Assessment and prediction of Chronic Kidney using an improved neutrosophic artificial intelligence model

Neyda Hernández Bandera[1], Jenny M. Moya Arizaga[2], Enrique Rodríguez Reyes[3]

[1]Docente de la carrera de Medicina de la Universidad Regional Autónoma de los Andes (UNIANDES), Ecuador

[2]Docente de la carrera de Medicina de la Universidad Regional Autónoma de los Andes (UNIANDES Santo Domingo), Ecuador

[3]Docente de la carrera de Medicina de la Universidad Regional Autónoma de los Andes (UNIANDES Ambato), Ecuador

Emails: ua.neydahernandez@uniandes.edu.ec; docentetp54@uniandes.edu.ec; ua.enriquerodriguez@uniandes.edu.ec

Abstract

CKD, or chronic kidney failure, is characterized by a gradual decline in kidney operation over time and may be linked to a wide range of medical conditions. Initial detection and therapy are the best tools for combating chronic kidney disease, although they often only delay the development of renal failure. The eGFR-based CKD grading system is useful for risk stratification, patient monitoring, and treatment strategy development. Personalized care and treatment planning will be possible if this research is successful in predicting how soon a CKD individual will need to begin dialysis. The machine learning methods used to predict CKD. But the dataset contains uncertain information, so the neutrosophic set is used to overcome this issue. This paper suggests a framework including the neutrosophic DEMATEL and machine learning method to predict CKD. The neutrosophic DEMATEL method is used to give weights to all variables of the dataset. Then conduct the preprocessing data to eliminate the variables with the least weight. The three machine learning methods used in this paper are Gradient Boosting (GB), Ada Boosting (AB), and Random Forest (RF). The results show the accuracy of the three algorithms. The AB has a 99.166% accuracy, and it is the highest accuracy in this paper followed by the GB has 98.3%, then RF has 92.85%.

Keywords: Neutrosophic Set; DEMATEL; Machine Learning; Kidney; Prediction

1. Introduction

CKD is characterized by a gradual decline in kidney capacity that eventually necessitates the use of substitute kidney treatment (dialysis or transplant). Having kidney failure or a predicted glomerular filtration rate of fewer than sixty milliliters per 1.73 m2 that lasts for at least three months is what medical professionals call chronic kidney disease. The global incidence of chronic kidney disease (CKD) is rising in tandem with other demographic and epidemiological shifts. The effects of this illness on people's daily lives and the long-term viability of national healthcare systems are devastating. In 2012, CKD was responsible for 2.96 million disability-adjusted life years and 2.5 million lost life

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

years, or 1% to 3% of the global total. Thus, evaluating how to effectively and rapidly detect and treat people with CKD is of the highest significance[1], [2].

In different parts of the world, CKD might have different causes. Approximately 70-90% of all cases of CKD and kidney damage may be attributed to only three basic diseases: hypertension, diabetes mellitus, and recurrent glomerulonephritis. Despite their prominence, other factors (such as pollutants, diseases, and autoimmune illnesses) contribute to the pathophysiology of CKD as well. Similarly, the development of chronic kidney disease (CKD) is influenced by a wide variety of variables, both those that cannot be changed (such as age, gender, and ethnicity) and those that can be changed (such as blood pressure levels and proteinuria)[3], [4].

AI refers to the ability of robots designed by humans to do tasks traditionally associated with human intellect, such as making complicated decisions or analyzing large amounts of data. ML, or machine learning, is a set of techniques that enables systems to learn from data and, hence, do difficult tasks by performing complex computations. Recent advances in artificial intelligence and machine learning have promised tantalizing answers to pressing clinical questions like how to make a diagnosis with little or conflicting evidence or how to foretell a patient's prognosis. Researchers have attempted to employ ML's skills to handle complicated issues like predicting CKD evaluation and outlook and controlling its therapy, given ML's immense potential and its ability to learn from data[5], [6].

The dataset has unreliable data. So the neutrosophic set is used to overcome this uncertain information. Smarandache introduced the neutrosophic set, a generalization of the IFS and FS, in 1995. Zhang developed a single-valued neutrosophic set (SVNS), a part of an NS, to make NSs more usable in practice[7], [8]. Inaccuracy (rejection), truth (acceptance), and indeterminacy (vague) are the three distinct membership functions that distinguish NSs, especially single-valued NSs. When compared to FSs and IFSs, NSs are more adaptable and do a better job of displaying positive, negative, and ambiguous information. Since indeterminacy can be measured directly and truth, indeterminacy, and falsity membership characteristics are independent, it can handle indeterminate and unreliable data[9], [10].

The objective of this paper is to introduce the neutrosophic set as a feature selection method. The NS hybrid with the DEMATEL method to compute the weights of variables in the dataset. Then the highest rank will be selected and others will be rejected. Then the machine learning algorithms will be applied to predict chronic kidney. Three algorithms are selected GB, AB, and RF.

2. Related work

This section presented some previous studies in chronic kidney. For example, An approach based on machine learning was presented for the detection of CKD by the authors [11]. A substantial number of errors plague the CKD data set that was retrieved from the machine learning repository at the University of California, Irvine (UCI). KNN imputation, which chooses many full examples with the most comparable results to analyze the data that is not present for every imperfect sample, has been employed to fill in the absent values. In practice, missing values are common in healthcare data because individuals may forget to take measures for a variety of reasons. Six machine learning methods (logistic regression, random forest, support vector machine, k-nearest neighbor, naive Bayes classifier, and feed-forward neural network) were employed to create algorithms once the insufficient data set was successfully populated.

The authors [12] proposed utilizing the support vector machines method to categorize individuals with chronic renal disease based on clinical characteristics. Clinical history, physical exams, and laboratory testing provide the basis of chronic kidney disease data. The classification success percentage for patients with renal disorders using these three experimental criteria was above 93%.

Using a method that makes use of ML approaches, the authors [13] set out to create effective tools for forecasting the onset of CKD. To address the imbalance between the two categories, they first apply class balancing, then rank and analyze the characteristics, and lastly, we train and compare several machine learning models using a variety of performance indicators.

The authors [14] provided a machine-learning approach to clinical data to forecast chronic renal disease. We investigate four different types of machine learning: K-nearest neighbors (KNN), support vector machines (SVM), logistic regression (LR), and decision trees (DTs).

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

The authors [15] set out to use machine learning techniques for the prediction of CKD and proposed a decision tree to get concrete outcomes with important accuracy by comparing its efficacy concerning its requirements and sensitivity.

3. Material and Methods

First, several machine learning methods were used to construct models for analyzing the data. The most effective models were chosen to be possible building blocks. The elements of the types were identified via a study of their errors in judgment. Then, to improve efficiency, a combined framework was set up. The steps of the suggested methodology are shown in Figure 1.

Figure 1: The steps of the proposed methodology.

The DEMATEL technique, developed at the Trial Evolution Laboratory, is now one of the most well-known MCDM approaches available. The DEMATEL technique was created in the 1970s to simplify difficult challenges in establishing causality. Developing the direct impact matrix, creating the direct influence matrix[16], [17], building the complete influence matrix, and creating the influential connection map are the four formal processes in DEMATEL. The power of DEMATEL lies in its ability to systematically build and assess the structure of intricate causal linkages among matrix or diagram variables[18]–[20].

The first step is, Build the direct relation matrix D

Then normalize the direct relation matrix as:

$$N_{ij} = \frac{A^{L-}+A^{U-}}{2} + \frac{A^{L+}+A^{U+}}{2} + \left(1 - \frac{B^{L-}+B^{U-}}{2}\right) + \left(1 - \frac{B^{L+}+B^{U+}}{2}\right) - \left(\frac{C^{L-}+C^{U-}}{2}\right) 1 - C^{U-} - \left(\frac{C^{L-}+C^{U-}}{2}\right) 1 - C^{U+} \quad (1)$$

Then compute the total relation matrix as:

$$T = D(I - D)^{-1} \quad (2)$$

Then set up the relationships between variables

Decision trees are the building blocks of Random Forest. To determine which subset of characteristics is most useful, it uses either the Information Gain or the Gini index. It uses majority voting on the results of many decision trees to choose how to categorize a given occurrence. The RF employs decision trees as its primary predictor. It rotates the

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

training data using a transformation matrix before beginning the training process. Subsets of the set of characteristics are selected at random[21].

Boosting algorithms iteratively integrate poor learners (those that perform just marginally better than chance) to create a powerful learner. For regression, the boosting-like approach known as gradient boosting has been developed. With a training dataset $D = \{x_i, y_i\}^{N_1}$, gradient boosting seeks to minimize the anticipated value of a loss function $L(y, F(x))$ to identify a function $F(x)$ that maps occurrences x to their results y. Using a weighted sum of operations, gradient boosting creates an additive estimate of $F(x)$ [22].

$$F_m(x) = F_{m-1}(x) + p_m h_m(x) \quad (3)$$

$$F_0(x) = \arg\min_a \sum_{i=1}^{N} L(y_i, a) \quad (4)$$

$$p_m h_m(x) = \arg\min_a \sum_{i=1}^{N} L(y_i, F_{m-1}(x) + ph(x_i)) \quad (5)$$

We can compute the pseudo-residuals as:

$$r_{mi} = \left[\frac{\partial L(y_i, F(x))}{\partial F(x)}\right]_{F(x)= F_{m-1}(x)} \quad (6)$$

To integrate the results of a series of L low-quality classification methods, the AdaBoostM1 uses a weighted majority voting system. Given a training set of N examples, and uniform starting weights, at every boosting phase r, every sample is weighted. The error value is used to calculate the weights. When the preceding categorization was wrong, the instance's importance is raised; otherwise, it is lowered. The magnitude of the inaccuracy determines how heavily the sample is relied upon. This is done until the error rate stabilizes. The ultimate forecast is calculated by[23], [24]:

$$G(x_i) = sign\left(\sum_{l=1}^{l} a_l G_l(x_i)\right)(\{-1, +1\}) \quad (7)$$

The derived coefficients a a take into account the classification mistake and give greater weight to more precise classifiers $G_l(x_i)$ [25]. [26].

4. Application

In this section, we apply the steps of the neutrosophic DEMATEL method as a feature selection. Then apply the machine learning algorithms like GB, AB, and RF. We apply these algorithms on the CKD dataset. The dataset has 28 features. Apply some descriptive statistics into a dataset like mean minimum and maximum. Table 1 shows the descriptive statistics.

Table 1: Descriptive statistics of CKD dataset.

	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
AGE	400	51.5625	16.983	2	42	55	64	90
BP	400	76.575	13.48979	50	70	80	80	180
SG	400	1.017712	0.005434	1.005	1.015	1.02	1.02	1.025
AL	400	0.9	1.31313	0	0	0	2	5
SU	400	0.395	1.040038	0	0	0	0	5
RBC	400	1.2625	0.655491	0	1	1	2	2
PC	400	0.9725	0.593823	0	1	1	1	2
PCC	400	0.125	0.360138	0	0	0	0	2
BA	400	0.075	0.299331	0	0	0	0	2
BGR	400	145.0625	75.26077	22	101	121	150	490
...	...	...	...	...	...	...	...	...

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

PCV	400	29.8525	10.52915	0	22	30	39.25	44
WC	400	64.43	28.14962	0	49	71	92	92
RC	400	34.8375	13.39738	0	26	36	49	49
HTN	400	0.3775	0.495588	0	0	0	1	2
DM	400	3.31	0.591417	0	3	3	4	5
CAD	400	1.09	0.319618	0	1	1	1	3
APPET	400	0.21	0.413918	0	0	0	0	2
PE	400	0.195	0.402965	0	0	0	0	2
ANE	400	0.155	0.36921	0	0	0	0	2
CLASSIFICATION	400	0.755	0.968152	0	0	0	2	2

Then we applied some plots on the dataset to show the information from it like histogram and scatter plots as shown in Figures 2 and 3.

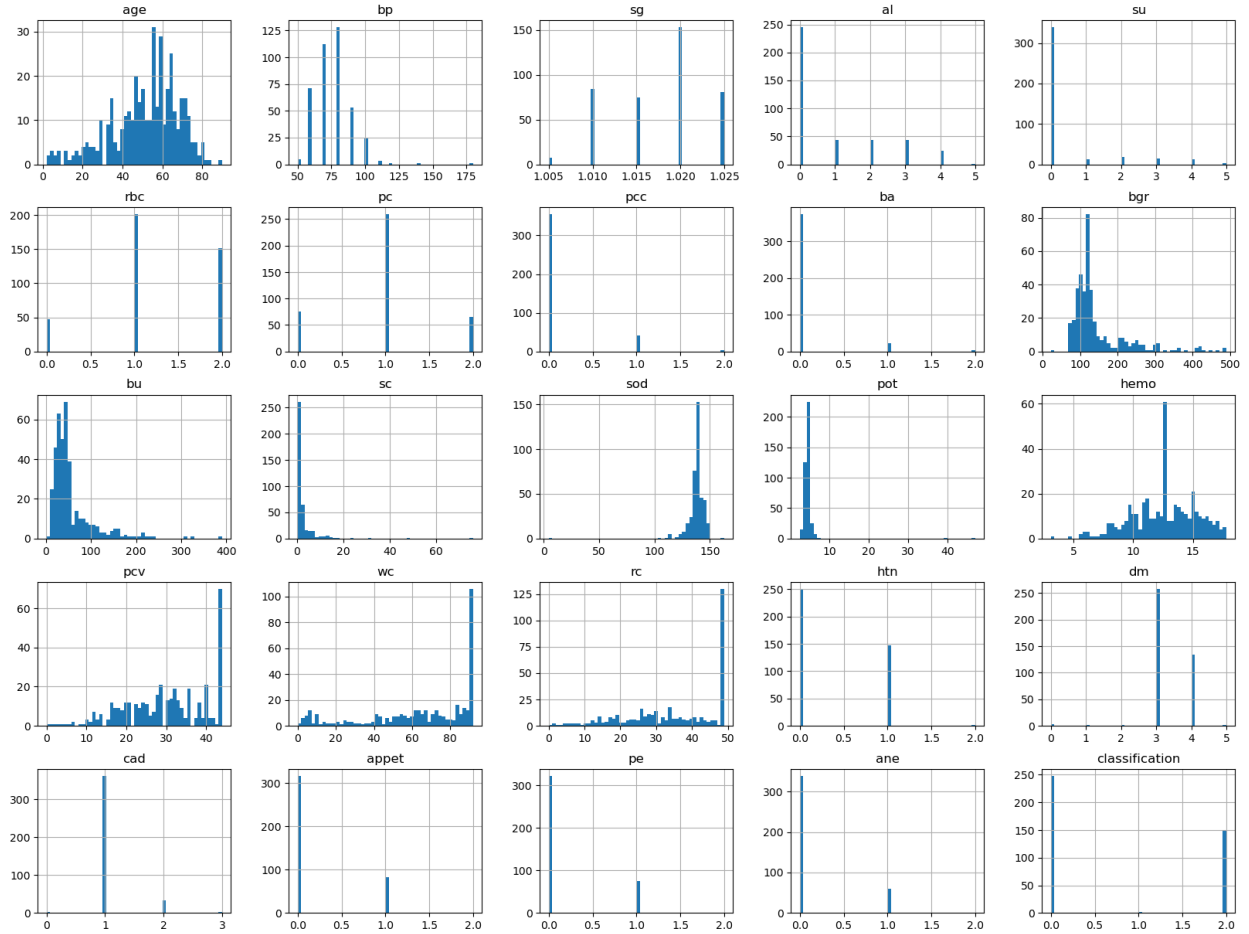

Figure 2: The histogram of the dataset

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

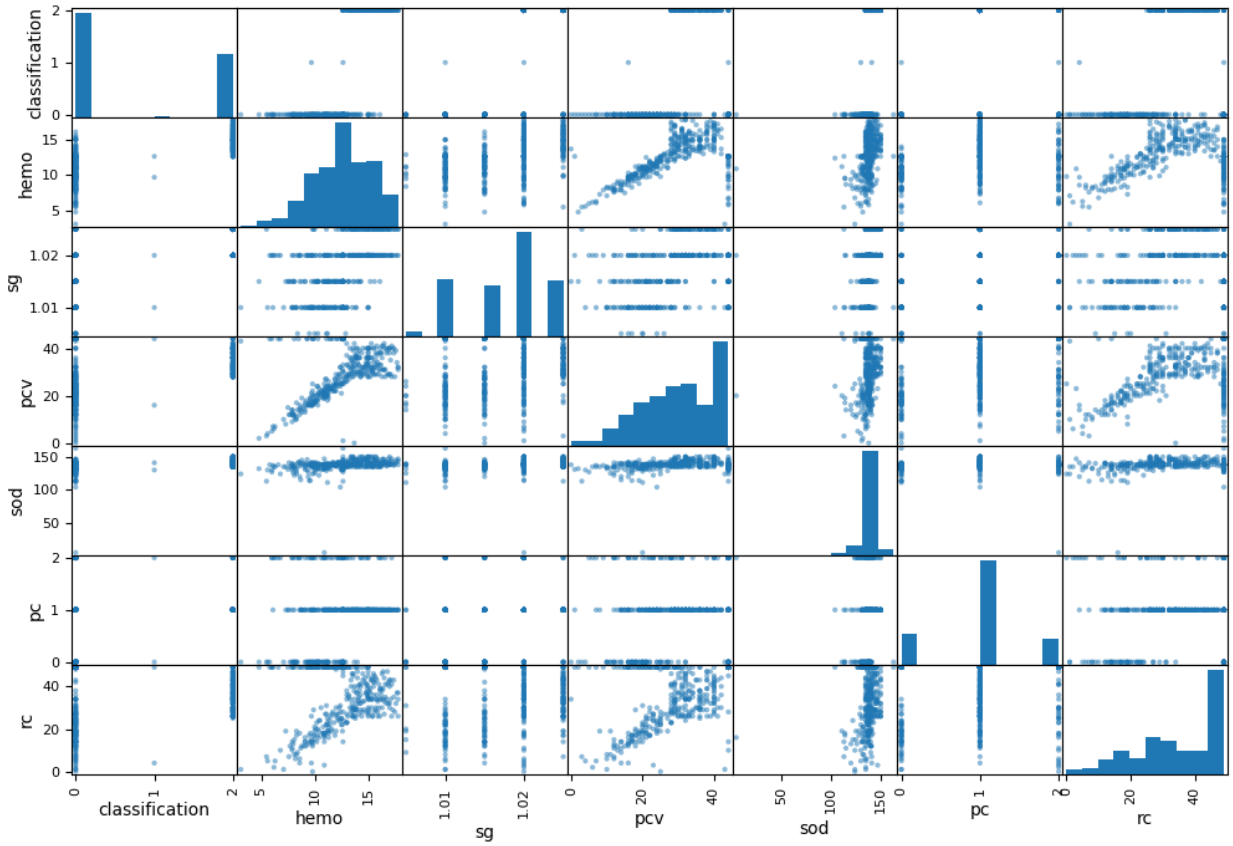

Figure 3: Scatter plot of the dataset.

Then obtain the heat map to show the relationship between variables and target variables.

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

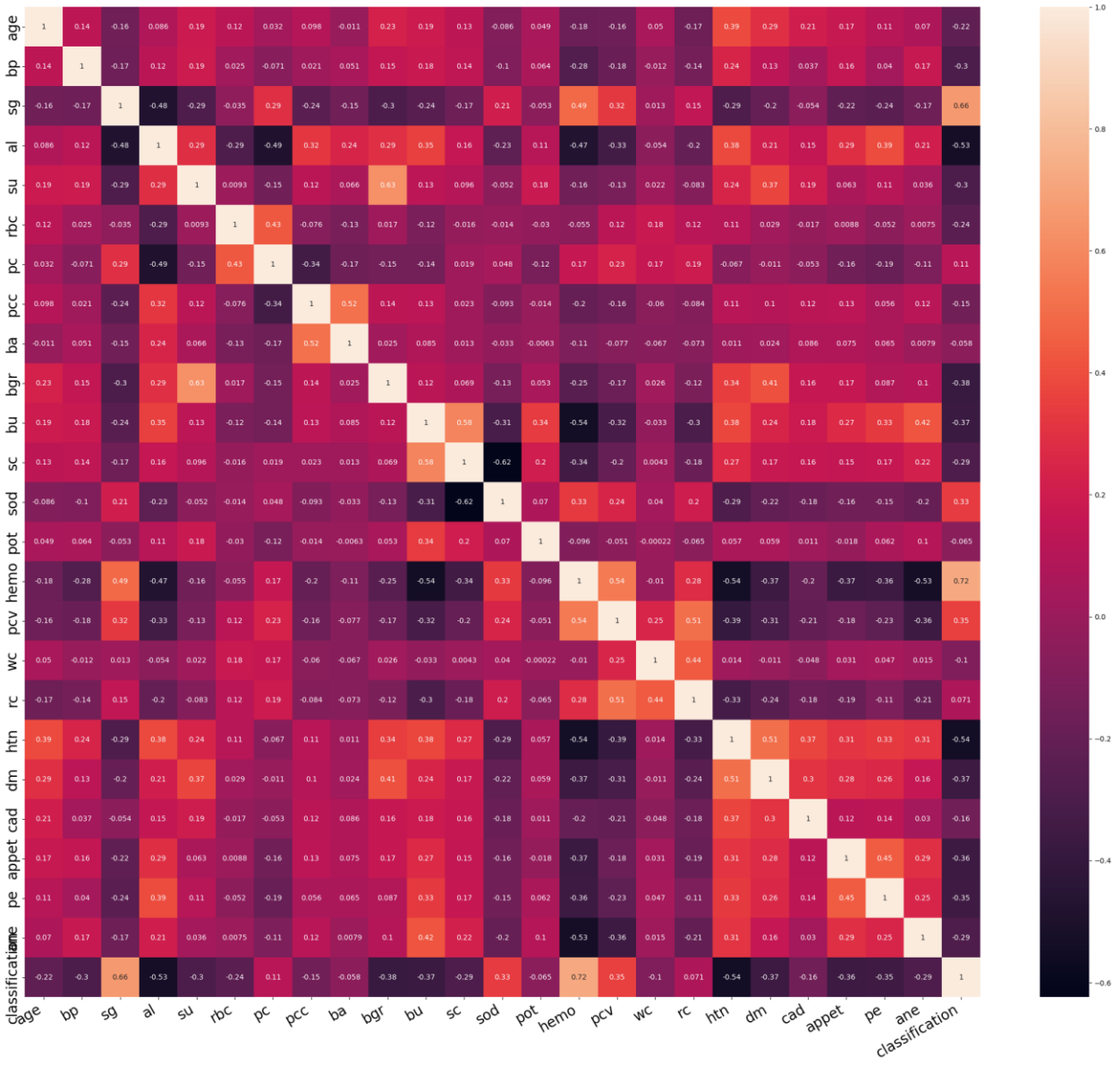

Figure 4: The heat map of the variables in the dataset.

Then apply the neutrosophic DEMATEL to show the weights of features in the dataset. The dataset has 28 variables so; we obtain the weights of 28 variables as shown in Figure 5. The least weight will be eliminated.

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

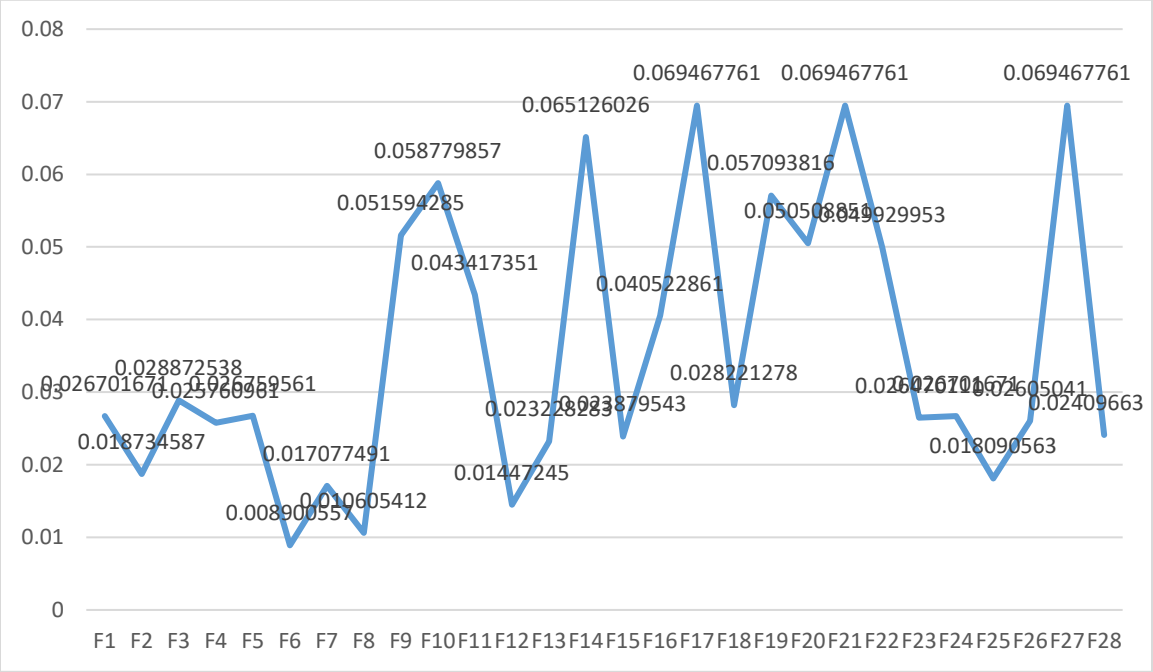

Figure 5: The weights of 28 variables in the dataset.

Then apply the machine learning models. Figure 6 shows the comparison between 3 models. The highest model is AB then GB, and finally the RF.

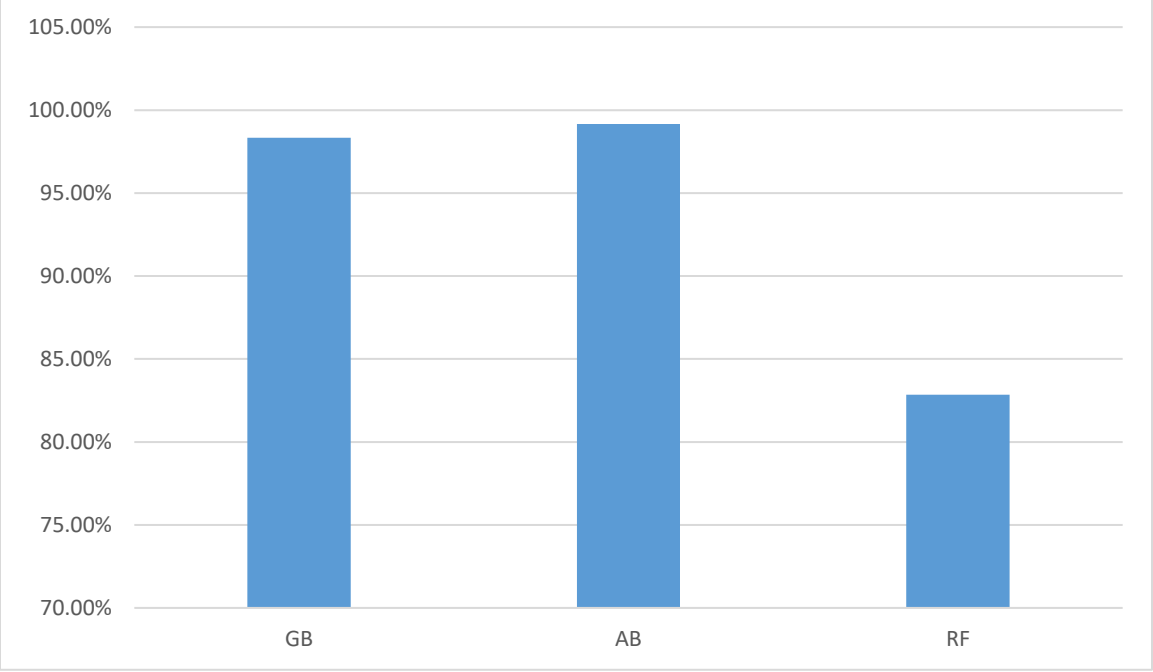

Figure 6: The comparative between GB, AB, and RF.

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

5. Conclusion

Reduction of kidney function occurs gradually over time in patients with chronic renal disease. Most people with this condition have no outward signs of illness. Because of the gravity of the problem, the medical community has turned to ML theory to craft an effective approach for the rapid detection and treatment of CKD. The current study describes a supervised learning-based technique to develop accurate models for forecasting the likelihood of CKD development, with a particular emphasis on probabilistic, tree-based, and ensemble learning-based approaches. The neutrosophic DEMATEL method is used in this paper as a feature selection method. The DEMATEL method was used to compute the weights of variables in the CKD dataset. Then the least weight will be removed from the dataset. Then three models namely GB, AB, and RF are applied to the CKD dataset. The AB has the highest accuracy, then GB and RF.

References

[1] M. U. Emon, R. Islam, M. S. Keya, and R. Zannat, "Performance Analysis of Chronic Kidney Disease through Machine Learning Approaches," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, IEEE, 2021, pp. 713–719.

[2] F. Sanmarchi, C. Fanconi, D. Golinelli, D. Gori, T. Hernandez-Boussard, and A. Capodici, "Predict, diagnose, and treat chronic kidney disease with machine learning: A systematic literature review," *J. Nephrol.*, pp. 1–17, 2023.

[3] P. Ghosh, F. M. J. M. Shamrat, S. Shultana, S. Afrin, A. A. Anjum, and A. A. Khan, "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, pp. 1–6.

[4] 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, pp. 1–13, 2019.

[5] M. Almasoud and T. E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors," *Int. J. Soft Comput. Its Appl.*, vol. 10, no. 8, 2019.

[6] 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, 2020.

[7] Mahmoud Ismail , Shereen Zaki, Neutrosophic MCDM Methodology to Select Best Industrial Arc Welding Robot, Neutrosophic and Information Fusion, Vol. 1 , No. 1 , (2023) : 08-16 (Doi : https://doi.org/10.54216/NIF.010101)

[8] Ahmed M. Ali, Ranking Renewable Energy Alternatives by using Triangular Neutrosophic Sets Integrated with MCDM, Neutrosophic and Information Fusion, Vol. 1 , No. 1 , (2023) : 17-26 (Doi : https://doi.org/10.54216/NIF.010102

[9] H. S. Kilic, P. Yurdaer, and C. Aglan, "A leanness assessment methodology based on neutrosophic DEMATEL," *J. Manuf. Syst.*, vol. 59, pp. 320–344, 2021.

[10] Ahmed Sleem , Ibrahim Elhenawy, An Interval Valued Neutrosophic Sets Integrated with the AHP MCDM Methodology to Assess the Station of 5G Network, Neutrosophic and Information Fusion, Vol. 1 , No. 1 , (2023) : 34-40 (Doi : https://doi.org/10.54216/NIF.010104)

[11] 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, 2019.

[12] Y. Amirgaliyev, S. Shamiluulu, and A. Serek, "Analysis of chronic kidney disease dataset by applying machine learning methods," in *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*, IEEE, 2018, pp. 1–4.

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023

[13] E. Dritsas and M. Trigka, "Machine learning techniques for chronic kidney disease risk prediction," *Big Data Cogn. Comput.*, vol. 6, no. 3, p. 98, 2022.

[14] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," in *2016 management and innovation technology international conference (MITicon)*, IEEE, 2016, p. MIT-80.

[15] K. R. A. Padmanaban and G. Parthiban, "Applying machine learning techniques for predicting the risk of chronic kidney disease," *Indian J. Sci. Technol.*, vol. 9, no. 29, pp. 1–6, 2016.

[16] Ahmed M. Ali , Ahmed Abdelmouty, Ranking Sustainable Technologies in Wave Energy: Multi-Criteria Decision-Making Approach under Neutrosophic Sets, American Journal of Business and Operations Research, Vol. 10 , No. 2 , (2023) : 14-22 (Doi : https://doi.org/10.54216/AJBOR.100202) [17] C. Veeramani, R. Venugopal, and S. A. Edalatpanah, "Neutrosophic DEMATEL approach for financial ratio performance evaluation of the NASDAQ Exchange," *Neutrosophic Sets Syst.*, vol. 51, pp. 766–782, 2022.

[18] Mahmoud Ismail , Mahmoud Ibrahiem, Multi-Criteria Decision-Making Approach based on Neutrosophic Sets for Evaluating Sustainable Supplier Selection in the Industrial 4.0, American Journal of Business and Operations Research, Vol. 7 , No. 2 , (2022) : 41-55 (Doi : https://doi.org/10.54216/AJBOR.070204)

[19] Abduallah Gamal , Nehal Nabil Mostafa, Sustainable Supplier Selection using Neutrosophic Sets and MCDM Framework, Neutrosophic and Information Fusion, Vol. 1 , No. 1 , (2023) : 27-33 (Doi : https://doi.org/10.54216/NIF.010103)

[20] Myvizhi M. , Samah I. Abdel Aal, A Multi-Criteria Decision-Making Model Based on Bipolar Neutrosophic Sets for the Selection of Battery Electric Vehicles, American Journal of Business and Operations Research, Vol. 10 , No. 2 , (2023) : 23-31 (Doi : https://doi.org/10.54216/AJBOR.100203)

[21] S. Revathy, B. Bharathi, P. Jeyanthi, and M. Ramesh, "Chronic kidney disease prediction using machine learning models," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 6364–6367, 2019.

[22] A. J. Aljaaf *et al.*, "Early prediction of chronic kidney disease using machine learning supported by predictive analytics," in *2018 IEEE congress on evolutionary computation (CEC)*, IEEE, 2018, pp. 1–9.

[23] S. Tekale, P. Shingavi, S. Wandhekar, and A. Chatorikar, "Prediction of chronic kidney disease using machine learning algorithm," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 7, no. 10, pp. 92–96, 2018.

[24] E. Dovgan *et al.*, "Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients," *PLoS One*, vol. 15, no. 6, p. e0233976, 2020.

[25] Hernández, Noel Batista, et al. "A new method to assess entrepreneurship competence in university students using based on plithogenic numbers and SWOT analysis." International Journal of Fuzzy Logic and Intelligent Systems, vol 21, pp.280-292, 2021.

[26] Ricardo, J. E., Fernández, A. J. R., Martínez, T. T. C., & Calle, W. A. C, "Analysis of Sustainable Development Indicators through Neutrosophic Correlation Coefficients." Neutrosophic Sets and Systems, vol 52 , 2022

DOI: https://doi.org/10.54216/IJNS.210116
Received: March 30, 2023 Revised: April 19, 2023 Accepted: May 14, 2023