accuracy	**95.9%**	**93.88%**	**97%**
G-mean	**94.86%**	**93.48%**	**94.91%**
Kappa Coefficient	**91.68%**	**87.68%**	**91.8%**
$(AUC)_{ROC}$	**97.1%**	**94.1%**	**100%**

Table 5 highlights that the accuracy of KNN was 95.9%. The performance of the approach decreased while using the DT model, which achieved an accuracy of 93.88%, whereas the RF method attained an accuracy of 97%. This suggests a preference for RF due to its precision; the more precise the approach, the more superior it is considered. The G-mean criteria for KNN was 94.86 percent. The DT model yielded a value of 93.48, but the geometric mean criteria for the RF approach was 94.91%. This promotes the RF based on the G-mean criteria. Furthermore, the Kappa coefficient for the RF approach was the highest, at 91.8%, indicating a superior level of agreement compared to other methods.On the other hand,The area under the curve (ROC) for the KNN was 97.1%. The model's performance decreased to 94.1% when utilizing the decision tree (DT), whereas the area under the curve (ROC) for random forest (RF) models reached 100%. A higher value of the area under the curve (ROC) indicates superior performance.

3.SIGNIFICANT VARIABLES INFLUENCING RFD PATIENTS

Upon assessing and identifying the ideal solution using the RF model, it was essential to ascertain the most significant variables influencing RFD patients in our research. Utilizing the VarImp function inside the caret package to assess variable significance for the ROC curve in the machine learning models.

Table 6 illustrates the relevance of the independent variables in the KNN model.

Table (6): Variables Importance for KNN

ROC curve variable importance

	Importance
urea	100.0000
albumin	83.5859
mg	74.6533
k	72.7507
protein	58.3360
ca	32.6024
na	30.4418
age	24.8952
zn	15.9303
cu	13.8987
smoking	1.6769
clubin	0.4837
sex	0.0000

Table 6 indicates that the factors of urea, albumin, and magnesium had the most significant impact on the KNN model, with respective influences of 100%, 83.5%, and 75%, compared to other independent variables.

Figure 4 shows the importance of variables for KNN Model.

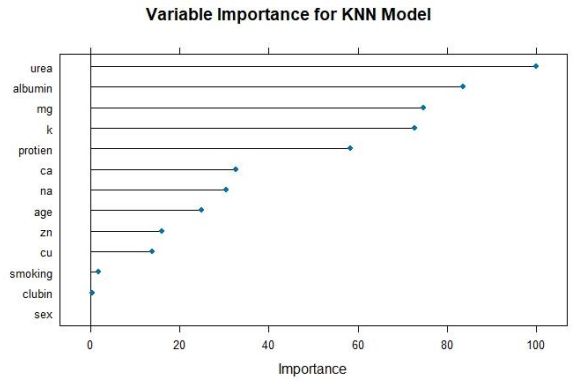

Figure (4) Importance Variables of KNN model

Conversely, using Garson's technique to assess relative variable relevance revealed significant variables for the Decision Tree and Random Forest models, as shown in Tables 7 and 8.

Table (7): Variables Importance for DT

	Overall
albumin	**36.63052**
k	19.94836
mg	**30.61771**
protein	14.50000
urea	**49.88510**
sex	0.00000
age	0.00000
smoking	0.00000
clubin	0.00000
na	0.00000
ca	0.00000
zn	0.00000
cu	0.00000

Table (8): Variables Importance for RF

	Overall
sex	0.6749094
age	7.9085772
smoking	0.5321570
urea	**27.9977939**
protein	5.7392805
albumin	**12.2887285**
clubin	3.5358083
na	7.0830485
k	11.1262586
ca	3.6614641
zn	0.7278021
cu	2.9394394
mg	**11.2872162**

Tables 7 and 8 indicate that urea, albumin, and magnesium are the variables with the highest importance and contribution in patients with chronic renal failure.

Figures 5 and 6 illustrate the importance of the above variables.

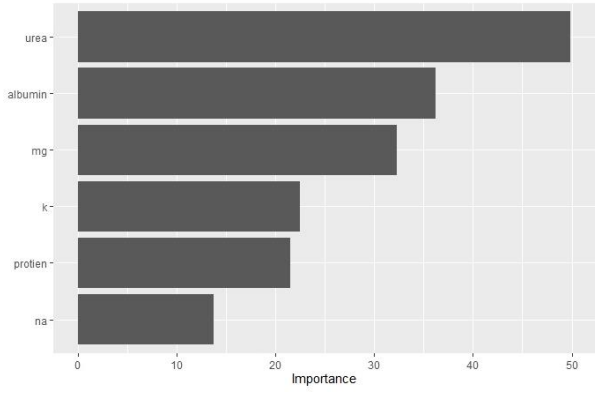

Figure (5) Importance Variables of DT model

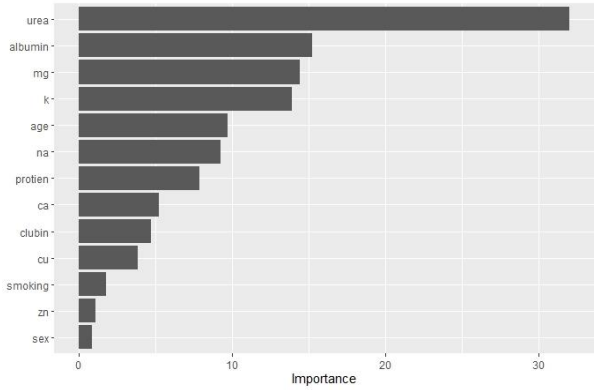

Figure (6) Importance Variables of RF model

4. Conclusions

Renal failure disease holds significant importance due to its potential to cause mortality and health crises in the community. In addition, since RFD is one of the illnesses that has been on the rise in prevalence over the last several years, this research was successful in establishing that the RF method is the most effective approach for classifying and diagnosing RFD patients. Utilizing the three approaches (KNN, DT, and RF), we came to the conclusion that urea, albumin, and magnesium are the most effective and significant variables. This was our conclusion after finding that the factors had the greatest impact on chronic renal failure data.

References

[1] Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., . . . Yang, C.-W. (2013). Chronic kidney disease: global dimension and perspectives. *The Lancet*, *382*(9888), 260-272.

[2] NCHS (National Center for Health Statistic Survey Data) Centers for Disease Control and Prevention (CDC). Hyattsville:Department of Health and Human Services, Centers for Disease Control and Prevention. 2012.

[3] Basturk, T., Sari, O., Koc, Y., Eren, N., Isleem, M., Kara, E., . . . Hasbal, N. (2017). Prognostic significance of NGAL in early stage chronic kidney disease. *Minerva Urologica e Nefrologica*, *69*(3).

[4] Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of chemical information and computer sciences*, *43*(6), 1882-1889.

[5] Chen, S.-T., Hsiao, Y.-H., Huang, Y.-L., Kuo, S.-J., Tseng, H.-S., Wu, H.-K., & Chen, D.-R. (2009).

Comparative analysis of logistic regression, support vector machine and artificial neural network for the differential diagnosis of benign and malignant solid breast tumors by the use of three-dimensional power Doppler imaging. *Korean Journal of Radiology*, *10*(5), 464-471.

[6] Bhatla, N., & Jyoti, K. (2012). An analysis of heart disease prediction using different data mining techniques. *International Journal of Engineering*, *1*(8), 1-4.

[7] Sarwar, A., & Sharma, V. (2014). Comparative analysis of machine learning techniques in prognosis of type II diabetes. *AI & society*, *29*, 123-129.

[8] George, Y. M., Zayed, H. H., Roushdy, M. I., & Elbagoury, B. M. (2013). Remote computer-aided breast cancer detection and diagnosis system based on cytological images. *IEEE Systems Journal*, *8*(3), 949-964.

[9] Saxena, R. (2016). Knn classifier, introduction to K-nearest neighbor algorithm. *DataAspirant [online], available at: https://dataaspirant. com/2016/12/23/k-nearest-neighbor-classifier-intro*.

[10] Swain, D., Mehta, U., Bhatt, A., Patel, H., Patel, K., Mehta, D., . . . Manika, S. (2023). A robust chronic kidney disease classifier using machine learning. *Electronics*, *12*(1), 212.

[11] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*, 81-106.

[12] Ibrahim, I. M., & Abdulazeez, A. M. (2021). The role of machine learning algorithms for diagnosing diseases. *learning*, *4*(5), 6.

[13] Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

[14] Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, *9*(1), 109.

[15] Alshebly, O., & Abdullah, S. N. (2024). Proposed Two-Steps Procedure of Classification High Dimensional Data with Regularized Logistic Regression. *Statistics, Optimization & Information Computing*, *12*(2), 325-342.

[16] Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, *42*(8), 1778-1790.

[17] Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, *43*(6), 1223-1232.

[18] Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *arXiv preprint cmp-lg/9602004*.

[19] Alshebly, O. Q., & Abdullah, S. N. (2024). The Fuzziness Models with The Proposed New Conjugate Gradient Method for The Classification of High-Dimensional Data in Bioinformatics. *Journal of Economics and Administrative Sciences*, *30*(142), 425-448.

[20] Hosmer, D., & Lemeshow, S. (2000). Applied Logistic Regression, New York: Johnson Wiley & Sons: Inc.