Evidential Hybrid Re-sampling
for Multi-class Imbalanced Data

Fares Grina[1,2(✉)], Zied Elouedi[1], and Eric Lefevre[2]

[1] Institut Supérieur de Gestion de Tunis, LARODEC, Université de Tunis,
Tunis, Tunisia
grina.fares2@gmail.com, zied.elouedi@gmx.fr
[2] Univ. Artois, UR 3926, Laboratoire de Génie Informatique et d'Automatique de
l'Artois (LGI2A), 62400 Béthune, France
eric.lefevre@univ-artois.fr

Abstract. Learning from class-imbalanced datasets has gained substantial attention in the machine learning community, leading to solutions for healthcare, security, banking, etc. Specifically, binary imbalanced problems has received the most interest in the field. Yet, there has been little emphasis given to dealing with multi-class imbalance learning. Data imbalance can significantly worsen the classification performance, especially in the presence of other data difficulties such as uncertainty, i.e., ambiguous samples and noise. In this paper, we present an evidential hybrid re-sampling method for dealing with class imbalance in the multi-class setting. This technique uses the evidence theory to assign a soft label to each object. This evidential modeling provides more information about each object's region, which improves the selection of objects in both undersampling and oversampling. Our approach firstly selects ambiguous majority instances for undersampling, then oversamples minority objects through the generation of synthetic examples in borderline regions to better improve minority class borders. An adjustment has also been integrated in order to avoid excessive oversampling and undersampling. Benchmarking results have shown significant improvement of G-Mean of AUC metrics over other popular re-sampling methods.

Keywords: Resampling · Multi-class imbalance · Evidence theory · Data uncertainty

1 Introduction

Class imbalance is a very common situation in classification problems, specifically in many real-world scenarios such as intrusion detection [4], medical diagnosis [15], and fraud detection [16]. Formally, an imbalanced dataset contains at least one class with much fewer number of examples than the other classes. The underrepresented classes are called minority classes while the others with larger number of instances are referred to as majority classes. In most cases,

classification models tend to favor majority classes due to their large presence, while incorrectly classifying the instances from the minority classes. This raises a problem since minority classes often have more importance than the majority ones. For instance, in the medical domain, our interest is to detect patients with diseases, which could be a rare pattern (minority).

A lot of studies has been carried out in order to enhance classification performance on binary datasets, in which there is only one minority class and one majority class. Among many strategies, re-sampling is an effective approach for learning from class-imbalanced datasets. It aims at rebalancing the training set at the preprocessing level by adding synthetic minority objects (oversampling), removing majority samples (undersampling), or both (hybrid resampling). As a consequence, it improves the effects. The most traditional resampling techniques are random oversampling (ROS) and random undersampling (RUS). The former randomly selects minority objects and simply replicates them, while the latter randomly removes majority instances. Although it might seem that these random solutions are effective since the class distribution is balanced, they can lead to different issues. For instance, ROS can lead to overfitting, and RUS can potentially remove meaningful samples from the data [11]. Therefore, other methods have been suggested to make re-sampling less naive.

The most popular solution based on this logic is Synthetic Minority Oversampling Technique (SMOTE) [6], which is considered as the baseline for most contributions. Unlike ROS, SMOTE interpolates between minority objects that are close to each other, to create new synthetic minority instances. Unfortunately, SMOTE has some issues regarding overgeneralization and amplification of problems already present in the data, that is, data uncertainty involving class overlapping and noise [23]. In [12], authors suggested BorderlineSMOTE, a SMOTE variant with the goal of identifying borderline minority class objects to generate new samples. In fact, applying oversampling on the borders of the minority class allows the improvement of the visibility of minority objects. Safe-Level-SMOTE [5] is another technique which only selects objects in safe region for oversampling. Clustering has also been introduced in many oversampling methods [8] to smartly select the regions where to create synthetic points. More recently, some interests have been pointed towards generative adversarial networks as the basis for oversampling [24].

Regarding undersampling, contributions focused on intelligent selection of unwanted majority samples to discard, rather than randomly removing them. Traditionally, Editing Nearest Neighbors (ENN) [29] and Tomek Links (TL) [14] are occasionally used for undersampling. Similar to oversampling, clustering-based undersampling was presented in a number of works [27] to improve the selection of majority objects to remove.

The combination of oversampling and undersampling is also a useful strategy to re-balance the dataset. Traditionally, methods combine SMOTE's oversampling with undersampling filtering techniques. In [3], for example, the authors proposed two hybrid sampling approaches: SMOTE-ENN and SMOTE-TOMEKLINKS. Since SMOTE can potentially expand the minority set regions
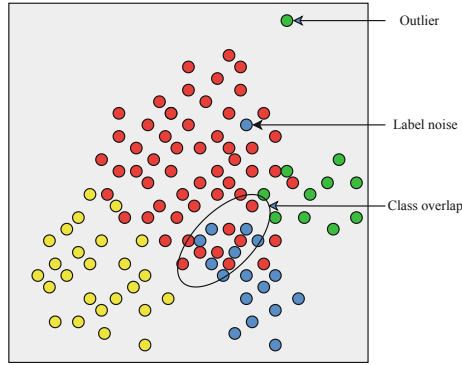
Fig. 1. Data difficulties in multi-class imbalanced datasets.

by interpolating new synthetic samples into the majority set clusters. This is called over-generalization. The application of ENN [29] to the SMOTE-oversampled training set tends to remove examples from both classes, meaning that every observation misclassified by its nearest neighbors is cleaned from the training data. Similar to SMOTE-ENN, instead of eliminating only the majority examples that form a Tomek Link, examples from both classes are deleted. SMOTE-RSB* [21] is another method that combines SMOTE for oversampling with the Rough Set Theory as a cleaning technique. Nevertheless, learning from multi-class imbalanced datasets has not been as heavily researched. This is due to the complexity of multi-class relationships compared to two-class problems. Rather than directly applying these methods to the multiple classes, many methods focus on decomposing multi-class problems into binary ones. For instance, the One-Versus-One (OVO) approach [13] is a decomposition scheme developed to modify the multi-class problem into multiple binary sub-problems, one for each pair of classes. Each sub-problem is trained on a binary classifier, ignoring the remaining instances not belonging the pair. Similarly to OVO, One-Versus-All (OVA) [22] is another decomposition framework which transforms the multi-class data into multiple binary sub-problems. However, OVA trains a classifier for each class in the training dataset. Other variants of these methods were also suggested [19], mostly to improve the combination of classifiers decisions. Even though binarization is simple and straightforward approach for learning from multiple class problems, it may lead to some regions being ignored and left unlearned. Specifically, when there is high data uncertainty, such as ambiguity created by high overlapping, and noise. These issues cannot be dealt with using decomposition techniques.

To handle the drawbacks of existing re-sampling algorithms, this paper presents an algorithm named Multi-Class Evidential Hybrid re-Sampling (MC-EVHS). It is an extension of the method proposed in [10] to specifically deal with imbalanced datasets with multiple classes and other data difficulties, i.e., overlapping classes, label noise and outliers, as illustrated in Fig. 1. This app-

roach uses evidence theory in order to represent memberships, before combining oversampling and undersampling. Utilizing this theory provides us with more information in order to better choose the locations of newly generated objects and which majority instances to remove. Consequently, we apply an evidential version of SMOTE on the minority classes, and evidential undersampling on majority classes. Since we are dealing with multiple classes, we also propose a mechanism to identify which classes to consider for oversampling or undersampling, in addition to an improved way of controlling the amount of re-sampling that should be performed for each class. This adds an adaptive behavior to our approach.

This paper will be divided as follows. First, evidence theory is recalled in Sect. 2. Section 3 details each step of our idea. Experimental evaluation is discussed in Sect. 4. We finish our paper with a conclusion and an outlook on future work in Sect. 5.

2   Evidence Theory

The theory of evidence [7,25,26], also referred to as Dempster-Shafer theory (DST) or belief function theory, is a flexible and well-founded framework for the representation and combination of uncertain knowledge. The frame of discernment defines a finite set of M exclusive possible events, e.g., possible labels for an object to classify, and is denoted as follows:

$$\Omega = \{w_1, w_2, ..., w_M\} \tag{1}$$

A basic belief assignment (bba) denotes the amount of belief stated by a source of evidence, committed to 2^Ω, i.e., all subsets of the frame including the whole frame itself. Precisely, a bba is represented by a mapping function $m : 2^\Omega \rightarrow [0, 1]$ such that:

$$\sum_{A \in 2^\Omega} m(A) = 1 \tag{2}$$

Each mass $m(A)$ quantifies the amount of belief allocated to an event A of Ω. A focal element is a subset $A \subseteq \Omega$ where $m(A) \neq 0$.

The *Plausibility* function is another representation of knowledge defined by *Shafer* [25] as follows:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall \ A \in 2^\Omega \tag{3}$$

$Pl(A)$ represents the total possible support for A and its subsets.

3   Evidential Hybrid Re-sampling for Multi-class Imbalanced Data

MC-EVHS is a re-sampling method which combines oversampling and undersampling to re-balance multi-class datasets. It firstly assigns soft evidential labels

to each object in the dataset. The theory of evidence is used here to represent memberships towards each class, and also each meta-class (overlapping between classes). This allows us to represent data uncertainty. The computed soft labels are then used to select unwanted majority samples for undersampling, and pick the right regions for oversampling the minoirty classes.

The definitions of majority and minority classes in multi-class imbalance have been discussed in many works. In this paper, we consider a minority each class that has a number of objects less than the mean s of the number of objects in each class.

For all classes with a number of objects higher than the mean s, the assigned memberships are used to smartly perform undersampling. Our version of undersampling has an adaptive behavior, since the number of removed objects depends on the amount of overlap and noise present in the corresponding majority class. However, each majority class size should not get inferior to the calculated mean s.

Regarding all classes with a number of objects lower than the mean s, we use the calculated evidential memberships in order to perform oversampling in the borders of the minority class. Similarly to undersampling, our version of oversampling adapts to each class and generates synthetic minority instances only in the wanted locations. The only stopping criterion is not exceeding the mean s.

3.1   Computation of Evidential Labels

Our proposed approach proceeds by determining the centers of each class and meta-class (the overlapping region), then creating a *bba* based on the distance between each object and each class center.

The class centers are calculated using the mean value of the training set in the corresponding class. Regarding the overlapping regions represented by meta-classes, the centers are defined by the barycenter of the involved class centers as follows:

$$C_U = \frac{1}{|U|} \sum_{\omega_i \in U} C_i \qquad (4)$$

where ω_i are the classes in U, U represents the meta-class, and C_i is the corresponding center.

After the creations of centers, we assign to each example a soft evidential label represented by a *bba* over the frame of discernment $\Omega = \{\omega_1, ..., \omega_M, \omega_0\}$, where the M classes are represented. The proposition ω_0 is included in the frame of discernment to represent the outlier, i.e., assignment of objects that are far from any class in the data. It is important to note that not all meta-classes should be considered as potential focal elements. Indeed, some classes do not overlap, and so no object needs to be assigned to the meta-class involving them. Additionally, it would be more computationally efficient to not calculate the mass value for this type of meta-classes. As enforced in [17], the meta-class center should be

closer to the centers of its involved classes than to other incompatible classes' centers.

Let x_s be an object belonging to the training set. The idea is that each class or meta-class center represents a piece of evidence to the evidential membership of x_s. Accordingly, the mass values for each focal element in regard to x_s's memberships should depend on $d(x_s, C)$, that is, the distance between the respective class center C and x_s. The farther the center is, the lower the mass value for the corresponding class. By analogy, the closer x_s is to a class/meta-class center, the more likely it belongs to it. Hence, the initial unnormalized masses should be represented by decreasing distance based functions. We use the Mahalanobis distance [18], in this work, as recommended by [17] in order to deal with anisotropic datasets. Meta-classes U are chosen based on the constraint given above. The unnormalized masses are calculated accordingly:

$$\hat{m}(\{\omega_i\}) = e^{-d(x_s, C_i)} \tag{5}$$

$$\hat{m}(U) = e^{-\gamma\, \lambda\, d(x_s, C_U)}, \quad for\ |U| \geq 1 \tag{6}$$

$$\hat{m}(\{\omega_0\}) = e^{-t} \tag{7}$$

where $\lambda = \beta\, 2^\alpha$. A value of $\alpha = 1$ is fixed as recommended to obtain good results on average, and β is a parameter such that $0 < \beta < 1$. It is used to tune the number of objects committed to the overlapping region. The value of γ is equal to the ratio between the maximum distance of x_s to the centers in U and the minimum distance. It is used to measure the degree of distinguishability among classes in U. The smaller γ indicates a poor distinguishability degree between the classes of U for x_s. The outlier class ω_0 is taken into account in order to deal with objects far from all classes, and its mass value is calculated according to an outlier threshold t.

Finally, the unnormalized belief masses are normalized as follows:

$$m(A) = \frac{\hat{m}(A)}{\sum_{B \subseteq \Omega} \hat{m}(B)} \tag{8}$$

3.2   Evidential Adaptive Undersampling

This part consists of downsampling the majority classes. As mentioned above, this is dedicated to the classes whose size is higher than the mean size, corresponding to a majority. The created *bbas* are used here to determine whether an object is necessary for the learning phase or not. The logic behind our idea is to discard the samples which have a high uncertainty, that is, samples which present a relatively higher difficulty to correctly classify. These types of instances involve high ambiguity (class overlapping samples), outliers, and label noise. The evidential membership is used to detect those samples.

Class Overlapping. In our framework, overlapping objects have high masses assigned to meta-class focal elements, i.e., non-singleton propositions. For example, a sample with the maximum mass assigned to $U = \{\omega_1, \omega_2, \omega_3\}$ signifies that it is located in the region intersecting the three classes ω_1, ω_2, and ω_3. This specific instance can be removed in the undersampling phase, in order to reduce the data ambiguity and reduce majority classes' sizes, at the same time.

Since class overlap has not been mathematically well characterized, some control over the number of examples removed should be set up. Consequently, the selected objects for undersampling are sorted in a descending order based on the average mass value attributed to non-singleton elements $\bar{\mu}$. Formally, for a selected object x_i:

$$\bar{\mu}_{x_i} = \frac{\sum_{|A|>1} m(A)}{k}, \quad A \in 2^{\Omega} \tag{9}$$

where k represents the number of non-singleton focal elements. In other words, the more ambiguous objects (higher imprecision) are firstly removed until the size of the corresponding majority class reaches the mean s.

Regarding majority objects whose highest mass is not assigned to a non-singleton proposition (meta-class), we can safely say that they are not located in an overlapping region. However, they could be situated far from all classes (outlier), or in a different class (label noise). To further detect those types of samples, the maximum plausibility $Pl_{max} = max_{\omega \in \Omega} Pl(\{\omega\})$ is used.

Outliers. This type of objects are located far from any class in the data. Typically, this could be described as the state of ignorance in our framework. Thus, objects with maximum plausibility assigned to ω_0, i.e., $Pl_{max} = Pl(\{\omega_0\})$, are eliminated from the dataset.

Label Noise. Reasonably, a safe object should have the maximum plausibility assigned to its label. Otherwise, it could be considered as located in another class, which could be described as label noise. Following this logic, each object, with the maximum plausibility affected to another label than its own, is discarded from the dataset.

3.3   Evidential Adaptive Oversampling

In order to strengthen the presence of minority classes in the dataset, an oversampling phase is added to make the borders of each minority class more robust. Our objective, in this phase, is to emphasis the borders of each minority class, much like other oversampling techniques such as BorderlineSMOTE [12]. Another aspect of our approach is avoiding oversampling noisy examples and outliers.

The previously calculated *bbas* are used in the phase to smartly pick the regions where synthetic minority objects should be created. Minority instances are sorted into three probable categories, similar to the cleaning step: overlapping, label noise, or outlier. If an object does not correspond to one of the three

categories, it is considered as located in a safe region and is not selected for the creation of new synthetic objects. The same is valid for label noise and outliers. Indeed, selecting noisy objects and outliers to generate new samples could lead to overgeneralization, which is a significant disadvantage of many oversampling techniques [12].

Our evidential approach to oversampling consists of generating synthetic minority data near the borderline objects of the minority class. The idea is to empower the minority class borders in order to avoid the misclassification of difficult objects. Formally speaking, only objects whose highest mass is committed towards an overlapping region are selected for oversampling. This procedure also helps us avoid selecting objects which are committed towards label noise and outlier. Indeed, selecting those objects would amplify the problems already present in the dataset.

As mentioned above, the number of generated examples is also controlled and the size of each minority class should not exceed the mean s. In fact, the objects in the corresponding minority class are sorted in descending order based on Eq. 9. The idea behind this is to give priority towards minority objects with higher uncertainty in order to generate synthetic object in difficult-to-classify locations.

4   Experimental Study

In this section, we will present firstly the setup of the conducted experiments in Subsect. 4.1. Lastly, we will show the results and discuss them in Subsect. 4.2.

4.1   Set-up

Datasets. In this work, seven multi-class imbalanced datasets were selected from the KEEL repository [2]. The details of each dataset are summarized in Table 1, where we describe the number of samples, number of features, number of classes, and the class distributions.

Table 1. Description of the imbalanced datasets selected from the KEEL repository.

Datasets	Class Distrib.	Features	Samples	#Class
Dermatology	112; 61; 72;49; 52; 20	34	366	6
Wine	71; 59; 48	13	178	3
Pageblocks	492; 33; 8; 12; 3	10	548	5
Thyroid	17; 37; 666	21	720	3
Hayes-roth	51; 51; 30	4	132	3
Contraceptive	629; 511; 333	9	1473	3
Yeast	463; 429; 244; 163; 51; 44; 35; 30; 20; 5	8	1484	10

Evaluation Metric. We evalute the performance of the compared techniques multi-class Geometric-Mean measure (G-Mean), which is the standard G-Mean calculated for each class separately by means of One-v-All strategy. The latter strategy was also used to compute the Area Under the ROC Curve (AUC), which is a popular metric to evaluate a model based on how good it separates the classes. For better assessment of the different performances, statistical analysis was run using the Wilcoxon's signed rank tests [28] for the significance level of $\alpha = 0.05$.

Baseline Classifier. As a baseline classifier, we use the decision tree classifier, more specifically CART. For all experiments, the implementation provided in the scikit-learn machine learning python library [20] was used, with the default parameters unchanged.

Compared Methods and Parameters. We aimed at comparing our method against re-sampling approaches specifically proposed for multi-class imbalanced datasets: Mahalanobis Distance Oversampling (MDO) [1], Static-SMOTE (S-SMOTE) [9], the basic version of SMOTE paired with the One-Versus-One strategy (SMOTE-MC), in addition to baseline (BL) with no re-sampling performed.

The considered parameters for our proposed method MC-EVHS are: α was set to 1 as recommended in [17], the tuning parameter t for $m(\{\omega_0\})$ was fixed to 2 to obtain good results in average, and we tested three different values for β in $\{0.3, 0.5, 0.7\}$ and selected the most performing one for each dataset, since the amount of class overlap differs in each case. For the other reference methods, we used the recommended parameters in their respective original papers.

4.2   Results and Discussion

In order to evaluate whether any of the methods perform consistently better than the other, we use a 10-fold stratified cross validation. The G-Mean and AUC results are presented respectively in Table 2 and Table 3. The two tables indicate that our approach MC-EVHS consistently produced the best results, in terms of G-Mean and AUC, when applying to these benchmarking datasets. Our proposal obtains the highest metric in 5 out of 7 datasets for both G-Mean and AUC. We can notice that all performances deteriorates with the increase of noise and overlap present in the dataset. However, our approach performed significantly better in datasets where there are many difficult-to-classify objects, in a consistent manner. It is safe to say that our approach is robust when applied on complex datasets.

The results for Wilcoxon's pairwise test are shown in Table 4. $R+$ represents the sum of ranks in favor of MC-EVHS, $R-$, the sum of ranks in favor of the reference methods, and exact p-values are calculated for each comparison. All pairwise comparisons can be considered as statistically significant with a level of 5% since all p-values are lower than the threshold 0.05. Thus, we can safely say that our method performed significantly better than MDO, Static-SMOTE, and SMOTE-MC.

Table 2. G-Mean results for KEEL datasets using CART.

Datasets	BL	SMOTE-MC	S-SMOTE	MDO	MC-EVHS
Dermatology	0.921	0.937	0.904	0.931	**0.952**
Wine	0.887	**0.936**	0.919	0.887	0.920
Thyroid	0.933	0.919	0.933	0.933	**0.981**
Hayes-roth	0.804	0.788	0.779	0.804	**0.820**
Contraceptive	0.442	0.441	0.448	0.442	**0.460**
Pageblocks	0.492	0.460	**0.553**	0.492	0.542
Yeast	0.599	0.635	0.642	0.599	**0.699**

Table 3. AUC results for KEEL datasets using CART.

Datasets	BL	SMOTE-MC	S-SMOTE	MDO	MC-EVHS
Dermatology	0.917	0.950	0.933	0.947	**0.955**
Wine	0.888	0.938	0.894	0.888	**0.950**
Thyroid	0.983	0.983	**0.989**	0.983	0.985
Hayes-roth	0.796	0.801	0.757	0.796	**0.811**
Contraceptive	0.470	0.464	0.476	0.470	**0.477**
Pageblocks	0.852	0.943	0.938	0.952	**0.954**
Yeast	0.664	0.674	**0.715**	0.664	**0.715**

Table 4. Pairwise comparisons of obtained G-Mean and AUC scores based on Wilcoxon's signed ranks test.

Comparisons	G-Mean			AUC		
	$R+$	$R-$	p-value	$R+$	$R-$	p-value
MC-EVHS vs BL	28.0	0.0	0.0078125	28.0	0.0	0.0078125
MC-EVHS vs SMOTE-MC	26.0	2.0	0.0234375	28	0.0	0.0078125
MC-EVHS vs S-SMOTE	26.0	2.0	0.0234375	19.0	9.0	0.0373677
MC-EVHS vs MDO	28.0	0.0	0.0078125	28.0	0.0	0.0078125

5 Conclusion

The aim of this paper was to specifically develop an approach for handling multi-class imbalanced datasets. Our method MC-EVHS can exploit the computed evidential memberships to better choose the locations for oversampling minority classes, and improve the selection of objects to eliminate in the undersampling phase. Evidence theory is an ideal tool in this case to express the different data difficulties that could be present in the dataset, i.e., class overlap, label noise and outliers. The conducted experiments confirmed the effectiveness of our proposed method and on the basis of a thorough statistical analysis, we may confirm

that MC-EVHS performed significantly better than some popular methods. Further, the combination of undersampling and oversampling paired with evidential memberships, can reduce the multi-class imbalance problem, without excessive use of re-sampling.

As future work, we propose to investigate the wider applicability of MC-EVHS with very large datasets and extreme levels of imbalance.

References

1. Abdi, L., Hashemi, S.: To combat multi-class imbalanced problems by means of over-sampling techniques. IEEE Trans. Knowl. Data Eng. **28**(1), 238–251 (2015)
2. Alcala-Fdez, J., et al.: Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. J. Multiple-Valued Logic Soft Comput. **17**, 255–287 (2010)
3. Batista, G., Prati, R., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor. **6**, 20–29 (2004)
4. Bedi, P., Gupta, N., Jindal, V.: I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems. Appl. Intell. **51**(2), 1133–1151 (2020). https://doi.org/10.1007/s10489-020-01886-y
5. Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C.: Safe-level-SMOTE: safe-level-synthetic minority over-sampling TEchnique for handling the class imbalanced problem. In: Theeramunkong, T., Kijsirikul, B., Cercone, N., Ho, T.-B. (eds.) PAKDD 2009. LNCS (LNAI), vol. 5476, pp. 475–482. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01307-2_43
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
7. Dempster, A.P.: A generalization of Bayesian inference. J. Roy. Stat. Soc. Ser. B (Methodol.) **30**(2), 205–232 (1968)
8. Douzas, G., Bacao, F., Last, F.: Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Inf. Sci. **465**, 1–20 (2018)
9. Fernández-Navarro, F., Hervás-Martínez, C., Gutiérrez, P.A.: A dynamic over-sampling procedure based on sensitivity for multi-class problems. Pattern Recogn. **44**(8), 1821–1833 (2011)
10. Grina, F., Elouedi, Z., Lefèvre, E.: Uncertainty-aware resampling method for imbalanced classification using evidence theory. In: Vejnarová, J., Wilson, N. (eds.) ECSQARU 2021. LNCS (LNAI), vol. 12897, pp. 342–353. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86772-0_25
11. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: review of methods and applications. Expert Syst. Appl. **73**, 220–239 (2017)
12. Han, H., Wang, W.-Y., Mao, B.-H.: Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 878–887. Springer, Heidelberg (2005). https://doi.org/10.1007/11538059_91
13. Hastie, T., Tibshirani, R.: Classification by pairwise coupling. Adv. Neural. Inf. Process. Syst. **10**, 507–513 (1997)
14. Ivan, T.: Two modifications of CNN. IEEE Trans. Syst. Man Commun. **SMC 6**, 769–772 (1976)

15. Khushi, M., et al.: A comparative performance analysis of data resampling methods on imbalance medical data. IEEE Access **9**, 109960–109975 (2021)
16. Li, Z., Huang, M., Liu, G., Jiang, C.: A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. Expert Syst. Appl. **175**, 114750 (2021)
17. Liu, Z.g., Pan, Q., Dezert, J., Mercier, G.: Credal classification rule for uncertain data based on belief functions. Pattern Recogn. **47**(7), 2532–2541 (2014)
18. Mahalanobis, P.C.: On the generalized distance in statistics, vol. 2, pp. 49–55. National Institute of Science of India (1936)
19. Murphey, Y.L., Wang, H., Ou, G., Feldkamp, L.A.: OAHO: an effective algorithm for multi-class learning from imbalanced data. In: 2007 International Joint Conference on Neural Networks, pp. 406–411. IEEE (2007)
20. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
21. Ramentol, E., Caballero, Y., Bello, R., Herrera, F.: SMOTE-RSB *: a hybrid pre-processing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. Knowl. Inf. Syst. **33**(2), 245–265 (2012)
22. Rifkin, R., Klautau, A.: In defense of one-vs-all classification. J. Mach. Learn. Res. **5**, 101–141 (2004)
23. Sáez, J.A., Luengo, J., Stefanowski, J., Herrera, F.: SMOTE-IPF: addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. Inf. Sci. **291**(C), 184–203 (2015)
24. Salazar, A., Vergara, L., Safont, G.: Generative adversarial networks and Markov random fields for oversampling very small training sets. Expert Syst. Appl. **163**, 113819 (2021)
25. Shafer, G.: A Mathematical Theory of Evidence, vol. 42. Princeton University Press, Princeton (1976)
26. Smets, P.: The transferable belief model for quantified belief representation. In: Smets, P. (ed.) Quantified Representation of Uncertainty and Imprecision. HDRUMS, vol. 1, pp. 267–301. Springer, Dordrecht (1998). https://doi.org/10.1007/978-94-017-1735-9_9
27. Tsai, C.F., Lin, W.C., Hu, Y.H., Yao, G.T.: Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. Inf. Sci. **477**, 47–54 (2019)
28. Wilcoxon, F.: Individual comparisons by ranking methods. In: Kotz, S., Johnson, N.L. (eds.) Breakthroughs in statistics. PSS, pp. 196–202. Springer, New York (1992). https://doi.org/10.1007/978-1-4612-4380-9_16
29. Wilson, D.L.: Asymptotic properties of nearest neighbor rules using edited data. IEEE Trans. Syst. Man Cybern. **3**, 408–421 (1972)