Modeling Influenced Criteria in Classifiers' Imbalanced Challenges Based on TrSS Bolstered by The Vague Nature of Neutrosophic Theory

Ibrahim El-Henawy [1] , Shrouk El-Amir [1,*], Mona Mohamed[2] Florentin Smarandache[3]

[1]Faculty of Computers and Informatics, Zagazig University; Sharkia, Egypt, Emails: ielhenawy@zu.edu.eg, shelshazly@fci.zu.edu.eg
[2]Higher Technological Institute, 10th of Ramadan City 44629, Egypt, Email: mona.fouad@hti.edu.eg
[3] University of New Mexico, 705 Gurley Ave., Gallup, NM 87301, USA,Email: smarand@unm.edu

* Correspondence: shelshazly@fci.zu.edu.eg; Tel: +201017461959

Abstract: Because of the advancements in technology, classification learning has become an essential activity in today's environment. Unfortunately, through the classification process, we noticed that the classifiers are unable to deal with the imbalanced data, which indicates there are many more instances (majority instances) in one class than in another. Identifying an appropriate classifier among the various candidates is a time-consuming and complex effort. Improper selection can hinder the classification model's ability to provide the right outcomes. Also, this operation requires preference among a set of alternatives by a set of criteria. Hence, multi-criteria decision-making (MCDM) methodology is the appropriate methodology can deploy in this problem. Accordingly, we applied MCDM and supported it through harnessing neurotrophic theory as motivators in uncertainty circumstances. Single value Neutrosophic sets (SVNSs) are applied as branch of Neutrosophic theory for evaluating and ranks classifiers and allows experts to select the best classifier So, to select the best classifier (alternative), we use MCDM method called Multi-Attributive Ideal-Real Comparative Analysis (MAIRAC) and the criteria weight calculation method called Stepwise Weight Assessment Ratio Analysis (SWARA) where these methods consider single-value neutrosophic sets (SVNSs) to improve and boost these techniques in uncertain scenarios. All these methods are applied after modeling criteria and its sub-criteria through a novel technique is Tree Soft Sets (TrSS). Ultimately, the findings of leveraging these techniques indicated that the hybrid multi-criteria meta-learner (HML)-based classifier is the best classifier compared to the other compared models.

Keywords: Neutrosophic theory; Multi-Criteria Decision Making; Class Imbalance; Meta-Learner; Ranking Classifiers; Single Values Neutrosophic Sets, Tree Soft Set (Trss).

1. Introduction

Currently, Artificial Intelligence (AI) techniques have been applied in several spheres. As in [1] where AI techniques are leveraged in healthcare, Fiscal fraud [2], and agriculture[3]. As well, machine learning (ML) techniques subset of AI are gleaning valuable knowledge from massive, complicated, diverse, and hierarchical data[4]. Also, ML techniques can be used as a classifier. Just like [5] described classification as ML techniques wherein a computer program learns from historical data and then applies that knowledge to forecast the class label for data that hasn't yet been observed. These techniques are represented in k Nearest Neighbors (KNN), Support Vector Machines (SVM), and random forest (RF) [5]. Moreover, Johnson et al.[6] affirmed that the majority or minority class is forecasted using binary classification. Besides that [7] categorized data in binary classification into balanced or imbalanced. And [8] demonstrated that the majority of classifiers travail optimally when the response variable's distribution in the dataset is balanced. In spite of that, the techniques mentioned in [5] encounter significant difficulties due to imbalanced data. Due to [9] one class has more instances than the other classes in imbalanced data , and the distribution of classes is skewed toward that class. From perspective of [10] positive instances are typically referred to as the minority class, whereas negative instances are typically called the majority class. Also, [11] indicated that in handling an imbalanced hurdle, the conventional algorithms exhibit a bias in favor of the dominant class.

Hence, [12] exhibited that class imbalance is still a perplexing problem that needs further study to be properly understood and skillfully managed. All that motivated[13] for developing algorithms or techniques that are very dependable and efficient is essential to properly addressing the problems brought on by the imbalanced datasets.

As per the prior literature [13, 14], several techniques have been suggested to tackle the problem of class disparity, and these may be roughly categorized into (1) algorithm-level methods [15]included cost-sensitive learning which employs the expenses of incorrect classifying samples and make an effort to improve the classifiers' favorability for the minority class by adding various cost variables into the algorithms. (2) sampling methods [15] which encompasses over sampling, random under sampling, synthetic minority oversampling technique (SMOTE), and edited nearest neighbour.(3) ensemble learning, this technique applied boosted based methods, pre-processing ensemble, and boosted imbalanced data solving toward endeavor to improve the unbalanced data classification's accuracy by fusing many classifiers to produce a novel, more potent classifier. Additionally, Chamlal et al. [13] proposed Hybrid Multi-criteria Meta-learner (HML) which includes an ensemble-based meta-learner component and a multi-objective optimization component as its two primary parts.

General speaking, selection of optimal and suitable classifier for treating with imbalanced data amongst these techniques is crucial process.

1.1 Motivation of Study

According to the surveys conducted for prior studies. From perspective of [13] it’s difficult to foresee the unpredictable, notably when tackling the problem of class imbalance, which occurs when the training data's class distribution is biased in favor of one particular class. In the same vein [16] demonstrated that the majority and minority samples are included in the imbalance datasets. Comparatively speaking, there are much less sample instances in the minority class than in the majority class. Accordingly, severe skews in the distribution of classes and inadequate rendition of specific data are persistent challenges in many domains as medicine [17], predicting defects for software[18],and in financial services [19].Hence, [20] stated that the performance of conventional classifiers may suffer when there is an imbalanced distribution of classes in a dataset.

Another aspect discussed by [21] in classification issues, depending on only a single criteria is widely used for evaluation. However, the evaluation of only one aspect may mis select the best performance classifier. To select the best performance classifier, several evaluation criteria, including proficiency, time consuming, uniformity, and others, need to be utilized. The multicriteria evaluation aims to achieve a balance between these criteria instead of depending on only one criterion [22] . So, we need an efficient multi-criteria decision-making method that assesses and ranks classifiers and allows experts to select the best classifier for their applications by using the previously mentioned criteria.

1.2 Contribution

Herein, we are evaluating the optimal classifier based on a set of criteria. For conducting this process, we are leveraging MCDM techniques which have ability to treat with such problems. Especially, Stepwise Weight Assessment Ratio Analysis (SWARA) for obtaining criteria's weights. Also, Multi-Attributive Ideal-Real Comparative Analysis (MAIRAC) is applied for ranking the alternatives of classifiers and select optimal classifier.

Neutrosophic theory is deploying in this study and contribute to MCDM techniques for bolstering and supporting expert in ambiguity situations as incomplete data and uncertainty [23]. Due to, ability of this theory to measure membership function as truth (T), also non-membership function false (F) whilst take into consideration indeterminacy (I). Thereby, single value neutrosophic sets (SVNSs) as type of Neutrosophic is implementing in evaluation process.

Also, Tree soft set (TrSS) is leveraging in this problem to model the identified criteria and its sub-criteria in set of nodes which resident into set of levels. TrSS is introduced Smarandache [24] who is founder of this approach as well as introduced Neutrosophic theory.

1.3 Study Outline

This study is organized into a set of sections; each section exhibits the benefits of our study and the followed steps toward achieving study's objectives.

Section one: illustrated the main idea of our study, motivations and the main contributions which are provided through our study. For completing our objectives' study, we conducted survey for prior techniques and studies in section 2. Through the conducted surveys, we determined the effective techniques to treat our problem through conducting SDMM. To validate the accuracy of this model, it forced us to apply the constructed model on real case study in section four. Finally, we recorded the results and conclusions which we reached in this study research through section five.

2. Literature review

In this section we exhibited the earlier studies which related to our study's objectives. Therefore, this section divides into set of sub-sections. Each sub-section introduces previous studies and techniques have been harnessed.

2.1 Around classification of imbalanced data

Several strategies [22, 25] have been mentioned by researchers to deal with the imbalanced data problem. These strategies can be categorized into data-level approaches, algorithm-level approaches, cost-sensitive strategies, and boosting strategies. The first strategy rebalances the data, utilizing the resampling technique to improve accuracy. In the second strategy, the standard classifiers are biased towards the minor class by adjusting their methodology. The third strategy gathers data-level and algorithm-level strategies by giving higher costs to positive samples and decreasing these costs. The fourth strategy combines multiple learners and then aggregates their predictions.

The Random Over-Sampling (ROS)[26] technique is the simplest over-sampling approach that randomly generates minor instances from the imbalanced data set until the class distribution more balanced. The Random Under-Sampling Strategy (RUS) is the simplest under-sampling approach that picks negative examples at random and discards them from the dataset until the class distribution is balanced. Wang et al. [27] used the SMOTE method to synthesize data with the Tomek Links technique to eliminate some of the majority of cases. Wang et al. [28] introduced Focal-XGBoost and Weighted-XGBoost, which blend the XGBoost algorithm with focal and weighted strategies to cope with imbalanced classification issues by minimizing the significance of well-classified cases. Ref [29] applied integrated optimization and sampling presumptions to address class imbalances. The researchers employed simulated annealing to choose the optimal subset of negative class records based on F-score. The under-sampled training set was then trained using several core classifiers, including SVM, KNN, DA, and DT. Boosted Random Forest [30] is constructed from two components: the boosting technique and the random forest classifier, in which each decision tree in the forest is created based on misclassification penalties. HICD [31] depends on data density; it is a hybrid, unbalanced classification model. It creates subsets for various instance classes, builds ensemble models, splits the data space using a density-based resampling technique, and chooses suitable models according to the instance distribution. Liu et al. [32] developed the fuzzy SVM algorithm and began dealing with borderline noise by employing a new strategy of measuring distance and gaussian fuzzy to decrease the influence of this noise. Zhang et al. [33]trained several classifiers on balanced subsets obtained by POS (perturbation-based oversampling) and used majority voting for ensemble learning. Barua et al. [34]presented the MWMOTE technique, which addresses imbalanced learning through determining key minority class instances, providing penalties based on proximity to the majority class, and creating synthetic instances from the minority class. Choudhary et al. [35] provided a method that employs a fuzzy clustering technique to segment the complex imbalance challenge into smaller issues before allocating ratings to each sub-classifier for a majority vote.

2.2 Influenced Neutrosophic Theory in Evaluation Process

In the recent studies, Neutrosophic with its various types are emerged in various vital domains toward supporting the stakeholders with valuable decisions in anxiety ambience. As Elhenawy et al. [36] employed neutrosophic especially, Triangular Neutrosophic Sets (TriNSs) for weighting criteria which contributed to evaluate alternatives of Metaverse in healthcare domain.Also, SVNSs are merged with TrSS in [37] for modeling criteria and its sub-criteria based blockchain technology (BCT).Portfolio selection model is established in [38] through adopting neutrosophic theory and entropy objective function and this model is applied on real time case study to validate the model accuracy. The optimal warehouse management software is selected through utilizing various MCDM with SVNSs for evaluating warehouse management software programs and select optimal toward achieving sustainable logistics systems [39]

As mentioned previously, we need efficient MCDM techniques to be implemented under authority of SVNSs and modeling the identified criteria and sub-criteria using TrSS which applied to model the criteria and sub-criteria in various applications as indicate the best location for solar hydrogen production [40]. Also, TrSS is leveraged in [41] for recommending the secure enterprise based on modeling blockchain criteria and its sub-criteria using TrSS.

Overall, These techniques are leveraged for constructing a robust decision-making model for assessing and ranking classifiers and allow experts to select the best classifier. So, we will use HML and MESA (boost ensemble imbalanced learning using a meta-sampler) with other familiar solutions based on sampling and cost-sensitive techniques for selecting the best classifier.

3. Soft Decision-Making Model (SDMM)

The objective of this section is to cover the following points:

- o How is the evaluation process conducting?
- o What techniques are used for serving the objective of model and study? And what is the role of each technique?
- o What are the influenced criteria which impact on the quality and performance of classifiers?

The previous questions will be answered through the following sub-sections.

3.1 Preliminaries

The utilized techniques and its basic concept are exhibited in this sub-section.

3.1.1 Tree Soft Sets [37]

A novel technique of TrSS is proposed by Smarandache who is founder for Neutrosophic theory Smarandache [24]. This technique has several concepts which described as:

- Assuming that $\aleph$ be a universe of discourse which includes ϱ a non-empty as subset of $\aleph$, thus the powerset of ϱ expressed as p (ϱ).
- Let TrSS encompasses set of levels, each one has a multitude of nodes as:
  - ➢ Level 1: consists of a multitude of nodes where each node represents main criteria, then expressed as: C={C1, C2,..Cn} for integer n ≥ 1.
  - ➢ Level 2: includes several sub-nodes of {C1, C2,..Cn} and stated as { C1-1, …C1-n } branched of C1 , and { C2-1, …C2-n } branched of C2 , finally { Cn-m, …Cn-m } branched of Cn .
  - We call the leaves of the graph-tree, all terminal nodes (nodes that have no descendants). Then, Tree Soft Set: F: P (Tree (C)) → p (ϱ).
  - Tree (C) is the set of all nodes and leaves (from level 1 to level n) of the graph-tree, and P (Tree(δ)) is the powerset of the Tree (Ind). All node sets of TrSS of level n as: Tree (C) = {C nm | nm= 1, 2, ...}.

3.1.2 Single-Valued Neutrosophic Sets (SVNSs)[42]

SVNSs are a branch of Neutrosophic theory that originated from Smarandache's work. Whilst SVNSs consider three measurement and probabilities as Truth (ϑ), Falsity (ν), and Indeterminacy (δ). Hence, three measurements are deployed and represented as:

- Assume that χ is universal set and κ is element in χ and this element is formed as: $\vartheta_\kappa(\omega)$, $\nu_\kappa(\omega)$, $\delta_\kappa(\omega)$.

- $0 \leq sup\, \vartheta_{\kappa}(\omega)$+sup $\nu_{\kappa}(\omega)$+ sup$\delta_{\kappa}(\omega) \leq 3$.
- The operations in SVNSs formed as:
  - Addition of two sets : $\widetilde{\mathrm{Ne}_1}+\widetilde{\mathrm{Ne}_2}=$

$$\langle(\tau_1 + \tau_2 - \tau_1 \tau_2, \gamma_1, \gamma_2, \varphi_1\varphi_2)\rangle$$

  - Multiplication of two sets : $\widetilde{\mathrm{Ne}_1}+\widetilde{\mathrm{Ne}_2}=$

$$\langle(\tau_1 \tau_2, \gamma_1 + \gamma_2 - \gamma_1 \gamma_2, \varphi_1 + \varphi_2 - \varphi_1 \varphi_2)\rangle$$

3.2 Development of SDMM: Evaluation classifiers and selecting the best classifier in the imbalanced data problem

Developing SDMM for evaluating determined alternatives of classifiers required follow set of steps to implement the mentioned techniques.

Step 1: Structuring criteria and sub-criteria into set of levels.

1.1 Determining set of alternatives of classifiers which involve into evaluation process.

1.2 Determining the influenced criteria and sub-criteria which contribute to evaluating process.

1.3 Modeling and structuring these criteria and its sub-criteria as nodes into several levels.

1.4 DMs panel is formed for rating enterprises based on modelled criteria and sub-criteria.

Step 2: SVNSs based SWARA for generating weighting [43]**.**

2.1 SWARA is deployed for obtaining criteria weights as:

- Expert panel is rating criteria through using SVN scale. Each decision maker (DM) rates the criteria in his/her decision matrix.
- Deneutrosophic rating of each DM according to Eq.(1).

$$\mathrm{De_j} = \frac{2+\vartheta-\nu-\delta}{3} \tag{1}$$

- Where: ϑ, ν , δ indicated to truth, false, and indeterminacy respectively.
- The constructed deneutrosophic matrices are aggregated into an aggregated matrix through employing Eq.(2).

$$\wp_{\mathbf{j}} = \frac{\sum_{\mathbf{j}}^{\mathrm{n}} \mathrm{De}_{\mathbf{j}}}{\mathbf{U}} \tag{2}$$

- Where: **U** refers to number of DMs
- Comparative importance of Average value (Sj) is obtaining according to Eq.(3) [44]. The values of Sj facilitate obtaining the values of coefficient (Kj) according to Eq.(4).

$$\mathbf{S_j} = \begin{cases} \mathbf{0} & \mathbf{j = 1} \\ \wp_{\mathbf{j-1}} - \wp_{\mathbf{j}} & \mathbf{j > 1} \end{cases} \tag{3}$$

$$\mathbf{K_j} = \begin{cases} \mathbf{1} & \mathbf{j = 1} \\ \mathbf{s_j} & \mathbf{j > 1} \end{cases} \tag{4}$$

- Generating recalculated weights (qj) through implementing Eq.(5).

$$\mathbf{q_j} = \begin{cases} \mathbf{1} & \mathbf{j = 1} \\ \dfrac{\mathbf{q_j - 1}}{\mathbf{k_j}} & \mathbf{j > 1} \end{cases} \tag{5}$$

- Accordingly, qj contributed to obtain final weights ($\boldsymbol{w_j}$)based on Eq.(6).

$$\boldsymbol{w_j} = \frac{\boldsymbol{q_j}}{\sum_{k=1}^{n} \boldsymbol{q_k}} \tag{6}$$

2.2 **SWARA is deployed for obtaining sub- criteria weights.** We follow the steps of 2.1 to generate sub- criteria weights.

Step 3: Recommending Optimal classifier based on MAIRAC and SVNSs.

3.1 Constructing Neutrosophic decision matrix for DM based on SVN scale.

3.2 Denutrosophic decision matrix for each DM according to Eq.(2). Also, Eq.(3) has vital role for aggregating these matrices into an aggregated matrix.

3.3 Eq.(6) responsible for calculating theoretical evaluation matrix(TP) toward estimating preferences of alternatives.

$$P_{A_j} = \frac{1}{m} \tag{6}$$

Where: m indicates the number of alternatives

3.4 Calculating real evaluation matrix (TR) according to Eq.s(7),(8).

$$\mathbf{tr_{ij}} = \mathbf{tp_{ij}} \left(\frac{\mathbf{x_{ij} - \bar{x}_i}}{\mathbf{x_i^+ - x_i^-}}\right), \text{ for maximum} \tag{7}$$

$$\mathbf{tr_{ij}} = \mathbf{tp_{ij}} \left(\frac{\mathbf{x_{ij} - x_i^+}}{\mathbf{x_i^- - x_i^+}}\right), \text{ for minimum} \tag{8}$$

3.5 Calculating criteria function (Q) based on Eq.(9).

$$Q_i = \sum_{i=1}^{m} g_{ij} \tag{9}$$

Where,

$$g_{ij} = t_{p_{ij}} - t_{r_{ij}} \tag{10}$$

4 Empirical Case Study.

We are implementing our SDMM for evaluating and ranking classifiers. Herein,we are leveraging five classifiers (alternatives) are Smote-Tomek Link (STL), Focal-XGBoost (FGB), Boosted Random Forest (BRF), MESA, and HML. The evalaution for five alternatives are conducting based on four criteria used for evaluation are described in Table 1.

The evalaution is conducting through implementing the steps are aforementioned.

4.1 Assigning each criterion to certain node and also, sub-criteria through leveraging TrSS technique as in Figure 1.

4.2 Three DMs are rating modeled criteria and sub-criteria through utilizing the SVN scale in Table 1.

4.2.1 For main criteria, DMs are rating the four criteria through SVN scale in Table 1.

- DMs'preferences and transform these preferences into deneutrosophic values according to Eq.(1). Eq.(2) used to aggregate these preferences as in Table 2.

- Sorting the criteria in descending order according toaggreget values of criteria.According to aggregated values in Table 2 , C4 is the more important than C2.Also, C2 is more important than C3.Accordingly, C3 is more important than C1.

- Employing Eq.(3) for genertaing Sj values .also, q_j and criteria weights are obtaining through Eq.s(5),(6).Table 3 involves the findings of applied Eq.s.Figure 2 showcases final weights for criteria.

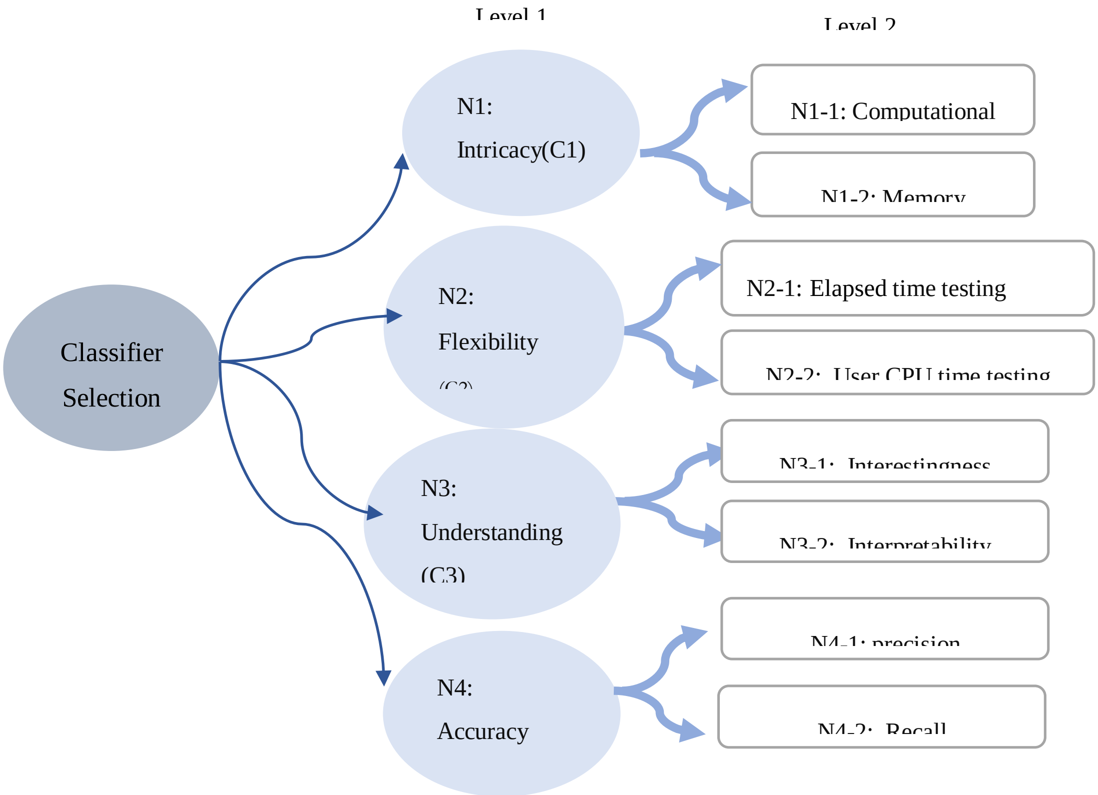

Figure 1.Tree soft for determined criteria and sub-criteria

4.2.2 For Sub- criteria, DMs are rating for each criterion is branched from main criterion according to structure of TrSS in Figure 1. SVN scale in Table 1 is utilized to rate the sub-criteria.

- The steps in 4.2.1 are followed for obtaining weights for these sub-criteria. Final findings formed in Figure 3,4,5,6 for sub-criteria of main criteria.

4.3 Three DMs are utilized SVN scale in Table 1 for second time for rating five alternatives of classifiers .

4.4 Aggregated decision matrix is generated through employing Eq.(2) after convreting three matrices form neutrosophic to deneutrosophic. Table 4 aggregets the three matrices into single matrix.

4.5 Calculating calculating theoretical evaluation matrix(TP) through Eq.(6) and represented in Table 5.

4.6 Calculating real evaluation matrix (TR)through employing Eq.(7) for maximum whlist Eq.(8) for minimum. The findings recorded in Table 6.

4.7 Calculating criteria function (Q) based on Eq.s (9),(10) and final ranking for alternatives is shown in Figure 7. This figure indicated that alternitive 5 (A5) HML is the optimal classifier. In contrast, alternative 2(A2) is the worst.

Table 1. Scale of SVN

Synonmy	Acronym	Scale		
		T	I	F
Extremly Weak	EW	0.00	1.00	1.00
Absolutely Weak	AW	0.10	0.90	0.90
Very Weak	VW	0.20	0.85	0.80
Weak	W	0.30	0.75	0.70
Fairly Weak	FW	0.40	0.65	0.60
Fairly	F	0.50	0.50	0.50
Fairly Well	FW	0.60	0.35	0.40
Well	W	0.70	0.25	0.30
Very Well	VW	0.80	0.15	0.20
Absolutely Well	AW	0.90	0.10	0.10
Extremly Well	EW	1.00	0.00	0.00

Table 2. Denutrosophic Matrix

Criteria	Expert Panel			Aggregeted values
	DM_1	DM_2	DM_3	
Intricacy(C1)	0.5	0.62	0.9	0.672222222
Flexibility (C2)	0.62	0.82	1	0.811111111

Understanding (C3)	0.82	0.9	0.5	0.738888889
Accuracy (C4)	0.9	1	0.82	0.905555556

Table 3. Final criteria weights

Criteria	**Sj**	**Kj**	**qj**	**wj**
Accuracy (C4)	0	1	1	0.280523
Flexibility (C2)	0.094444	1.094444	0.913706	0.256316
Understanding (C3)	0.072222	1.072222	0.852161	0.239051
Intricacy(C1)	0.066667	1.066667	0.798901	0.22411
		Sum	3.564767	1

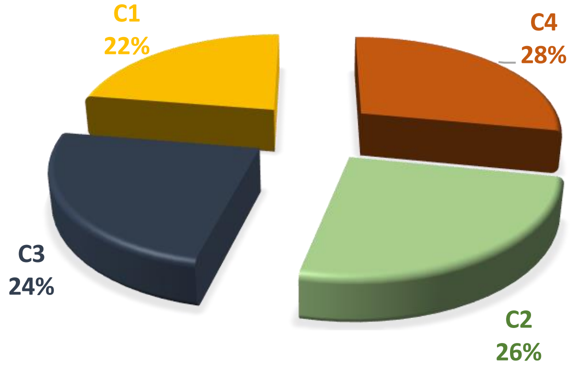

Figure 2. Final criteria weights

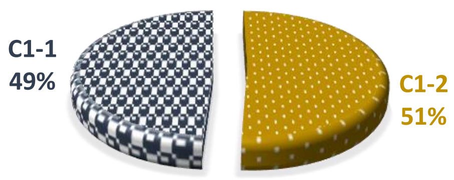

Figure 3.Sub-ceiteria of main criteria 1 weights

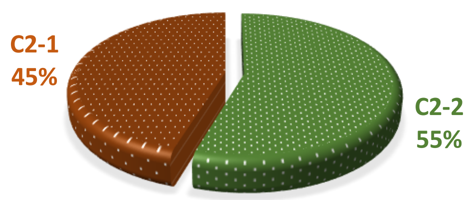

Figure 4. Sub-ceiteria of main criteria 2 weights

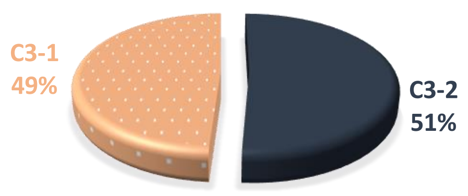

Figure 5.Sub-ceiteria of main criteria 3 weights

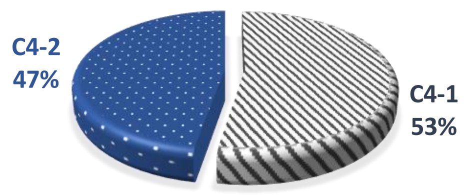

Figure 6.Sub-ceiteria of main criteria 4 weights

Table 4. Aggregeted matrix

Alternatives / criteria	Intricacy(C1)	Flexibility (C2)	Understanding (C3)	Accuracy (C4)
SMOTE Tomek Link	0.5	0.572222222	0.5	0.461111111
FGB	0.65	0.283333333	0.538888889	0.5
BRF	0.816666667	0.5	0.75	0.5
MESA	0.816666667	0.716666667	0.75	0.75
HML	0.9	0.75	0.844444444	0.844444444

Table 5. Theoretical Evaluation Matrix(TP)

Alternatives / criteria	Intricacy(C1)	Flexibility (C2)	Understanding (C3)	Accuracy (C4)
SMOTE Tomek Link	0.04482204	0.051263133	0.047810176	0.056104651
FGB	0.04482204	0.051263133	0.047810176	0.056104651
BRF	0.04482204	0.051263133	0.047810176	0.056104651
MESA	0.04482204	0.051263133	0.047810176	0.056104651
HML	0.04482204	0.051263133	0.047810176	0.056104651

Table 6. Real Evaluation Matrix (TR)

Alternatives / criteria	Intricacy(C1) MIN	Flexibility (C2) MAX	Understanding (C3) MAX	Accuracy (C4) MAX
SMOTE Tomek Link	0.04482204	0.03173432	0	0
FGB	0.028013775	0	0.005397923	0.005691776
BRF	0.009337925	0.02380074	0.034700934	0.005691776
MESA	0.009337925	0.047601481	0.034700934	0.005691776
HML	0	0.051263133	0.047810176	0.056104651

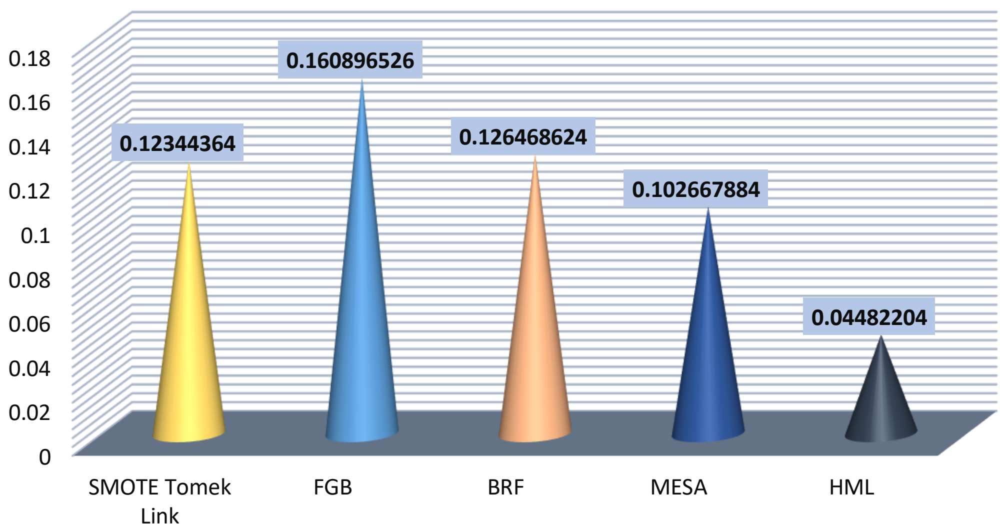

Figure 7. Final Ranking for Alternatives

5. Conclusions

Classification learning is a vital process. However, most classifiers can't deal with the imbalanced data problem, which indicates there are many more instances (majority instances) in one class than in another. Unfortunately, algorithms usually concentrate on only one criterion when facing this problem. Moreover, this became the catalyst for conducting this study and constructing SDMM based on MCDM techniques to compare different classifiers according to multiple dimensions in the case of an imbalanced data problem based on set of influenced criteria and sub-criteria. Hence, that motivates us for modeling and structuring these criteria and its sub-criteria to illustrate the relation between each other. Accordingly, SWARA as technique of MCDM, is applied to obtain criteria and sub-criteria weights, with the assistance of SVNSs. The findings of SVNSs based SWARA showcased in Figures 2 for main criteria whereas sub-criteria's weights illustrated in Figures 3,4,5,6. After that role of MAIRAC based on SVNSs initializes for rating five alternatives of classifiers through leveraging generated criteria's weights from SWARA -SVNSs. The findings recommended that A5 HML is the optimal classifier otherwise, A2 is the worst as in Figure 7.

References

[1] Alam, S., Raja, P., & Gulzar, Y. (2022). Investigation of machine learning methods for early prediction of neurodevelopmental disorders in children. *Wireless communications and mobile computing*, *2022*.

[2] Aburbeian, A. M., & Ashqar, H. I. (2023). Credit card fraud detection using enhanced random forest classifier for imbalanced data. *International conference on advances in computing research* (pp. 605–616). Springer.

[3] Hamid, Y., Wani, S., Soomro, A. B., Alwan, A. A., & Gulzar, Y. (2022). Smart seed classification system based on mobilenetv2 architecture. *2022 2nd international conference on computing and information technology (iccit)* (pp. 217–222). IEEE.

[4] Islam, A., Belhaouari, S. B., Rehman, A. U., & Bensmail, H. (2022). KNNOR: An oversampling technique for imbalanced datasets. *Applied soft computing*, *115*, 108288.

[5] Jadhav, A. S. (2020). A novel weighted TPR-TNR measure to assess performance of the classifiers. *Expert systems with applications*, *152*, 113391. DOI:10.1016/j.eswa.2020.113391

[6] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of big data*, *6*(1), 1–54.

[7] Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A. (2021). Classification of Imbalanced Data:Review of Methods and Applications. *IOP conference series: materials science and engineering*, *1099*(1), 012077. DOI:10.1088/1757-899x/1099/1/012077

[8] Li, J., Fong, S., Mohammed, S., & Fiaidhi, J. (2016). Improving the classification performance of biological imbalanced datasets by swarm optimization algorithms. *The journal of supercomputing*, *72*(10), 3708–3728.

[9] Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert systems with applications*, *73*, 220–239.

[10] Lutu, P. E. N. (2018). Improving classification performance for minority classes through the use of positive-versus-negative ensemble models. *2018 international conference on intelligent and innovative computing applications (iconic)* (pp. 1–6). IEEE.

[11] Kang, Q., Chen, X., Li, S., & Zhou, M. (2016). A noise-filtered under-sampling scheme for imbalanced classification. *IEEE transactions on cybernetics*, *47*(12), 4263–4274.

[12] Hasib, K. M., Iqbal, M. S., Shah, F. M., Mahmud, J. Al, Popel, M. H., Showrov, M. I. H., … Rahman, O. (2020). A survey of methods for managing the classification and solution of data imbalance problem. *ArXiv preprint arxiv:2012.11870*.

[13] Chamlal, H., Kamel, H., & Ouaderhman, T. (2024). A hybrid multi-criteria meta-learner based classifier for imbalanced data. *Knowledge-based systems*, *285*, 111367.

[14] Ayoub, S., Gulzar, Y., Rustamov, J., Jabbari, A., Reegu, F. A., & Turaev, S. (2023). Adversarial approaches

to tackle imbalanced data in machine learning. *Sustainability*, *15*(9), 7097.

[15] Castro, C. L., & Braga, A. P. (2013). Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data. *IEEE transactions on neural networks and learning systems*, *24*(6), 888–899.

[16] Ram, P. K., & Kuila, P. (2023). GAAE: a novel genetic algorithm based on autoencoder with ensemble classifiers for imbalanced healthcare data. *The journal of supercomputing*, *79*(1), 541–572.

[17] Chang, R.-I., Chiu, Y.-H., & Lin, J.-W. (2020). Two-stage classification of tuberculosis culture diagnosis using convolutional neural network with transfer learning. *The journal of supercomputing*, *76*(11), 8641–8656.

[18] Gao, Y., Zhu, Y., & Zhao, Y. (2022). Dealing with imbalanced data for interpretable defect prediction. *Information and software technology*, *151*, 107016.

[19] Abd El-Naby, A., Hemdan, E. E.-D., & El-Sayed, A. (2023). An efficient fraud detection framework with credit card imbalanced data in financial services. *Multimedia tools and applications*, *82*(3), 4139–4160.

[20] Al-Stouhi, S., & Reddy, C. K. (2016). Transfer learning for class imbalance problems with inadequate data. *Knowledge and information systems*, *48*, 201–228.

[21] Lavesson, N., Boeva, V., Tsiporkova, E., & Davidsson, P. (2014). A method for evaluation of learning components. *Automated software engineering*, *21*, 41–63.

[22] Zhao, J., Jin, J., Chen, S., Zhang, R., Yu, B., & Liu, Q. (2020). A weighted hybrid ensemble method for classifying imbalanced data. *Knowledge-based systems*, *203*, 106087.

[23] AL-baker, S. F., El-henawy, I., & Mohamed, M. (2024). Pairing New Approach of Tree Soft with MCDM Techniques: Toward Advisory an Outstanding Web Service Provider Based on QoS Levels. *Neutrosophic systems with applications*, *14*, 17–29.

[24] Smarandache, F. (2023). New Types of Soft Sets: HyperSoft Set, IndetermSoft Set, IndetermHyperSoft Set, and TreeSoft Set. *International journal of neutrosophic science*, *20*(4), 58–64. DOI:10.54216/IJNS.200404

[25] Guo, H., Diao, X., & Liu, H. (2019). Improving undersampling-based ensemble with rotation forest for imbalanced problem. *Turkish journal of electrical engineering and computer sciences*, *27*(2), 1371–1386.

[26] Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM sigkdd explorations newsletter*, *6*(1), 20–29.

[27] Hairani, H., Anggrawan, A., & Priyanto, D. (2023). Improvement performance of the random forest method on unbalanced diabetes data classification using Smote-Tomek Link. *JOIV: international journal on informatics visualization*, *7*(1), 258–264.

[28] Wang, C., Deng, C., & Wang, S. (2020). Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern recognition letters*, *136*, 190–197.

[29] Desuky, A. S., & Hussain, S. (2021). An improved hybrid approach for handling class imbalance problem. *Arabian journal for science and engineering*, *46*, 3853–3864.

[30] Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., & Jo, O. (2020). COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in public health*, *8*, 562169.

[31] Shi, S., Li, J., Zhu, D., Yang, F., & Xu, Y. (2023). A hybrid imbalanced classification model based on data density. *Information sciences*, *624*, 50–67.

[32] Liu, J. (2021). Fuzzy support vector machine for imbalanced data with borderline noise. *Fuzzy sets and systems*, *413*, 64–73.

[33] Zhang, J., Wang, T., Ng, W. W. Y., & Pedrycz, W. (2022). Ensembling perturbation-based oversamplers

for imbalanced datasets. *Neurocomputing*, *479*, 1–11.

[34] Barua, S., Islam, M. M., Yao, X., & Murase, K. (2012). MWMOTE--majority weighted minority oversampling technique for imbalanced data set learning. *IEEE transactions on knowledge and data engineering*, *26*(2), 405–425.

[35] Choudhary, R., & Shukla, S. (2021). A clustering based ensemble of weighted kernelized extreme learning machine for class imbalance learning. *Expert systems with applications*, *164*, 114041.

[36] Elhenawy, I., AL-baker, S. F., & Mohamed, M. (2023). Intelligent Healthcare: Evaluation Potential Implications of Metaverse in Healthcare Based on Mathematical Decision-Making Framework. *Neutrosophic systems with applications*, *12*, 9–21.

[37] Gharib, M., Smarandache, F., & Mohamed, M. (2024). *CSsEv: Modelling QoS Metrics in Tree Soft Toward Cloud Services Evaluator based on Uncertainty Environment*. Infinite Study.

[38] Islam, S., & Ray, P. (2018). Multi-Objective Portfolio Selection Model with Diversification by Neutrosophic Optimization. *Neutrosophic sets and systems: an international book series in information science and engineering*, *21*, 20.

[39] Kara, K., Yalçın, G. C., Simic, V., Önden, İ., Edinsel, S., & Bacanin, N. (2024). A single-valued neutrosophic-based methodology for selecting warehouse management software in sustainable logistics systems. *Engineering applications of artificial intelligence*, *129*, 107626.

[40] Mohamed, M. (2024). TrSo-BHARAT-V1 Recommender Model: Toward Eco-friendly Energy Based on Optimal Installation for Solar Hydrogen Production. *Plithogenic logic and computation*, *1*, 25–35. https://doi.org/10.61356/j.plc.2024.18860

[41] Smarandache, F. (2024). HyperSoft Set Methods in Engineering Evaluating Blockchain Cybersecurity Based on Tree Soft and Opinion Weight Criteria Method under Uncertainty Climate, *1*, 1–10. https://doi.org/10.61356/j.hsse.2024.17850.

[42] Smarandache, F. (2019). Neutrosophic set is a generalization of intuitionistic fuzzy set, inconsistent intuitionistic fuzzy set (picture fuzzy set, ternary fuzzy set), pythagorean fuzzy set, spherical fuzzy set, and q-rung orthopair fuzzy set, while neutrosophication is a genera. *Journal of new theory*, (29), 1–31.

[43] Debnath, B., Bari, A. B. M. M., Haq, M. M., de Jesus Pacheco, D. A., & Khan, M. A. (2023). An integrated stepwise weight assessment ratio analysis and weighted aggregated sum product assessment framework for sustainable supplier selection in the healthcare supply chains. *Supply chain analytics*, *1*, 100001.

[44] Cakmak, E. (2023). Supplier Selection for a Power Generator Sustainable Supplier Park: Interval-Valued Neutrosophic SWARA and EDAS Application. *Sustainability*, *15*(18), 13973.

Received: 11 Dec, 2023 Accepted: 15 Mar, 2024