Hybridization of Neutrosophic Logic with Quasi-Oppositional Chimp Optimization based Data Classification Model

Sundus Naji AL-Aziz[1,*], Reem Atassi[2], Abd Al-Aziz Hosni El-Bagoury [3]

[1] Department of Mathematical Sciences, Faculty of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[2] Faculty of Computer Information System, Higher Colleges of Technology, UAE

[3] Higher Institute of Engineering and Technology, El-Mahala El-Kobra, Egypt.

Emails: snalaziz@pnu.edu.sa[1]; ratassi@hct.ac.ae[2], azizhel2013@yahoo.com[3]

Abstract

Data classification is the procedure of investigating structured or unstructured data and forming it into distinct classes depending upon file types, size, etc. It assist the organizations to derive important solutions based on the data and helps decision making process. The computational intelligence techniques such as neural computing, fuzzy logic, machine learning, etc. can be used to design effective data classification models. This study offers a Hybridization of Neutrosophic Logic with Quasi-Oppositional Chimp Optimization based Data Classification (HNLQOCO) model. The presented HNLQOCO algorithm aims to integrate the concepts of NL and QOCO algorithm for improved data classification outcomes. Besides, the QOCO algorithm is designed by incorporating the concepts of quasi oppositional based learning (QOBL) with traditional chimp optimization algorithm (COA). Here, the NL is applied to represent various kinds of knowledge and the QOCO algorithm is applied to tune the produced NS rules. The experimental result analysis of the HNLQOCO model is tested using three benchmark medical dataset. The obtained results reported the significant performance of the HNLQOCO model over the other methods.

Keywords: Data classification, Neutrosophic Logic; Chimp optimization algorithm; Rule generation; QOBL.

1. Introduction

Because of the progression in innovation there are enormous measure of natural data. It is tedious to view or concentrate the required data [1]. In such a circumstance we are deprived to foster a technique which is helpful to get the fundamental data. Since there are enormous measure of data dynamic cycle is monotonous. To defeat these entanglements, the idea of data mining (DM) is utilized [2]. The strategies of DM will assist the clients with gaining the fundamental data. DM is broadly utilized in a wide assortment of uses, like clinical finding, designated showcasing, assessing the portions of TV crowds, monetary anticipating item configuration, computerized reflection, investigation of natural mixtures, Mastercard extortion recognition [3]. DM is a non-trifling extraction of certain, already obscure and likely valuable data about data [3]. To put it plainly, it is a course of dissecting data according to alternate point of view and assembling the information from it [4]. The found information can be utilized for various applications for instance medical care industry [5]. These days medical care industry produces enormous measure of data about patients, illness analysis and so on DM gives a bunch of strategies to find concealed designs from data. A significant test confronting Healthcare

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

industry is nature of administration. Nature of administration infers diagnosing illness accurately and gives successful medicines to patients. Unfortunate conclusion can prompt grievous results which are unsuitable. Classification involved two stages in the initial step a model is built in light of some preparation data set, in seconds step the model is utilized to group an obscure tuple into a class name [6].

The classification is one of the significant errands in DM. The thought behind this is to arrange the given data records into one of the numerous potential cases which are known as of now [7]. Classification assignments can utilize any one system. Assuming that the data are characterized without taking a gander at the preparation data, this sort of classification is known as priori classification [8]. Yet, in talk assuming the data were ordered with the assistance of preparing data this is known as posteriori classification. There are a few applications for Machine Learning (ML), the most critical of which is DM. Individuals are frequently inclined to committing errors during investigations or, potentially, while attempting to lay out connections between various highlights [9]. This makes it challenging for them to track down answers for specific issues. AI can frequently be effectively applied to these issues, working on the productivity of frameworks and the plans of machines. Various ML applications include undertakings that can be set up as regulated [10].

The authors in [11] presented a big health application scheme based optimum artificial neural network (OANN) for heart disease diagnoses, that is taken into account as a serious disease worldwide. The presented method involves a two major methods i.e., teaching and learning based optimization (TLBO) approach for ANN, named (TLBO-ANN) and distance based misclassified instance removal (DBMIR). The authors in [12] proposed a parallel classifier method based random projection and Bagging- SVM for processing higher-dimension information. The technique initially employs random projection for projecting the input information into the lower-dimension region. Next, employed the Bagging technique for constructing trained data subset and employed SVM for training the subset in parallel and generating different sub-classifiers.

The authors in [13] examine nine ML approaches for understanding the performance using Credibility-Based Fake News Detection. We employ a benchmark data set with feature relates to the credibility of news publisher. This feature is analyzed by this algorithm. The outcomes of this experiment is analyzed by the four estimation models. The authors in [14] employed k-means approach as the clustering technique for FS. The presented technique is utilized for categorizing feature that have similar features in single cluster, such that microarray information redundancy is detached. The outcome of clustering is rated by the Relief approach so that the optimal scoring component for all the clusters are attained. Each optimal element of cluster is carefully chosen and employed as a feature in the classifier method.

This study offers a Hybridization of Neutrosophic Logic with Quasi-Oppositional Chimp Optimization based Data Classification (HNLQOCO) model. The presented HNLQOCO algorithm aims to integrate the concepts of NL and QOCO algorithm for improved data classification outcomes. Besides, the QOCO algorithm is designed by incorporating the concepts of quasi oppositional based learning (QOBL) with traditional chimp optimization algorithm (COA). Here, the NL is applied to represent various kinds of knowledge and the QOCO algorithm is applied to tune the produced NS rules. The experimental result analysis of the HNLQOCO model is tested using three benchmark medical dataset.

2. The Proposed Model

This study has provided a new HNLQOCO algorithm for accurate data classification. The presented HNLQOCO algorithm aims to integrate the concepts of NL and QOCO algorithm for improved data classification outcomes. Besides, the QOCO algorithm is designed by incorporating the concepts of quasi QOBL with traditional COA.

2.1 Overview of NL

Neutrosophy is a subdivision of philosophy, handling neutralities, with ancient roots, along with the interaction with distinct ideational spectra. It is a logic where all the percentage is assessed for having the proportion of truth in a sub set T, the proportion of indeterminacy in a sub set I, and the proportion of falsity in a sub set F, whereas $T, I,$ and F denotes non-standard or standard real subset of $]-0, 1+[$ where $]-0, 1+[$ is non-standard unit intervals.

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

$$sup\,T = t_{sup}, inf\,T = t_{\inf} \tag{1}$$

$$sup\,I = i\_sup, inf\,I = i\_inf \tag{2}$$

$$sup\,F = t\_sup, inf\,F = f\_inf \tag{3}$$

and

$$n\_sup = t\_sup + i\_\,sup + f\_sup \tag{4}$$

$$n\_inf = t\_inf + i\_inf + f\_inf \tag{5}$$

T, I, and F denotes a neutrosophic mechanisms, represent correspondingly the truth, indeterminacy, and falsehood value refers to neutrosophic statistics, neutrosophic probability, neutrosophic logic, neutrosophic set, neutrosophy. In real time application, it is easy to employ standard real intervals [0,1] for T, land F rather than non-standard unit interval. Each factor states in neutrosophic logic are intrinsic to individual thought. It is uncommon that we tend to judge or conclude in definite environment. Inaccuracy of human system might be because of the deficiency of knowledge that human obtains from the outside world [15]. Fig. 1 shows the neutrosophic logic diagram [16].

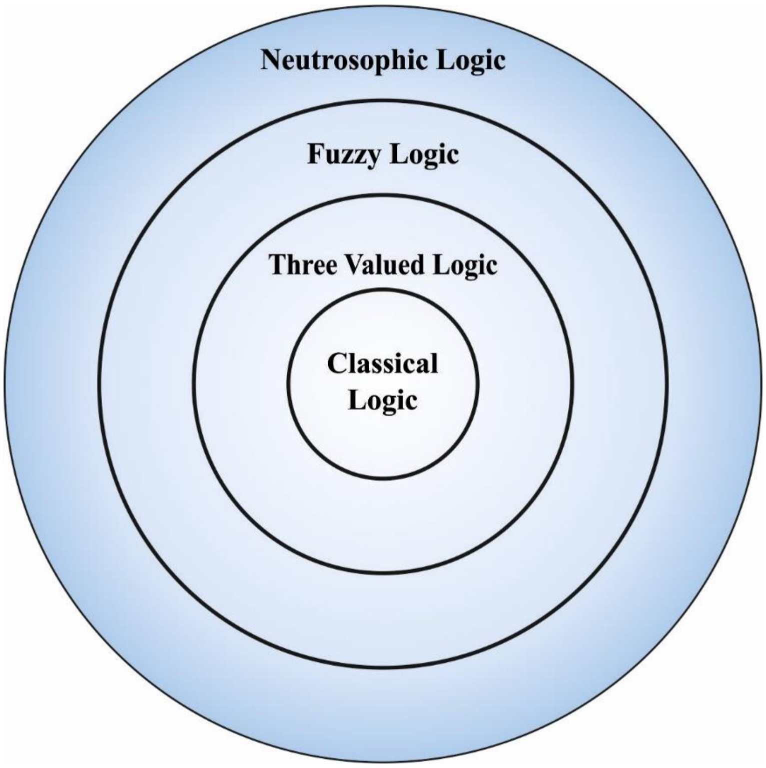

Figure 1: Neutrosophic logic diagram

2.2 Process involved in HNLQOCO Algorithm

In this work, the NL is applied to represent various kinds of knowledge and the QOCO algorithm is applied to tune the produced NS rules. COA is a current multi-group metaheuristic approach depending on the behavior of chimp hunting [17]. This approach comprises four approaches of hunting, i.e., attacking, driving, chasing, and blocking. The major phase of the COA approach is given below and is shown in Fig. 2:

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

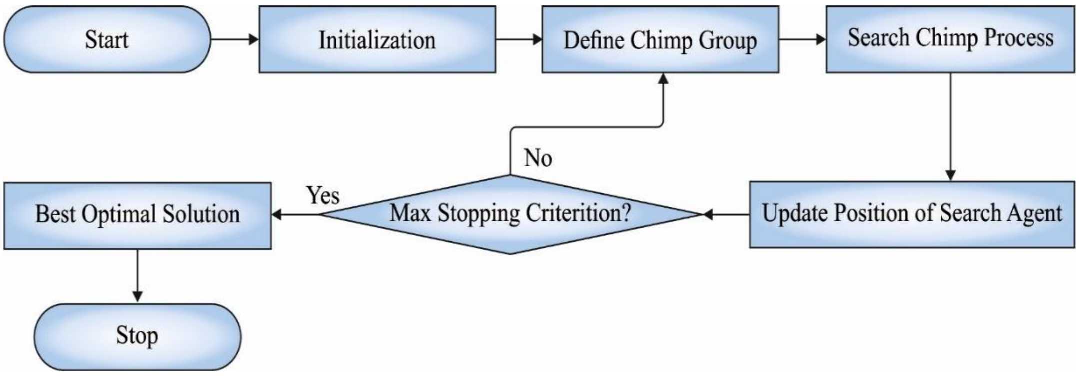

Figure 2: Process involved in COA

Exploration stage comprises of blocking, driving and chasing the prey that is characterized by:

$$x_{chimp}(t+1) = x_{prey}(t) - a\hat{A}\cdot d \tag{6}$$

In whichx_{prey} and x_{chimp} denotes the location vector of prey and chimp, correspondingly. The distance between the prey and chimp is represented as d:

$$d = |c.x_{prey}(t) - m \cdot x_{chimp}(t)| \tag{7}$$

Whereas , c, and m denotes coefficient vector and chaotic variable.

The exploitation stage amongst chimps is associated with the attacking method that is implemented by assuming the location of the driver, attacker, barrier and chaser chimps. They separated from the prey and later grouped to attacks as follows:

$$d_{attacker} = |c_1.x_{attacker} - m_1.x|, d_{barrier} = |c_2.x_{barrier} - m_2.x|$$

$$d_{chaser} = |c_3.x_{chaser} - m_3.x|, d_{driver} = |c_4.x_{driver} - m_4.x| \tag{8}$$

Furthermore, the four optimal solutions to evaluate the prey location are estimated by:

$$x_1 = |x_{attacker} - a_1.d_{attacker}|, x_2 = |x_{barrier} - a_2.d_{barrier}|$$

$$x_3 = |x_{chaser} - a_3.d_{chaser}|, x_4 = |x_{driver} - a_4.d_{driver}| \tag{9}$$

At last, the optimal location is estimated by the average value of attained four locations. The major step of the COA algorithms is shown below [18]:

- Initializing the chimp population.
- Initializing the , c, and m constants.
- Estimate the location of chimp.
- Divide chimps arbitrarily into separate group.
- Allocate the optimal chimp in all the groups.
- Upgrade the chimp location.
- Show the optimal chimp location and upgrade the fitness function until the maximal amount of iterations is obtained.

In order to improve the outcomes of the COA, the QOCOA is derived by the use of QOBL. In previous to emphasis on OBL, quasi-opposite and opposite values must be determined.

Opposite value: Assume $x \in [a, b]$ as a real number. In opposite value x^o is demonstrated as follows

$$x^o = a + b - x. \tag{10}$$

Similarly, the explanation is standardized as higher dimension as follows.

Opposite point: Assume $P(x_1, x_2, x_n)$ denotes a point in n-dimension space, whereas $x_j \in [a_j, b_j]; i = \{1, 2, , n\}$. The opposite point $OP(x_1^o, x_2^o, x_n^o)$ is demonstrated:

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

$$x_i^o = a_i + b_i - x_i. \tag{11}$$

Quasi-oppositional number and quasi-oppositional point: QOBL guarantee that a quasi-opposite point is near the optimum solution in contrast to opposite point.

Quasi-oppositional value: Assume $x \in [a, b]$ represent a real number. The quasi-opposite number x^{qo} is illustrated as follows

$$x^{qo} = rand(c, x^o), \tag{12}$$

Whereas $rand(c, x^o)$ denotes an arbitrary number in c and x^o; c implies the range $[a, b]$ as follows,

$$c = \frac{a+b}{2}. \tag{13}$$

Quasi-oppositional point: consider $P(x_1, x_2, x_n)$ denotes a point in n-dimension space, in which $x_i \in [a_i, b_i]; i = \{1, 2, \dots, n\}$. The quasi-oppositional point $QOP(x_1^{qo}, x_2^{qo}, \dots, x_n^{qo})$ is characterized by:

$$x_i^{qo} = rand(c_i, x_i^o), i = \{1, 2, , n\}, \tag{14}$$

Now

$$c_i = \frac{a_i + b_i}{2}. \tag{15}$$

3. Performance Validation

The simulation analysis of HNLQOCO model is tested using benchmark dataset from UCI repository [19]. Table 1 and Fig. 3 demonstrates the classifier outcomes of the HNLQOCO model on three distinct dataset. The experimental outcomes reported that the HNLQOCO model has classified all the class labels with maximum classification performance. For instance, on iris dataset, the HNLQOCO model has resulted $prec_n$, $reca_l$, $accu_y$, and $spec_y$ of 99.84%, 99.60%, 99.78%, and 99.70% respectively. In addition, on wine dataset, the HNLQOCO model has offered $prec_n$, $reca_l$, $accu_y$, and $spec_y$ of 97.83%, 96.33%, 96.89%, and 98.99% respectively. Also, on wdbc dataset, the HNLQOCO model has gained $prec_n$, $reca_l$, $accu_y$, and $spec_y$ of 97.56%, 96.87%, 96.39%, and 96.82% respectively.

Table 1: Classification outcomes of HNLQOCO model

Class Labels	Precision	Recall	Accuracy	Specificity
Iris Dataset				
setosa	99.84	99.64	99.79	99.70
versicolor	99.84	99.53	99.75	99.70
viginica	99.84	99.64	99.81	99.70
Average	99.84	99.60	99.78	99.70
Wine Dataset				
Class-1	99.65	95.96	96.83	99.74
Class-2	97.44	96.52	97.36	98.48
Class-3	96.40	96.52	96.48	98.73
Average	97.83	96.33	96.89	98.99
Wdbc Dataset				
Class M	97.89	94.53	94.95	99.06
Class B	97.24	99.22	97.83	94.58
Average	97.56	96.87	96.39	96.82

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

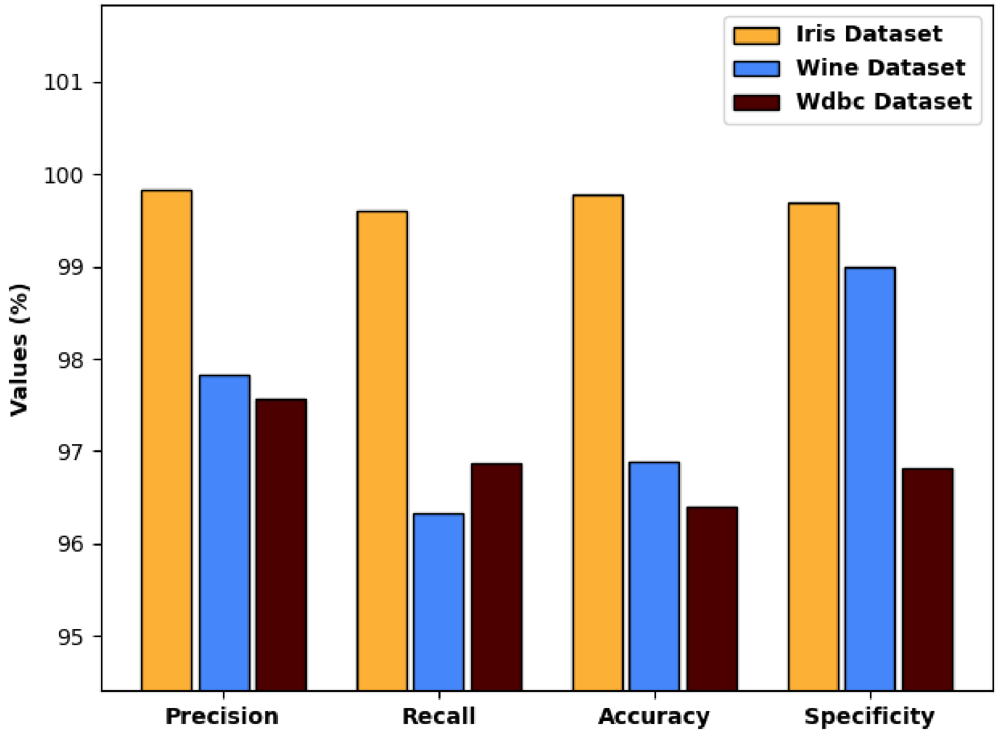

Figure 3: Overall classifier results of HNLQOCO model

Table 2 and Fig. 4 depicts a comparative study of the HNLQOCO model with other models on iris dataset. The results indicated that the HNLQOCO model has outperformed other methods on iris dataset. The NS model has resulted to lower precision values of 99.32%, 95.27%, and 92.20% on the classification of setosa, versicolor, and viginica classes respectively. Followed by, the GNS model has attained slightly enhanced precision values of 99.47%, 99.69%, and 99.47% on the classification of setosa, versicolor, and viginica classes respectively. However, the HNLQOCO model has reached maximum precision values of 99.84%, 99.84%, and 99.84% on the classification of setosa, versicolor, and viginica classes respectively.

Table 2 Comparison study of HNLQOCO model on iris dataset

Iris Dataset			
Class Labels	**Neutroshophic**	**Gen. Neutrosophic**	**HNLQOCO**
Precision			
setosa	99.32	99.47	99.84
versicolor	95.27	99.69	99.84
viginica	92.20	99.47	99.84
Recall			
setosa	99.07	98.96	99.64
versicolor	91.66	98.96	99.53
viginica	95.42	99.07	99.64
Specificity			
setosa	99.52	99.46	99.70
versicolor	97.66	99.40	99.70
viginica	95.86	99.40	99.70

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

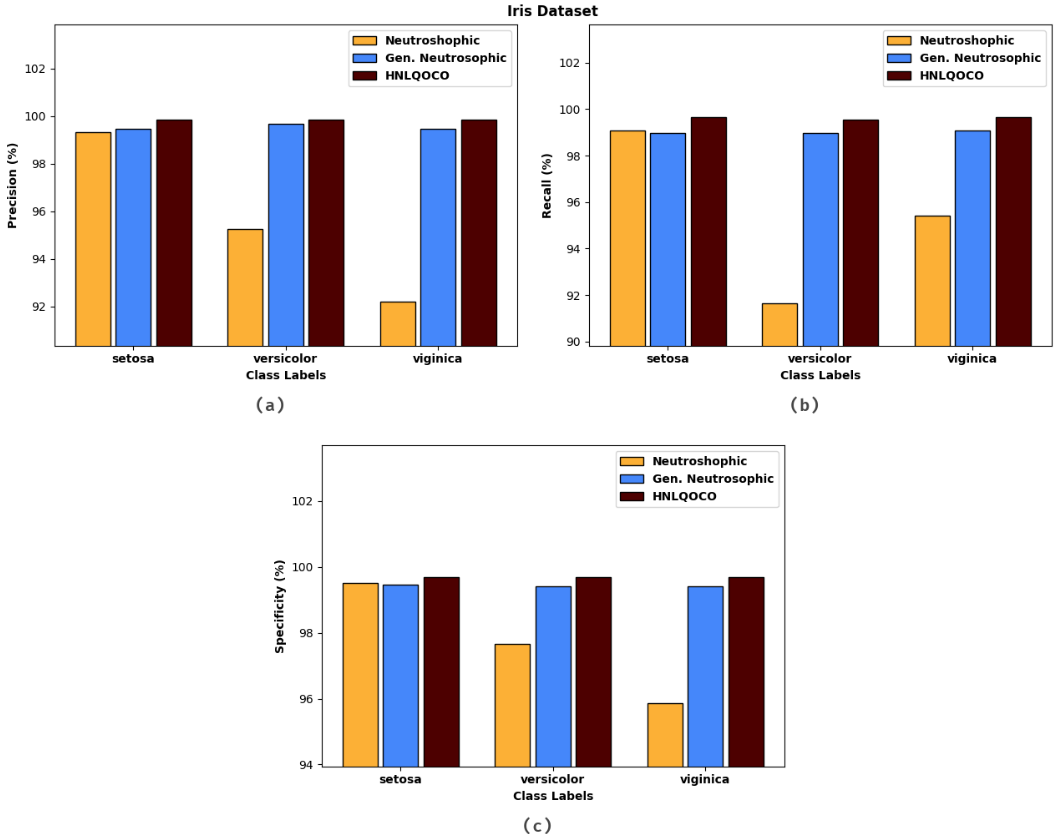

Figure 4: Comparative analysis of HNLQOCO model on iris dataset

Table 3 and Fig. 5 portrays a comparative study of the HNLQOCO model with other models [20] on wine dataset. The results showed that the HNLQOCO model has outperformed other methods on iris dataset. The NS model has resulted to lower precision values of 98.95%, 87.11%, and 95.12% on the classification of three classes respectively. Followed by, the GNS model has attained marginally enhanced precision values of 98.95%, 95.93%, and 95.47% on the classification of three classes respectively. However, the HNLQOCO model has reached maximum precision values of 99.65%, 97.44%, and 96.40% on the classification of three classes respectively.

Table 3: Comparison study of HNLQOCO model on wine dataset

Wine Dataset			
Class Labels	**Neutroshophic**	**Gen. Neutrosophic**	**HNLQOCO**
Precision			
Class-1	98.95	98.95	99.65
Class-2	87.11	95.93	97.44
Class-3	95.12	95.47	96.40
Recall			
Class-1	77.89	91.44	95.96
Class-2	89.75	91.44	96.52
Class-3	93.13	93.13	96.52
Specificity			
Class-1	99.24	99.49	99.74
Class-2	90.65	97.81	98.48
Class-3	97.81	97.81	98.73

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

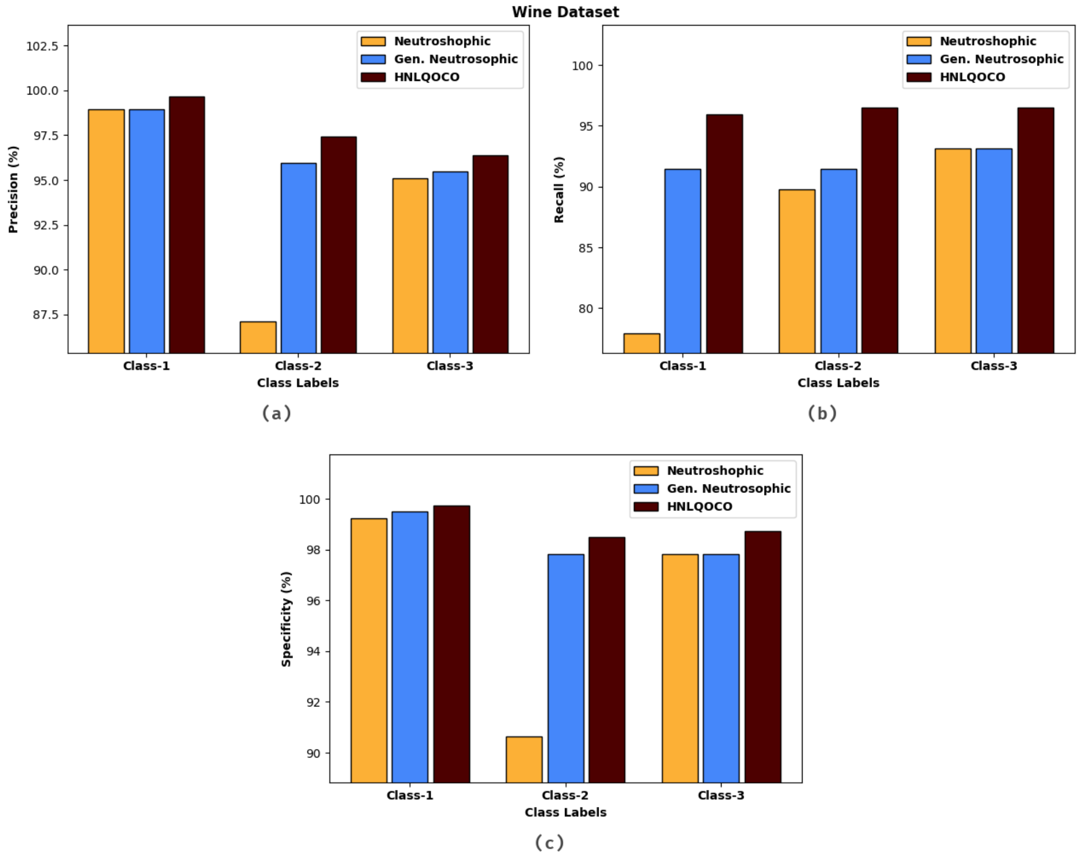

Figure 5: Comparative analysis of HNLQOCO model on wine dataset

Table 4 and Fig. 6 shows a comparative study of the HNLQOCO model with other models on WDBC dataset. The results designated that the HNLQOCO model has outperformed other methods on iris dataset. The NS model has resulted to lower precision values of 96.67% and 94.01% on the classification of M and B classes respectively. Followed by, the GNS model has attained slightly enhanced precision values of 94.53% and 99.22% on the classification of M and B classes respectively. However, the HNLQOCO model has reached maximum precision values of 97.89% and 97.24% on the classification of M and B classes respectively.

Table 4: Comparison study of HNLQOCO model on WDBC dataset

WDBC Dataset			
Class Labels	**Neutroshophic**	**Gen. Neutrosophic**	**HNLQOCO**
Precision			
Class M	96.67	97.53	97.89
Class B	94.01	96.70	97.24
Recall			
Class M	89.45	93.18	94.53
Class B	97.49	98.26	99.22
Specificity			
Class M	97.34	97.72	99.06
Class B	89.05	93.24	94.58

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022   Accepted: April 22, 2022

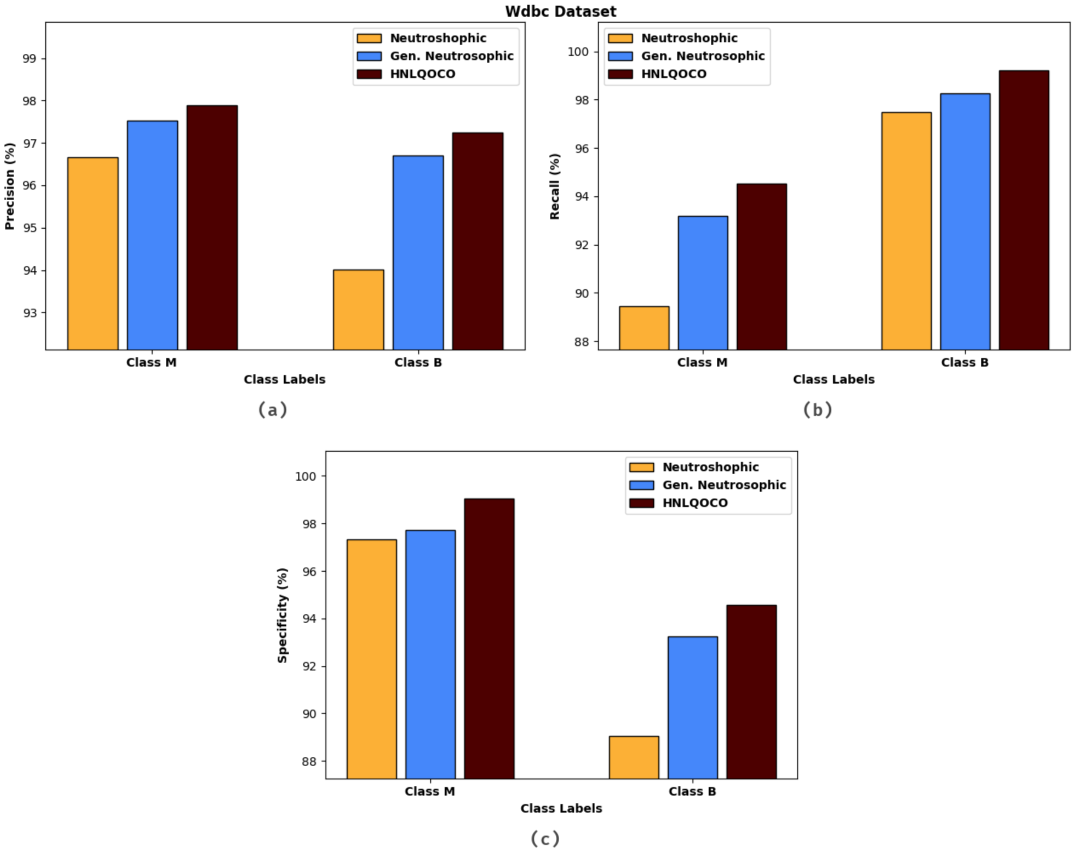

Figure 6: Comparative analysis of HNLQOCO model on WDBC dataset

Therefore, it can be concluded that the HNLQOCO model has accomplished maximum outcome on the test three datasets.

3. Conclusion

This study has provided a new HNLQOCO algorithm for accurate data classification. The presented HNLQOCO algorithm aims to integrate the concepts of NL and QOCO algorithm for improved data classification outcomes. Besides, the QOCO algorithm is designed by incorporating the concepts of quasi QOBL with traditional COA. Here, the NL is applied to represent various kinds of knowledge and the QOCO algorithm is applied to tune the produced NS rules. The experimental result analysis of the HNLQOCO model is tested using three benchmark medical dataset. The obtained results reported the significant performance of the HNLQOCO model over the other methods. In future, effective metaheuristic feature selection and reduction methodologies can be included into the presented HNLQOCO algorithm to boost the classification outcomes.

References

[1] Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N. and Asadpour, M., 2020. Boosting methods for multi-class imbalanced data classification: an experimental review. Journal of Big Data, 7(1), pp.1-47.

[2] Griffiths, D. and Boehm, J., 2019. A review on deep learning techniques for 3D sensed data classification. Remote Sensing, 11(12), p.1499.

[3] Basavegowda, H.S. and Dagnew, G., 2020. Deep learning approach for microarray cancer data classification. CAAI Trans. Intell. Technol., 5(1), pp.22-33.

[4] Liu, H., Zhou, M. and Liu, Q., 2019. An embedded feature selection method for imbalanced data classification. IEEE/CAA Journal of Automatica Sinica, 6(3), pp.703-715.

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022

[5] Wang, Y., Gan, W., Yang, J., Wu, W. and Yan, J., 2019. Dynamic curriculum learning for imbalanced data classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5017-5026).
[6] Sherwani, F., Ibrahim, B.S.K.K. and Asad, M.M., 2021. Hybridized classification algorithms for data classification applications: A review. Egyptian Informatics Journal, 22(2), pp.185-192.
[7] Xia, Y., Li, W., Zhuang, Q. and Zhang, Z., 2021. Quantum-enhanced data classification with a variational entangled sensor network. Physical Review X, 11(2), p.021047.
[8] Chaudhuri, A. and Sahu, T.P., 2021. A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification. Computers & Electrical Engineering, 90, p.106963.
[9] Jabbar, M.A., 2021. Breast cancer data classification using ensemble machine learning. Engineering and Applied Science Research, 48(1), pp.65-72.
[10] Mirzaei, B., Nikpour, B. and Nezamabadi-pour, H., 2021. CDBH: A clustering and density-based hybrid approach for imbalanced data classification. Expert Systems with Applications, 164, p.114035.
[11] Thanga Selvi, R. and Muthulakshmi, I., 2021. An optimal artificial neural network based big data application for heart disease diagnosis and classification model. Journal of Ambient Intelligence and Humanized Computing, 12(6), pp.6129-6139.
[12] Sun, Y. and Platoš, J., 2021. High-dimensional data classification model based on random projection and Bagging-support vector machine. Concurrency and Computation: Practice and Experience, 33(9), p.e6095.
[13] Ramkissoon, A.N., Mohammed, S. and Goodridge, W., 2021. Determining an Optimal Data Classification Model for Credibility-Based Fake News Detection. The Review of Socionetwork Strategies, 15(2), pp.347-380.
[14] Aydadenta, H. and Adiwijaya, A., 2018. A clustering approach for feature selection in microarray data classification using random forest. Journal of Information Processing Systems, 14(5), pp.1167-1175.
[15] Mikail Bal , Katy D. Ahmad , Arwa A. Hajjari , Rozina Ali, The Structure Of Imperfect Triplets In Several Refined Neutrosophic Rings, Journal of Neutrosophic and Fuzzy Systems, Vol. 2 , No. 1, (2022) : 21-30 (Doi : https://doi.org/10.54216/JNFS.020103)
[16] Jain, A., Nandi, B.P., Gupta, C. and KumarTayal, D., 2019, March. A hybrid framework based on PSO and neutrosophic set for document level sentiment analysis. In International Conference on Information Technology and Applied Mathematics (pp. 372-379). Springer, Cham.
[17] Khishe, M. and Mosavi, M.R., 2020. Chimp optimization algorithm. Expert systems with applications, 149, p.113338.
[18] Jia, H., Sun, K., Zhang, W. and Leng, X., 2021. An enhanced chimp optimization algorithm for continuous optimization domains. Complex & Intelligent Systems, pp.1-18.
[19] Dataset Source: https://archive.ics.uci.edu/ml/datasets.php
[20] Basha, S.H., Abdalla, A.S. and Hassanien, A.E., 2016, December. Gnrcs: hybrid classification system based on neutrosophic logic and genetic algorithm. In 2016 12th International Computer Engineering Conference (ICENCO) (pp. 53-58). IEEE.

Doi : https://doi.org/10.54216/IJNS.1803011

Received: February 03, 2022 Accepted: April 22, 2022