Uncertainty-aware resampling method for imbalanced classification using Evidence Theory

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

[1] Institut Supérieur de Gestion de Tunis, LARODEC, Université de 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), F-62400 Béthune, France
eric.lefevre@univ-artois.fr

Abstract. Class imbalance is a common issue in many real world classification problems. It refers to situations where the number of observations in the training dataset significantly differs for each class. Ignoring this issue will make it more challenging for classifiers to properly learn data characteristics, which results in poor performance. Many strategies have been proposed to deal with this issue. The most common one is tackling the imbalance at the preprocessing level, by re-sampling the training set. However, imbalanced classification can be affected by other data factors, such as uncertainty, i.e., ambiguous samples and noise. In this paper, we propose an uncertainty-aware hybrid re-sampling technique based on the theory of evidence to tackle imbalanced binary datasets in the presence of aleatoric uncertainty. A soft evidential structure is assigned to each object in the training set, which is later used to clean the dataset out of overlapping and noisy majority samples, and then selectively generate synthetic minority objects using a modified SMOTE algorithm. Experimental results on benchmark imbalanced datasets have shown significant improvement over popular re-sampling techniques.

Keywords: Resampling · Imbalanced datasets · Evidence theory · Data uncertainty

1 Introduction

Imbalanced classification is an active research topic in machine learning and data mining. It is a scenario in which class sizes are not equal making the class distribution imbalanced. Data imbalance exist in many real-life domains such as fraudulent credit card detection [25], medical diagnosis [5], drug discovery [18], etc. For instance, in imbalanced binary datasets, the class with the highest number of instances is referred to as the *majority* class, while the *minority* class is defined as the one with the fewest examples. In most cases, the minority class is more relevant than the majority one [7]. As an example, failing to detect intrusions in a company's network may result in huge financial losses.

A variety of variables may cause the class imbalance, such as the domain's nature (e.g. rare disease) or data collection factors (e.g. storage). Additionally, most classifier algorithms (such as decision trees, k-nearest neighbors, neural networks, etc) were designed with the presumption that training datasets have an even distribution, which reduces greatly their efficiency [13].

Many methods have been proposed over the years to cope with imbalanced datasets. Data resampling is one of the most efficient strategies for dealing with class imbalance [10]. This approach aims at fixing the uneven class distribution at the preprocessing level by re-balancing the training dataset. Being algorithm-independent, resampling is versatile and could be applied with any selected classifier.

In addition, recent findings show that class imbalance is not an issue in and of itself, but rather gets amplified by other data difficulties. Data uncertainty (can also be referred to as aleatoric uncertainty) refers to the imperfections present in the data. This type of uncertainty can include class overlapping and noise, which were proven to worsen the class imbalance issue [13].

To improve performance on imbalanced and uncertain binary datasets, we suggest an Uncertainty-Aware Hybrid reSampling (UAHS) method based on Evidence Theory, which was recently used for imbalanced classification [11, 12]. After creating soft evidential labels for each object, our method efficiently selects the majority instances to remove in an undersampling phase first, and the minority objects to focus on in the oversampling procedure lastly. The considered evidential label is appropriate for our goal, since it includes membership values towards single classes, in addition to a belief mass assigned to meta-classes (ambiguous region). This versatility allows us to create precise rules for the process of selecting undesirable samples in the undersampling phase, and intelligently select minority instances to generate new synthetic objects. It is important to note that our proposal is a hybrid resampling method, meaning that it performs both undersampling and oversampling, unlike [11] and [12] which are respectively pure oversampling and undersampling approaches.

The remainder of this paper will be divided as follows. First, related work for resampling methods is presented in Section 2. Evidence Theory is recalled in Section 3. Section 4 details each step of our idea. Experimental evaluation is discussed in Section 5. Our paper ends with a conclusion and an outlook on future work in Section 6.

2 Related work

Data resampling is one of the most common approaches for dealing with imbalanced classification [13]. In fact, data resampling deals with class imbalance at the preprocessing level by changing the class distribution of the training set. As a result, it alleviates the effects of distribution skewness of the learning process. These methods can be further categorized into three groups, namely:

- **Oversampling**: These techniques introduce new minority synthetic samples to re-balance the dataset. The most straightforward method is random

oversampling (ROS), which consists of selecting minority observations in the original data set and simply replicating them. Although it appears to be technically effective since the class balance is adjusted, it can lead to overfitting [16]. To cope with overfitting, the Synthetic Minority Oversampling Technique (SMOTE) was suggested in [6]. Unlike ROS, SMOTE generates new synthetic samples by interpolating among several minority objects that are close to each other. However, many studies [9, 30] have shown SMOTE's drawbacks which involve potential amplification of noise, overlap already present in the data. SMOTE's improvement include Borderline-SMOTE [14], which identifies borderline minority class examples to generate new samples. Clustering-based oversampling techniques were also proposed [9, 23] to smartly select the regions where to generate new points.
- **Undersampling**: These approaches create a subset of the original dataset by removing some majority class instances. Like random oversampling, the naive undersampling technique is to randomly remove majority objects, which may potentially remove meaningful information from the dataset. Therefore, other techniques have been suggested to smartly remove unwanted majority class instances. Commonly, traditional filtering techniques have been used to perform undersampling. For example, Neighborhood Cleaning Rule (NCL) discards majority class instances using the Edited Nearest Neighbors (ENN) introduced in [37]. Similarly, Tomek Links (TL) [15] is occasionally used as an undersampling method. Clustering has also been used for undersampling in a number of occasions [26, 34], to optimize the selection process of majority instances to eliminate.
- **Hybrid**: This strategy combines both oversampling and undersampling in order to re-balance the dataset. Typically, SMOTE is paired with an undersampling procedure to fix its drawbacks. For instance, SMOTE-ENN and SMOTE-TL were suggested in [3] to combine SMOTE with ENN and TL respectively. SMOTE-RSB* [29] is a method which combines SMOTE for oversampling with the Rough Set Theory [27] as a cleaning technique. In SMOTE-IPF [17], SMOTE is firstly executed, and then the Iterative-Partitioning Filter (IPF) [17] is performed to remove noisy original examples, and those introduced by SMOTE. Authors in [20] suggested a combination of a SMOTE-like algorithm with a cleaning procedure to reduce the effects of overlapping. Similarly, the class overlap issue is touched upon in [35] combining a soft clustering method with Borderline-SMOTE.

3 Theory of evidence

The theory of evidence [8, 31, 33], 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:

$$\Omega = \{\omega_1, \omega_2, ..., \omega_M\} \tag{1}$$

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:

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

$$m(\emptyset) = 0 \tag{3}$$

Each mass $m(A)$ quantifies the amount of belief allocated to an event A of Ω. A *bba* is unnormalized when $m(\emptyset) > 0$, and must be normalized under a closed-world assumption [32]. 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* [31] as follows:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall \ A \in 2^{\Omega} \tag{4}$$

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

4 Uncertainty-Aware Hybrid re-Sampling method (UAHS)

To tackle binary imbalanced datasets, we propose an Uncertainty-Aware Hybrid re-sampling method (UAHS). Observations are firstly assigned soft evidential labels (*bbas*) using the credal classification rule (CCR) introduced in [22].

CCR uses the centers of each class and meta-class as pieces of evidence for each example's membership, instead of using nearest neighbors as it has been employed in [11]. Unlike [11], our case deals with the soft labeling of both majority and minority classes. Thus, an evidential nearest neighbor-based approach might produce biased memberships towards the majority class, since the latter usually have a much higher density than the minority one.

The computed *bba* is later used for cleaning unwanted majority objects and selectively generating synthetic minority instances.

Each step is detailed in the following subsections.

4.1 Creating soft labels

UAHS proceeds by determining the centers of each class and meta-class (the overlapping region), then creating a *bba* based on the distance between the majority sample and each class center.

The class centers are simply computed by the mean value of the training set in the corresponding class. For the meta-class U, representing the overlapping region, the center is defined by the barycenter of the involved class centers as follows:

$$C_U = \frac{1}{|U|} \sum_{\omega_i \in U} C_i \tag{5}$$

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

Once the centers are created, the evidential soft label of each example is represented by a *bba* over the frame of discernment $\Omega = \{\omega_0, \omega_1, \omega_2\}$ where ω_1 and ω_2 represent respectively the majority and the minority class. The proposition ω_0 is included in the frame of discernment explicitly to represent the outlier class.

Let x_s be a sample belonging to the training set. Each class center represents a piece of evidence to the evidential membership of the sample. The mass values in regard to the class memberships of x_s should depend on $d(x_s, C)$, i.e., the distance between x_s and the respective class center. The greater the distance, the lower the mass value. Consequently, the closer x_s is to a specific class center, the more likely it belongs to the corresponding class. Hence, the initial unnormalized masses should be represented by decreasing distance based functions. We use the Mahalanobis distance [24], in this work, as recommended by [22] in order to deal with anisotropic datasets.

The unnormalized masses are calculated accordingly:

$$\hat{m}(\{\omega_i\}) = e^{-d(x_s, C_i)}, \qquad i \in \{1, 2\} \tag{6}$$

$$\hat{m}(U) = e^{-\gamma\ \lambda\ d(x_s, C_U)}, \qquad U = \{\omega_1, \omega_2\} \tag{7}$$

$$\hat{m}(\{\omega_0\}) = e^{t} \tag{8}$$

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 the majority and minority classes. 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 both classes, and its mass value is calculated according to an outlier threshold t.

As performed in [22], the unnormalized belief masses are finally normalized as follows:

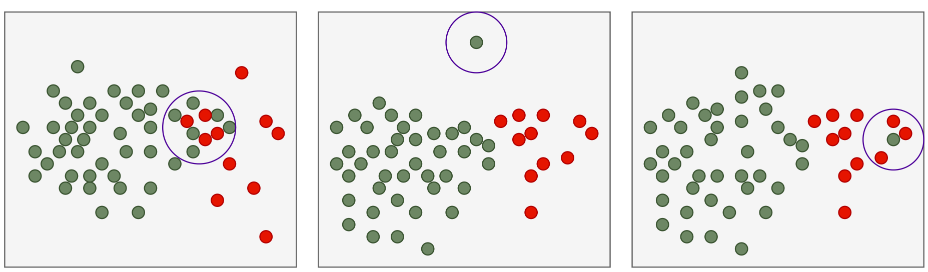
$$m(A) = \frac{\hat{m}(A)}{\sum_{B \subseteq \Omega} \hat{m}(B)} \tag{9}$$

As a result, a *bba* is created to formally represent each object's soft label.

4.2 Cleaning the majority class

As a result of bba creation, each majority object will have masses in 4 focal elements namely: $m(\{\omega_1\})$ for the majority class, $m(\{\omega_2\})$ for the minority class, $m(U)$ for the overlapping region U, and $m(\{\omega_0\})$ for the outlier class. These masses are used to remove problematic samples from the majority class. There are different types of unwanted samples which could be removed namely:

- **Overlapping:** Ambiguous samples are usually located in regions where there is strong overlap between classes as seen in Figure 1a. This situation could correspond to what is called "conflict" in Evidence Theory. In our framework, this type of examples will have a high mass value in $m(U)$. Thus, majority instances whose *bba* has the maximum mass committed to $m(U)$ are considered as part of an overlapping region, and are automatically discarded. To avoid excessive elimination and allow tuning, it is also possible to tune the parameter β. The higher value of β will result in fewer objects committed to the overlapping region. As for majority instances whose highest mass is not committed to $m(U)$ (i.e. not in overlapping regions), the observation is necessarily committed to one of the singletons in Ω ($\{\omega_1\}$, $\{\omega_2\}$, or $\{\omega_0\}$). In this situation, we make use of the *plausibility* function defined in eq. (4) to make a decision of acceptance or rejection. Each majority instance x_s is affected to the class with the maximum plausibility $Pl_{max} = max_{\omega \in \Omega} Pl(\{\omega\})$.
- **Label noise:** Normally, majority observations should have the maximum plausibility committed to $m(\{\omega_1\})$ which measures the membership value towards the majority class. By contrast, having Pl_{max} committed to $m(\{\omega_2\})$ signify that they are located in a minority location, as illustrated in Figure 1c. Consequently, these objects are eliminated from the dataset.
- **Outlier:** The final scenario occurs when the sample in question is located in a region far from both classes, as shown in Figure 1b. In our framework, this is characterized by the state of ignorance and could be discarded in the undersampling procedure. Hence, majority objects whose maximum plausibility Pl_{max} committed to $m(\{\omega_0\})$ are considered as outliers and removed from the dataset.



(a) Ambiguous samples in an overlapping area. (b) An outlier far from both classes. (c) A sample that could be characterized as label noise.

Fig. 1: Illustrations describing the different data difficulty factors that could worsen class imbalance. Green and red colors respectively represent the majority class and the minority one.

4.3 Applying selective minority oversampling

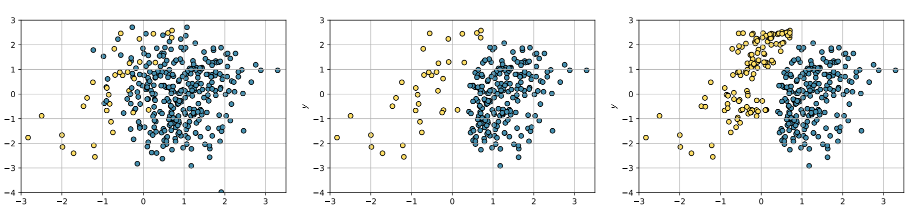
Once the cleaning procedure (undersampling) is performed, we execute the oversampling phase. The created *bba*s at the first step are further exploited in this stage, to intelligently create new synthetic minority samples.

Similarly to the cleaning step, the minority objects are categorized into three possible difficulties: overlapping, label noise, or outlier. The object is considered as "safe" if it does not belong to any of the three types. Thus, it does not need to participate in the oversampling phase. The same goes for both label noise and outlier, since using these types to create synthetic data could result in overgeneralization, which is a major drawback for oversampling [14]. Although, samples belonging to overlapping signify that they are located at the borders of the minority class. Hence, those are the samples that we are most uncertain about, and should focus on in the oversampling phase. Much like other popular oversampling methods such as BorderlineSMOTE [14], our goal is to emphasize the borders of the minority class in order to further improve its visibility.

More formally, let P denote the set of minority examples whose highest mass is committed to $m(U)$ (the set of ambiguous minority objects). We firstly compute the k minority nearest neighbors for each object in P. In this step, we generate $|P| * s$ synthetic minority points, where s is a value between 1 and k. In other words, for each minority instance in P, we randomly select s samples from its k minority nearest neighbors. Finally, s new synthetic points are generated anywhere on the lines joining the samples in P and its s randomly selected neighbors:

$$\overrightarrow{new} = \overrightarrow{a} + w * (\overrightarrow{b} - \overrightarrow{a}) \tag{10}$$

where $\overrightarrow{a}$ is the sample in P, $\overrightarrow{b}$ is a selected minority neighbor, and w represents a random value between 0 and 1. This procedure is repeated for each sample in P, similarly to the SMOTE algorithm (more details in [6]).



(a) Original imbalanced and uncertain dataset. (b) Cleaning the majority class. (c) Selective minority oversampling.

Fig. 2: An imbalanced binary example showing the behavior of our proposed algorithm at each step.

Furthermore, we tested our resampling method on a two-dimensional imbalanced dataset in order to showcase the behavior of the algorithm at each step (see Figure 2).

5 Experimental study

In this section, we will describe firstly the setup of the conducted experiments in subsection 5.1. Lastly, we will present the results and discuss them in subsection 5.2.

5.1 Setup

Datasets For the purpose of this evaluation, we selected binary imbalanced datasets from the keel repository [1]. The datasets are further detailed in Table 1. The imbalance ratio was calculated as $IR = \frac{\#majority}{\#minority}$. The variations of the different parameters (IR, features, and size) allowed for experimenting in different real world settings.

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

Datasets	Imbalance ratios (IR)	Features	Samples
wisconsin	1.860	9	683
glass0	2.060	9	214
vehicle3	2.990	18	846
ecoli1	3.360	7	336
yeast3	8.100	8	1484
page-blocks0	8.790	10	5472
ecoli-0-6-7 vs 3-5	9.090	7	222
yeast-0-3-5-9 vs 7-8	9.120	8	506
ecoli-0-2-6-7 vs 3-5	9.180	7	224
ecoli-0-1-4-7 vs 2-3-5-6	10.590	7	336
glass4	15.460	9	214
yeast-2 vs 8	23.100	8	482
yeast5	32.730	8	1484
kr-vs-k-zero vs eight	53.070	6	1460
abalone-20 vs 8-9-10	72.690	8	1916

Baseline classifier. As baseline, we use the decision tree classifier, more specifically CART [4]. The implementation provided in the scikit-learn machine learning python library [28] was used, with the default parameters.

Metrics and evaluation strategy To appropriately assess the methods in imbalanced scenarios, we use the G-Mean (GM) [2]. The GM is a popular measure for evaluating classifiers in imbalanced settings. It is calculated as the geometric mean of sensitivity and specificity:

$$\text{G-Mean} = \sqrt{sensitivity \times specificity} \tag{11}$$

In order to ensure the fairness of the observed results, we adopt a 10-fold stratified cross validation to eliminate inconsistencies. The dataset is split into 10 parts taking into account the class distribution, 90% of which is the training set, the rest is the test set, and the average of G-mean is taken as the final result. It is worth noting that at each fold, resampling was performed only on the training set.

To better evaluate the significance of the results, statistical analysis was run using the Wilcoxon's signed rank tests [36] for the significance level of $\alpha = 0.05$.

Reference methods and parameters. We compared our proposed method (UAHS) against 4 well known re-sampling methods, in addition to Baseline (BL). The compared methods are SMOTE [6], SMOTE-IPF [30], SMOTE-ENN [3], and SMOTE-TL [3]. The SMOTE-IPF implementation in Smote-variants [19] was used. For the rest of the methods, the implementations provided by the python toolbox imbalanced-learn [21] were applied.

The following parameters were considered for our proposed method UAHS: α was set to 1 as recommended in [22], 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.

5.2 Results discussion

Results on 15 binary imbalanced datasets are shown in Table 2. The best G-Mean value is marked in bold. Our proposed method UAHS achieved the top performances across 10 out of 15 datasets. It showed clear improvement over the compared resampling methods and baseline across complex datasets, especially with high imbalance degree and class overlapping. Furthermore, UAHS performed significantly better in cases where there is a high number of borderline instance. This confirms that our proposal succeeded at emphasizing on the visibility of the minority class, and improving its borders by cleaning the uncertain samples present in the overlapping and noisy regions of the majority class.

The results for Wilcoxon's pairwise test are shown in Table 3. $R+$ represents the sum of ranks in favor of UAHS, $R-$, the sum of ranks in favor of the reference methods, and exact p-values are calculated for each comparison. All comparisons

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

Datasets	BL	SMOTE	SMOTE-IPF	SMOTE-ENN	SMOTE-TL	UAHS
wisconsin	0.934	0.928	0.924	0.950	0.932	**0.964**
glass0	0.737	**0.789**	0.772	0.763	0.782	0.786
vehicle3	0.689	0.660	0.682	**0.727**	0.679	0.673
ecoli1	0.816	0.849	0.856	0.871	0.836	**0.883**
yeast3	0.791	0.834	0.854	0.878	0.849	**0.921**
page-blocks0	0.904	0.928	0.920	0.930	0.923	**0.958**
ecoli-0-6-7 vs 3-5	0.764	0.721	0.769	0.772	0.782	**0.790**
yeast-0-3-5-9 vs 7-8	0.532	0.620	0.606	**0.677**	0.582	0.613
ecoli-0-2-6-7 vs 3-5	0.762	**0.783**	0.781	0.751	0.780	0.761
ecoli-0-1-4-7 vs 2-3-5-6	0.735	0.797	**0.878**	0.857	0.861	0.835
glass4	0.735	0.637	0.706	0.658	0.706	**0.838**
yeast-2 vs 8	0.547	0.666	0.664	0.722	0.734	**0.737**
yeast5	0.840	0.830	0.812	0.912	0.855	**0.932**
kr-vs-k-zero vs eight	0.976	**1.000**	**1.000**	0.999	**1.000**	**1.000**
abalone-20 vs 8-9-10	0.477	0.695	0.602	0.679	0.587	**0.743**

can be considered as statistically significant under a level of 5% since all p-values are lower than the threshold 0.05. This reveals statistically significant improvements by our method against SMOTE, SMOTE-IPF, SMOTE-ENN, SMOTE-TL, and baseline.

Table 3: Wilcoxon's signed ranks test results comparing the G-Mean scores for CART.

Comparisons	$R+$	$R-$	p-value
UAHS vs BL	117.0	3.0	0.000153
UAHS vs SMOTE	98.0	22.0	0.002364
UAHS vs SMOTE-IPF	95.0	25.0	0.004187
UAHS vs SMOTE-ENN	94.0	26.0	0.027679
UAHS vs SMOTE-TL	83.5	36.5	0.004377

6 Conclusion

Solutions for imbalanced datasets are increasingly being applied to critical real world domains. In order to deal with such scenarios, the proposed methods should also handle the uncertainty in the data. In this work, we propose an Uncertainty-Aware Hybrid resampling which combines undersampling and oversampling phases to efficiently re-balance binary datasets. We use an evidential structure to represent soft labels for each sample in the dataset. These representations are later used to remove majority samples which are ambiguous and noisy, and to select minority observations at the borders to generate new minority points.

For future work, we plan to further optimize our method using heuristic methods in order to approximate the amount of instances which should be cleaned, and the number of instances to generate.

References

1. Alcala-Fdez, J., Fernández, A., Luengo, J., Derrac, J., Garcia, S., Sanchez, L., Herrera, F.: Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic and Soft Computing **17**, 255–287 (2010)
2. Barandela, R., Valdovinos, R.M., Sánchez, J.S.: New applications of ensembles of classifiers. Pattern Analysis & Applications **6**(3), 245–256 (2003)
3. Batista, G., Prati, R., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explorations **6**, 20–29 (2004)
4. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and regression trees. CRC press (1984)
5. Bridge, J., Meng, Y., Zhao, Y., Du, Y., Zhao, M., Sun, R., Zheng, Y.: Introducing the gev activation function for highly unbalanced data to develop covid-19 diagnostic models. IEEE Journal of Biomedical and Health Informatics **24**(10), 2776–2786 (2020)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002)
7. Chawla, N.V., Japkowicz, N., Drive, P.: Editorial : Special Issue on Learning from Imbalanced Data Sets. ACM SIGKDD Explorations Newsletter **6**(1), 1–6 (2004)
8. Dempster, A.P.: A generalization of bayesian inference. Journal of the Royal Statistical Society: Series B (Methodological) **30**(2), 205–232 (1968)
9. Douzas, G., Bacao, F., Last, F.: Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Information Sciences **465**, 1–20 (2018)
10. Feng, Y., Zhou, M., Tong, X.: Imbalanced classification: an objective-oriented review. arXiv preprint arXiv:2002.04592 (2020)
11. Grina, F., Elouedi, Z., Lefevre, E.: A preprocessing approach for class-imbalanced data using smote and belief function theory. In: International Conference on Intelligent Data Engineering and Automated Learning. pp. 3–11. Springer (2020)
12. Grina, F., Elouedi, Z., Lefevre, E.: Evidential undersampling approach for imbalanced datasets with class-overlapping and noise. In: The 18th International Conference on Modeling Decisions for Artificial Intelligence. Springer (2021)
13. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications **73**, 220–239 (2017)
14. Han, H., Wang, W.Y., Mao, B.H.: Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. Lecture Notes in Computer Science **3644**(PART I), 878–887 (2005)
15. Ivan, T.: Two modifications of cnn. IEEE transactions on Systems, Man and Communications, SMC **6**, 769–772 (1976)
16. Japkowicz, N.: Class imbalances: are we focusing on the right issue. In: Workshop on Learning from Imbalanced Data Sets II. vol. 1723, p. 63 (2003)

17. Khoshgoftaar, T.M., Rebours, P.: Improving software quality prediction by noise filtering techniques. Journal of Computer Science and Technology **22**(3), 387–396 (2007)
18. Korkmaz, S.: Deep learning-based imbalanced data classification for drug discovery. Journal of Chemical Information and Modeling **60**(9), 4180–4190 (2020)
19. Kovács, G.: Smote-variants: A python implementation of 85 minority oversampling techniques. Neurocomputing **366**, 352–354 (2019)
20. Koziarski, M., Woźniak, M., Krawczyk, B.: Combined cleaning and resampling algorithm for multi-class imbalanced data with label noise. Knowledge-Based Systems **204**, 106223 (2020)
21. Lemaître, G., Nogueira, F., Aridas, C.K.: Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research **18**(1), 559–563 (2017)
22. Liu, Z.g., Pan, Q., Dezert, J., Mercier, G.: Credal classification rule for uncertain data based on belief functions. Pattern Recognition **47**(7), 2532–2541 (2014)
23. Ma, L., Fan, S.: CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests. BMC Bioinformatics **18**(1), 1–18 (2017)
24. Mahalanobis, P.C.: On the generalized distance in statistics. vol. 2, pp. 49–55. National Institute of Science of India (1936)
25. Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.S., Zeineddine, H.: An experimental study with imbalanced classification approaches for credit card fraud detection. IEEE Access **7**, 93010–93022 (2019)
26. Ofek, N., Rokach, L., Stern, R., Shabtai, A.: Fast-CBUS: A fast clustering-based undersampling method for addressing the class imbalance problem. Neurocomputing **243**, 88–102 (2017)
27. Pawlak, Z.: Rough sets. International journal of computer & information sciences **11**(5), 341–356 (1982)
28. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
29. Ramentol, E., Caballero, Y., Bello, R., Herrera, F.: SMOTE-RSB *: A hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. Knowledge and Information Systems **33**(2), 245–265 (2012)
30. 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. Information Sciences **291**(C), 184–203 (2015)
31. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton university press (1976)
32. Smets, P.: The nature of the unnormalized beliefs encountered in the transferable belief model. In: Uncertainty in artificial intelligence. pp. 292–297. Elsevier (1992)
33. Smets, P.: The Transferable Belief Model for Quantified Belief Representation, pp. 267–301. Springer Netherlands, Dordrecht (1998)
34. Tsai, C.F., Lin, W.C., Hu, Y.H., Yao, G.T.: Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. Information Sciences **477**, 47–54 (2019)
35. Vuttipittayamongkol, P., Elyan, E.: Improved overlap-based undersampling for imbalanced dataset classification with application to epilepsy and parkinson's disease. International journal of neural systems **30**(08), 2050043 (2020)

36. Wilcoxon, F.: Individual comparisons by ranking methods. In: Breakthroughs in statistics, pp. 196–202. Springer (1992)
37. Wilson, D.L.: Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics (3), 408–421 (1972)