Heart Disease Prediction using Neutrosophic C-Means Clustering Algorithm

Piedad Acurio Padilla[1,*], Evelyn Betancourt Rubio[1], Walter Vayas Valdiviezo[1], Mohammed k. Hassan[2]

[1]Universidad Regional Autónoma de los Andes, Ecuador

[2]Mechatronics department, Faculty of Engineering, Horus university-Egypt (HUE), Egypt

Emails: ua.piedadacurio@uniandes.edu.ec; us.evelynbr17@uniandes.edu.ec; ua.waltervayas@uniandes.edu.ec; mkhassan@horus.edu.eg

Abstract

Heart disease, often known as cardiovascular illness, encompasses a broad range of heart-related disorders and has emerged as the leading cause of mortality during the last few decades everywhere in the globe. Numerous hazards are linked to cardiovascular disease, and timely, effective, and practical methods for making an early diagnosis are required for effective and efficient treatment. In this study, we describe a novel clustering technique for data that is unreliable clustering called neutrosophic c-means (NCM), which draws inspiration from both fuzzy c-means and the neutrosophic set architecture. The NCM is used to predict heart disease. There are four different databases included in the collection, all of which were created in 1988: Cleveland, Hungary, Switzerland, and Long Beach V. There are 76 qualities total, such as the anticipated characteristic, however only 14 have been used in any of the published trials.

Keywords: Neutrosophic C-Means Clustering; Clustering; Heart Disease Prediction.

1. Introduction

In the past ten years, cardiovascular disease has been the leading cause of mortality globally. According to the World Health Organization, cardiovascular disease is the leading cause of mortality globally, accounting for roughly 17.9 million fatalities annually. Eighty percent of these deaths are attributable to coronary artery disease and cerebral stroke[1], [2]. Countries with low and intermediate incomes tend to have the highest mortality rates. Heart disease has several causes, including lifestyle choices, family history, and even one's occupation[3]–[5]. Many physiological variables, including obesity, hypertension, high cholesterol levels, and preexisting cardiac diseases, as well as behavioral risk factors including smoking, excessive alcohol and caffeine usage, stress, and physical inactivity, predispose individuals to cardiovascular disease. Diagnosing heart disease quickly, accurately, and early is critical for initiating lifesaving preventative steps[6]–[8].

Iteratively minimizing a cost function yields membership levels of the data, with the caveat that the sum of membership levels across the clusters for every information must equal 1. The fuzzy c-means algorithm (FCM) is a popular fuzzy clustering algorithm[9], [10]. The FCM technique has various shortcomings, including the fact that it attempts to minimize intra-cluster variance as well, sharing the same issues as the k-means approach, where the minimum is a local minimum and the results are very sensitive to the initializations. The FCM method is also very vulnerable to background noise[11]–[13]. There might be a large population of "noise points." Because it is sensitive

DOI: https://doi.org/10.54216/IJNS.210206
Received: February 14, 2023 Revised: April 21, 2023 Accepted: May 27, 2023

to the distance measure used, the method known as FCM cannot differentiate between extremely probable and highly unlikely outcomes[14], [15].

There's a new school of thought called neutrophilic set (NS) that studies the genesis, range, and impact of neutralities, as well as how they interact with various ideational spectrum[16],[17]. Neutrosophic sets consist of three elements: the element E, its antithesis Anti-E, and the neutrality Neut-E, which is neither E nor Anti-E[18]. Three memberships are used to quantify the degree to which E is true, uncertain, or false[19], [20]. Neutrosophic theory, built on this defining feature, offers a potent instrument for dealing with uncertainty and has been put to use in a wide range of fields[21].

When it comes to border pixels and anomalies, the Neutrosophic C-means Clustering Method (NCM) offers a significant improvement over the standard FCM. However, the NCM model is incapable of producing reliable segmentation results in the absence of geographical information[22], [23].

Millions of individuals suffer from cardiovascular disease, and it continues to be the leading global killer. Professional, trustworthy, and computer-aided medical diagnosis is essential for lowering the successful price of diagnostic testing. The software technique known as "NCM" allows computers to construct and categorize a wide range of properties. In this study, we use clustering methods to the problem of cardiac illness prognosis.

2. Neutrosophic C-Means Clustering

Conventional approaches to cluster analysis just characterize the extent to which each group exists. Indeed, for certain samples, especially those at the transition zone between two groups, it is not always clear to what category they should be assigned. The centers of the various clusters will be off if an unbreakable partition is formed[24]–[26].

Taking into account clustering with uncertainty, we create an entirely novel function with objectives and membership as[27],[28].

$$
\begin{aligned}
J(A,B,C,U) = {} & \sum_{i=1}^{N}\sum_{j=1}^{U}(w_1 A_{ij})^m || \ x_i - u_j \ ||^2 + \sum_{i=1}^{N}\sum_{j=1}^{2}(w_2 B_{2ij})^m || \ x_i - u_{2j} \ ||^2 \\
& + + \sum_{i=1}^{N}\sum_{j=1}^{3}(w_3 B_{3ij})^m || \ x_i - u_{3j} \ ||^2 \\
& + \sum_{i=1}^{N}\sum_{j=1}^{4}(w_4 B_{4ij})^m || \ x_i - u_{4j} \ ||^2 + + \sum_{i=1}^{N}\sum_{j=1}^{5}(w_5 B_{5ij})^m || \ x_i - u_{5j} \ ||^2 \\
& + + \sum_{i=1}^{N}\sum_{j=1}^{6}(w_6 B_{6ij})^m || \ x_i - u_{6j} \ ||^2 + + \sum_{i=1}^{N}\sum_{j=1}^{7}(w_7 B_{7ij})^m || \ x_i - u_{7j} \ ||^2 + \cdots \\
& + \sum_{i=1}^{N}\sum_{j=1}^{U}(w_U B_{Uij})^m || \ x_i - u_{Uj} \ ||^2 + \sum_{i=1}^{N} \vartheta^2 (w_{U+1} \, C_i)^2 \qquad (1)
\end{aligned}
$$

Where u_{4j} refers to mean of any two groups, ϑ used to manage the amount of objects.

We can simply the previous equation into a simple equation as:

DOI: https://doi.org/10.54216/IJNS.210206
Received: February 14, 2023 Revised: April 21, 2023 Accepted: May 27, 2023

$$J(A,B,C,U) = \sum_{i=1}^{N}\sum_{j=1}^{U}(w_1 A_{ij})^m || \ x_i - u_j \ ||^2 + \sum_{i=1}^{N}\sum_{j=1}^{\frac{U}{2}}(w_2 B_{2ij})^m || \ x_i - u_{i\,max} \ ||^2 + \sum_{i=1}^{N}\vartheta^2(w_3 \ C_i)^m \quad (2)$$

$$u_{i\,max} = \frac{U_{pi}+U_{qi}}{2} \quad (3)$$

$$U_{pi} = \arg\max(A_{ij}) \quad (4)$$

$$j = 1,2,\ldots.U$$

$$U_{pi} = \arg\max(A_{ij}) \quad (5)$$

$$j \neq p_i \cap j = 1,2,\ldots.U$$

$$\sum_{j=1}^{U} A_{ij} + B_{ij} + C_{ij} = 1 \quad (6)$$

3. Analysis and Results

The purpose of this study is to determine the likelihood that a certain patient would acquire cardiovascular disease. The focus of this study was NCM. First make data description of dataset as shown in Table 1. From Table 1, the number of patients are 302. The 25% of target class are 0. The 50% of target class are 1.

Table 1: Data description of dataset.

	age	trestbps	chol	thalach	oldpeak	target
count	302.00000	302.000000	302.000000	302.000000	302.000000	302.000000
mean	54.42053	131.602649	246.500000	149.569536	1.043046	0.543046
td	9.04797	17.563394	51.753489	22.903527	1.161452	0.498970
min	29.00000	94.000000	126.000000	71.000000	0.000000	0.000000
25%	48.00000	120.000000	211.000000	133.250000	0.000000	0.000000
50%	55.50000	130.000000	240.500000	152.500000	0.800000	1.000000
75%	61.00000	140.000000	274.750000	166.000000	1.600000	1.000000
max	77.00000	200.000000	564.000000	202.000000	6.200000	1.000000

Figure 1. Shows the analysis of the dataset by the boxplot.

DOI: https://doi.org/10.54216/IJNS.210206
Received: February 14, 2023 Revised: April 21, 2023 Accepted: May 27, 2023

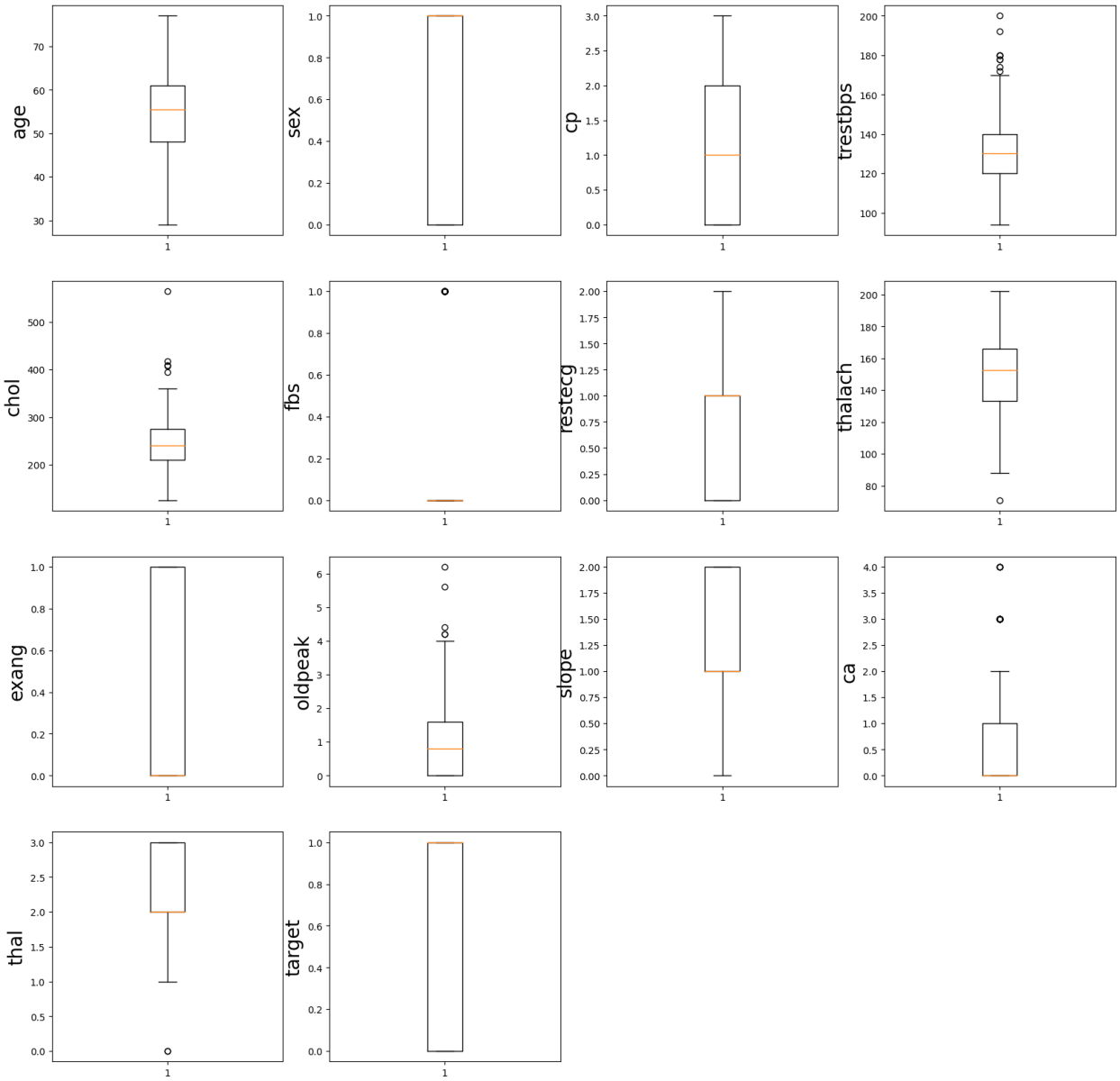

Figure 1: The boxplot analysis of the dataset.

Table 2 shows the classes with true and classes with NCM. In cell 0 to 4 the NCM predicted 0 class and the output is 0. From cell 5-9 the NCM predicted with 1 class and the true class is 1.

Table 2: The predicted class with NCM and true class.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	NCM
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0	0
1	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0	0
2	55	1	0	160	289	0	0	145	1	0.8	1	1	3	0	0
3	46	1	0	120	249	0	0	144	0	0.8	2	0	3	0	0
4	54	1	0	122	286	0	0	116	1	3.2	1	2	2	0	0
5	58	0	0	100	248	0	0	122	0	1.0	1	0	2	1	1
6	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1	1
7	51	0	2	140	308	0	0	142	0	1.5	2	1	2	1	1

DOI: https://doi.org/10.54216/IJNS.210206
Received: February 14, 2023 Revised: April 21, 2023 Accepted: May 27, 2023

8	50	0	1	120	244	0	1	162	0	1.1	2	0	2	1	1
9	58	1	2	140	211	1	0	165	0	0.0	2	0	2	1	1

Figure 2 shows the number of classes 0 and 1 to target class. The class 0 is less than the class 1.

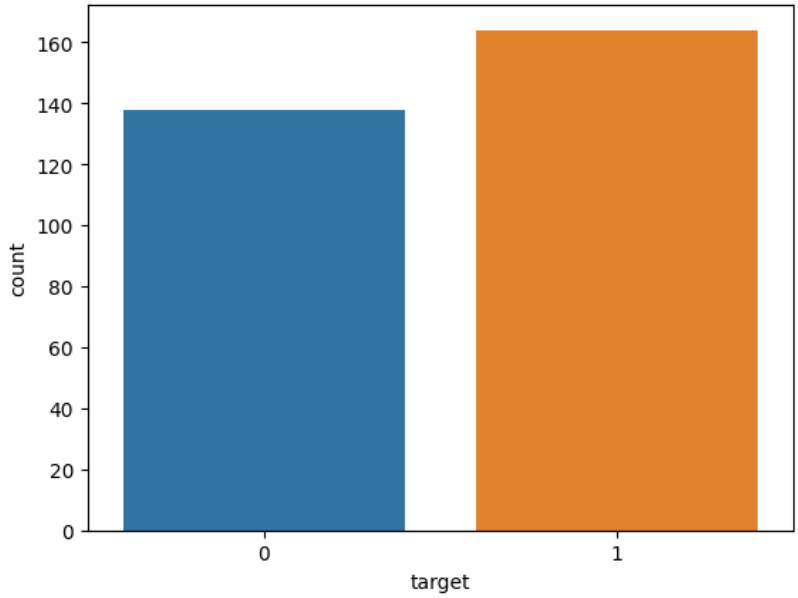

Figure 2: The target class.

4. **Conclusion**

Heart attacks and other consequences of heart disease are a serious hazard to human life. The relevance of the NCM approach in illness prediction is based on its potential for a high accuracy rate. Here, we put the NCM method for heart disease diagnosis to the test using a real-world dataset. NCM has been offered as a means to segment data, particularly material that is imprecise or ambiguous. The extent of each group is all that can be described using the standard techniques. Some samples lie on the cusp between two categories, making it hard to assign a definitive label to them. Furthermore, calculating memberships leads to off-target group zeniths. To counteract these shortcomings of conventional partitioning methods, NCM was developed. The suggested NCM method is evaluated on two different tasks: data clustering and picture segmentation.

References

[1] L. Parthiban and R. Subramanian, "Intelligent heart disease prediction system using CANFIS and genetic algorithm," *Int. J. Biol. Biomed. Med. Sci.*, vol. 3, no. 3, 2008.

[2] A. Singh and R. Kumar, "Heart disease prediction using machine learning algorithms," in *2020 international conference on electrical and electronics engineering (ICE3)*, IEEE, 2020, pp. 452–457.

[3] N. Bhatla and K. Jyoti, "An analysis of heart disease prediction using different data mining techniques," *Int. J. Eng.*, vol. 1, no. 8, pp. 1–4, 2012.

DOI: https://doi.org/10.54216/IJNS.210206
Received: February 14, 2023 Revised: April 21, 2023 Accepted: May 27, 2023

[4] A. Yadav, L. Gediya, and A. Kazi, "Heart disease prediction using machine learning," *Int. Res. J. Eng. Technol. (IRJET*, vol. 8, no. 09, 2021.

[5] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in *IOP conference series: materials science and engineering*, IOP Publishing, 2021, p. 12072.

[6] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," *Int. J. Comput. Appl.*, vol. 17, no. 8, pp. 43–48, 2011.

[7] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE access*, vol. 7, pp. 81542–81554, 2019.

[8] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Comput. Sci.*, vol. 1, pp. 1–6, 2020.

[9] A. M. Anter, A. E. Hassanien, M. A. A. ElSoud, and M. F. Tolba, "Neutrosophic sets and fuzzy c-means clustering for improving ct liver image segmentation," in *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014*, Springer, 2014, pp. 193–203.

[10] M. N. Qureshi and M. V. Ahamad, "An improved method for image segmentation using K-means clustering with neutrosophic logic," *Procedia Comput. Sci.*, vol. 132, pp. 534–540, 2018.

[11] J. Estupiñán Ricardo, J. J. Domínguez Menéndez, I. F. Barcos Arias, J. M. Macías Bermúdez, and N. Moreno Lemus, "Neutrosophic K-means for the analysis of earthquake data in Ecuador," *Neutrosophic Sets Syst.*, vol. 44, no. 1, p. 29, 2021.

[12] B. B. Yousif, M. M. Ata, N. Fawzy, and M. Obaya, "Toward an optimized neutrosophic K-means with genetic algorithm for automatic vehicle license plate recognition (ONKM-AVLPR)," *IEEE Access*, vol. 8, pp. 49285–49312, 2020.

[13] A. S. Ashour, A. R. Hawas, Y. Guo, and M. A. Wahba, "A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images," *Signal, Image Video Process.*, vol. 12, no. 7, pp. 1311–1318, 2018.

[14] F. Zamani, M. H. Olyaee, and A. Khanteymoori, "NCMHap: a novel method for haplotype reconstruction based on Neutrosophic c-means clustering," *BMC Bioinformatics*, vol. 21, pp. 1–15, 2020.

[15] B. Ji, X. Hu, F. Ding, Y. Ji, and H. Gao, "An effective color image segmentation approach using superpixel-neutrosophic C-means clustering and gradient-structural similarity," *Optik (Stuttg).*, vol. 260, p. 169039, 2022.

[16] H. K. M. Ahmed Abdelhafeez, "Enhance the Performance of Bus Rapid Transit (BRT) through the Evaluation of Alternatives under an Integrated MCDM Neutrosophic Environment," *Neutrosophic Inf. Fusion*, vol. 1, no. 2, pp. 08–15, 2023, doi: 10.54216/NIF.010201.

[17] Y. Guo, R. Xia, A. Şengür, and K. Polat, "A novel image segmentation approach based on neutrosophic c-means clustering and indeterminacy filtering," *Neural Comput. Appl.*, vol. 28, pp. 3009–3019, 2017.

[18] A. M. Ali, "Ranking Renewable Energy Alternatives by using Triangular Neutrosophic Sets Integrated with MCDM," *Neutrosophic Inf. Fusion*, vol. 1, no. 1, pp. 17–26, 2023, doi: 10.54216/nif.010102.

[19] Z. Lu, Y. Qiu, and T. Zhan, "Neutrosophic C-means clustering with local information and noise distance-based kernel metric image segmentation," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 269–276, 2019.

[20] J. E. Ricardo, J. J. D. Menéndez, I. F. B. Arias, J. M. M. Bermúdez, and N. M. Lemus, "Neutrosophic K-means for the analysis of earthquake data in Ecuador," *Neutrosophic Sets Syst.*, vol. 44, pp. 255–262, 2021.

[21] I. E. Ahmed Sleem, "An Interval Valued Neutrosophic Sets Integrated with the AHP MCDM Methodology to Assess the Station of 5G Network," *Neutrosophic Inf. Fusion*, vol. 1, no. 1, pp. 34–40, 2023, doi: 10.54216/NIF.010104.

DOI: https://doi.org/10.54216/IJNS.210206
Received: February 14, 2023 Revised: April 21, 2023 Accepted: May 27, 2023

[22] Y. Guo and A. Sengur, "NCM: Neutrosophic c-means clustering algorithm," *Pattern Recognit.*, vol. 48, no. 8, pp. 2710–2724, 2015.

[23] Y. Akbulut, A. Şengür, Y. Guo, and K. Polat, "KNCM: Kernel neutrosophic c-means clustering," *Appl. Soft Comput.*, vol. 52, pp. 714–724, 2017.

[24] Y. Guo and A. Sengur, "NECM: Neutrosophic evidential c-means clustering algorithm," *Neural Comput. Appl.*, vol. 26, pp. 561–571, 2015.

[25] J. L. Zeema and D. F. X. Christopher, "Evolving optimized neutrosophic C means clustering using behavioral inspiration of artificial bacterial foraging (ONCMC-ABF) in the prediction of Dyslexia," *J. King Saud Univ. Inf. Sci.*, vol. 34, no. 5, pp. 1748–1754, 2022.

[26] H. D. Cheng, Y. Guo, and Y. Zhang, "A novel image segmentation approach based on neutrosophic set and improved fuzzy c-means algorithm," *New Math. Nat. Comput.*, vol. 7, no. 01, pp. 155–171, 2011.

[27] RICARDO, J. E.; FERNÁNDEZ, A. J. R.; VÁZQUEZ, M. Y. L. Compensatory Fuzzy Logic with Single Valued Neutrosophic Numbers in the Analysis of University Strategic Management. International Journal of Neutrosophic Science, p. 151-159, 2022.

[28] LEYVA, M.; DEL POZO, P.; PEÑAFIEL, A. Neutrosophic DEMATEL in the Analysis of the Causal Factors of Youth Violence. International Journal of Neutrosophic Science, vol. 18, no 3, p. 199-207, 2022.

DOI: https://doi.org/10.54216/IJNS.210206
Received: February 14, 2023 Revised: April 21, 2023 Accepted: May 27, 2023