Evidential undersampling approach for imbalanced datasets with class-overlapping and noise

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. The class imbalance issue involves many real world domains such as fraud detection, medical diagnosis, intrusion detection, etc. Most classification algorithms tend to perform poorly when the training dataset is class-imbalanced. This problem gets more challenging in the presence of other factors such class-overlapping and noise. Among many methods, undersampling is a simple and efficient approach which re-balances the imbalanced dataset by removing majority samples. In this paper, we propose a novel method named Evidential Undersampling (EVUS), which is a re-sampling approach based on the theory of evidence. To avoid removing meaningful samples, each majority object is assigned a soft evidential label to gain more information about its location, then majority samples which are considered ambiguous or noisy by our framework, are eliminated from the training set. The conducted results with CART and SVM show that our proposal outperformed other well-known undersampling methods according to the AUC metric.

Keywords: Undersampling · Imbalanced data · Evidence theory

1 Introduction

In real-world classification tasks, instances are not always evenly distributed among classes. This form of distributions is typically referred to as *imbalanced* or *skewed*. In a binary class-imbalanced dataset, the class with the higher size is called the *majority* class, whereas the rare class is regarded as the *minority* class. Imbalanced classification has been reported in many domains such as medical diagnosis [4], fraudulent credit card detection [25], drug discovery [18], etc. From an application perspective, misclassifying a minority example is more critical than misclassifying a majority example [6]. For instance, failing to recognize a rare disease can be crucial. Class-imbalanced distributions can be due to a lot of factors, including the domain's background (e.g. rare fraudulent transactions) or data collection (e.g. storage). This significantly deteriorates the performance

of most classifier algorithms, since most of them assume an even distribution of the classes [14].

Over the years, many methods have been proposed to deal with imbalanced classification [12]. Generally, existent solutions can be categorized into two main groups: data-level [9], algorithm-level [14], as well as combinations of the two strategies. Algorithm-level methods involve modifying the classifier algorithm to adapt it for imbalanced datasets. It can also contain cost-sensitive solutions [1] and ensemble methods [29]. Data-level approaches typically change the class distribution of the training dataset by adding synthetic samples (oversampling), removing majority examples (undersampling), or both. One advantage is that data-level methods are independent of the used classifier. In other words, they are more flexible and do not require deep understanding of learning algorithms. Oversampling techniques are becoming more expensive in terms of complexity and memory as the amount of data is substantially increasing. In many classification cases, it is more effective to perform undersampling.

The most naive form of undersampling is random undersampling (RUS) [12], which randomly eliminates majority class objects to improve the class distribution of the training set. Nonetheless, it is possible that this method may remove potentially meaningful information from the dataset. To avoid this, many methods have been proposed to intelligently select unessential majority points for elimination. Some works used traditional filtering techniques, such as Edited Nearest Neighbors (ENN) [37], Nearest Neighbor Rule (CNN) [3], and Tomek Links (TL) [15]. These strategies discard majority points based on their nearest neighbors. Undersampling based on evolutionary algorithms were also proposed. In [10], authors implemented evolutionary prototype selection to create an improved subset of the majority class. ACOSampling [40] is another evolutionary-based undersampling technique, which makes use of ant colony optimization (ACO) [7]. More recently, clustering-based methods were developed [21,22,27,35]. In [22], the authors used the k-means algorithm [16] to reduce the majority class size by only selecting the cluster centers for training. CBIS [35] is another clustering-based undersampling approach which uses clustering analysis to divide the majority class into sub-classes, and remove the unessential points in each sub-class.

Although studies have confirmed that a fair distribution of the classes typically performs better [5,14], it is important to note that class imbalance is usually not a concern when the classes are easily separable. However, real-world datasets tend to have class-overlapping regions. This issue can elevate the complexity of the classification, especially when allied with class imbalance. In addition, noise can also amplify the class imbalance problem [30], since rare instances and noise have similar characteristics. Thus, they may be treated as the same pattern.

In this paper, we propose a new undersampling method based on the evidence theory [31], which was recently used for oversampling [11]. The intuition of our proposed method is to improve the visibility of the minority class region in binary imbalanced datasets. To do that, we assign a soft evidential label to each majority class sample, in order to acquire information about their locations. Then, we

eliminate the majority objects that are considered ambiguous (in overlapping regions), label noise (in the minority area), or outliers (far from both classes). This will not only improve the imbalance ratio, but also reduce the amount of overlap and noise in the training dataset. The considered evidential structure based on the theory of evidence, is suitable for our objective, since it provides membership values towards classes, in addition to a belief mass assigned to meta-classes (both classes). This flexibility helps us develop precise rules to detect the unwanted samples for undersampling, with the possibility of tuning each rule individually.

The remainder of this paper will be divided as follows. The theory of evidence will be recalled in Section 2. Section 3 presents our idea, detailing each step. Experimental evaluation and discussion are conducted in Section 4. Our paper ends with a conclusion and an outlook on future work in Section 5.

2 Evidence theory

The theory of evidence [8, 31, 33], also referred to as belief function theory or Dempster-Shafer theory (DST), is a flexible and well-founded framework for representing and combining uncertain information. The frame of discernment denotes a finite set of M exclusive possible events, e.g., possible class labels for an object in a classification problem. The frame of discernment is denoted as follows:

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

A basic belief assignment (*bba*) represents the amount of belief given by a source of evidence, committed to 2^Ω, that is, all subsets of the frame including the whole frame itself. Formally, 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 $m(A)$ measures the amount of belief allocated to a proposition A of Ω. A *bba* is called unnormalized if the sum of its masses is not equal to 1, and should 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:

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

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

3 Evidential undersampling approach (EVUS)

EVUS starts by assigning soft labels to each majority point using the credal classification rule (CCR) introduced in [23]. Generally, it firstly consists of

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 computed *bba* is later used for undersampling.

The remaining of this section will provide detailed descriptions of each step.

3.1 Determination of centers

The simple approach to calculating class centers is to compute the mean value of the training data in the corresponding class. For the overlapping region, which is represented by a meta-class, the center is defined by the barycenter of the involved class centers as follows:

$$C_U = \frac{1}{|U|} \sum_{\omega_i \in U} C_i \tag{4}$$

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

3.2 Computing the soft labels

The evidential membership of each majority 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 element ω_0 is included in the frame explicitly to represent the outlier, i.e., the unknown class.

Let x_s be a sample belonging to the majority class. Each class center represents a piece of evidence to the evidential membership of the majority sample. The mass values regarding the class memberships of x_s should depend on $d(x_s, C)$, i.e., the distance between x_s and the corresponding center of the class. The greater the distance, the lower the mass value. Hence, if x_s is more close to a specific class center, it means that x_s belongs very likely to the respective class. Thus, the initial (unnormalized) masses should be represented by decreasing distance based functions. To deal with anisotropic datasets, the Mahalanobis distance [24] is used in this work as recommended by [23].

The unnormalized masses are calculated as follows:

$$\hat{m}(\{\omega_i\}) = e^{-d(x_s, C_i)}, \quad i \in [1, 2] \tag{5}$$

$$\hat{m}(U) = e^{-\gamma\ \lambda\ d(x_s, C_U)}, \quad U = \{\omega_1, \omega_2\} \tag{6}$$

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

where $\lambda = \beta\ 2^{\alpha}$. A recommended value for $\alpha = 1$ can be used 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 (see Section 3.3). 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.

Finally, the previous unnormalized masses are normalized as follows:

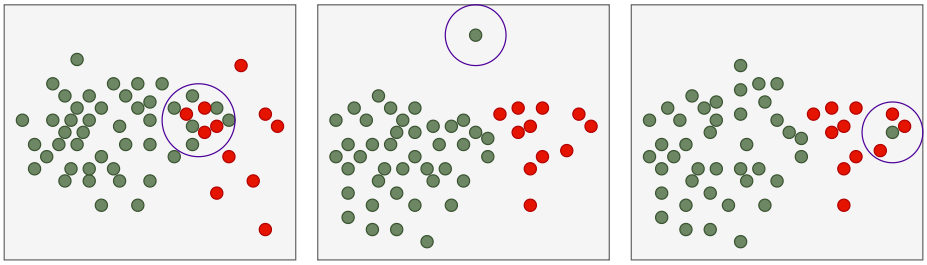
$$m(A) = \frac{\hat{m}(A)}{\sum_{B \subseteq \Omega} \hat{m}(B)}, \qquad \forall A \subseteq \Omega \tag{8}$$

3.3 Selecting majority samples for elimination

Once basic belief assignments are created, the soft memberships are used to reject samples that are unwanted from the majority class. Rejection strategies in evidence theory are common in many applications [17, 34]. The amount of information provided by evidential functions helps us determine whether a sample should be rejected or valid for classification.

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.

Overlapping rejection. Ambiguous samples are usually located in regions where there is strong overlap between classes as seen in Figure 1a. Consequently, this type of objects will have a high mass value in $m(U)$ in our framework. Thus, majority samples whose *bba* has the maximum mass committed to $m(U)$ are considered as part of an overlapping region, and are automatically discarded. Additionally, to avoid excessive elimination and allow tuning, it is possible to



(a) Ambiguous samples in an overlapping area.

(b) An outlier far from both classes.

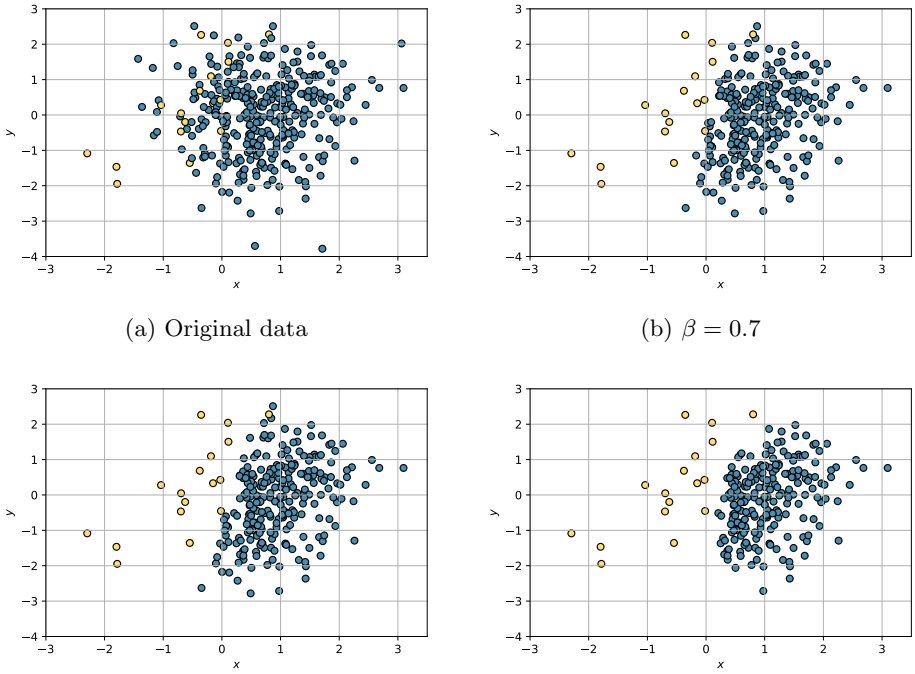
(c) A sample that could be characterized as label noise.

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.

tune the parameter β. The bigger value of β will result in smaller number of objects committed to the overlapping region as seen in Figure 2.

As for majority objects not in overlapping areas (i.e. the highest mass is not committed to $m(U)$), the object is necessarily committed to one of the singletons in Ω ($\{\omega_1\}$, $\{\omega_2\}$, or $\{\omega_0\}$). To make a decision of acceptance or rejection, the *plausibility* function defined in eq. (3) is used. Each majority object x_s is assigned to the class with the maximum plausibility $Pl_{max} = max_{\omega \in \Omega} Pl(\{\omega\})$.

Label noise In EVUS, majority objects should normally have the maximum plausibility committed towards $m(\{\omega_1\})$ which represents the membership value towards the majority class. Accordingly, objects with Pl_{max} committed to $m(\{\omega_2\})$ signify that they are located in the minority region, as illustrated in Figure 1c. In other words, this situation could be characterized by label noise, which is another data difficulty factor that amplifies the class imbalance issue [19]. In our undersampling framework, these types of majority objects are eliminated from the dataset.



(a) Original data

(b) $\beta = 0.7$

(c) $\beta = 0.5$

(d) $\beta = 0.3$

Fig. 2: Undersampling made by EVUS on a synthetic imbalanced dataset with overlapping. Different tunings of the overlapping parameter β were tested.

Outlier rejection. The final possibility occurs when Pl_{max} is committed to ω_0. This situation describes indecisive samples that are far from both classes and could be considered outliers as shown in Figure 1b. In our framework, majority objects with the maximum plausibility committed towards ω_0 are eliminated from the training dataset. The parameter t in eq (7) can be used for tuning the outlier rejection, although $t = 2$ is recommended for good results on average. The bigger t results in smaller number of outliers, and it is recommended to take $t \in [2, 5]$.

4 Experimental study

In this section, we describe our experimental setup and present the observed results.

4.1 Setup

Datasets. Binary imbalanced datasets were selected from the KEEL repository [2] to conduct the experimental study. Specifically, we have chosen a total of 20 datasets which vary in imbalanced ratios (1.87 to 129.44), number of instances (173 to 4174), number of features (6 to 41). The characteristics are further detailed in Table 1. The imbalance ratios (IR) are calculated as $\frac{\#majority}{\#minority}$. These variations allowed for comparisons in a range of different scenarios. In the case of *kr-vs-k-zero vs eight* and *kddcup-rootkit-imap vs back* datasets, categorical features were encoded as integers before applying undersampling. There was no further preprocessing done.

Baseline classifiers. CART decision tree and Support Vector Machine (SVM) were chosen as baseline classifiers to conduct the comparisons. These learning methods are considered as one of the most used classifiers in class-imbalanced problems [12]. For all experiments, the implementations provided in the scikit-learn machine learning python library [28] were used, with the default parameters unchanged (RBF kernel was used for SVM).

Compared methods. EVUS was compared against 5 other undersampling methods, in addition to baseline (BL). The approaches used are: random undersampling (RUS), Cluster Centroid undersampling (CC) [39], Condensed Nearest Neighbour editing (CNN) [13], One- Sided Selection (OSS) [20], and Near Miss undersampling (NM) [26]. The implementations provided by the python toolbox imbalanced-learn [21] were used for the compared methods.

Metric and evaluation strategy. The Area Under the ROC-Curve (AUC) was used as an evaluation measure. This metric provides a score to determine how well a classifier compensates its true positive and false positive rates. It has been shown to be a reliable assessment criterion for imbalanced classification problems

[38]. The AUC scores were averaged through a 10-fold stratified cross validation to eliminate inconsistencies. It is worth mentioning that undersampling was performed only on the training set at each fold. Finally, statistical comparisons were carried out using the Wilcoxon's signed rank tests [36] to further evaluate the significance of the results.

Parameters The following parameters were considered for EVUS: α was set to 1 as recommended in [23], the outlier tuning parameter t was fixed to 2 to obtain averagely good results, and we tested three different values for β in $\{0.3, 0.5, 0.7\}$ and selected the most performing value each time, since the amount of overlapping differs in each dataset. For the other reference methods, the recommended parameter in the respective papers were used.

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
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
glass-0-1-4-6 vs 2	11.060	9	205
glass4	15.460	9	214
yeast-2 vs 8	23.100	8	482
winequality-red-4	29.170	11	1599
winequality-red-8 vs 6	35.440	11	656
kr-vs-k-zero vs eight	53.070	6	1460
winequality-white-3-9 vs 5	58.280	11	1482
poker-8-9 vs 6	58.400	10	1485
poker-8 vs 6	85.880	10	1477
kddcup-rootkit-imap vs back	100.140	41	2225
abalone19	129.440	8	4174

4.2 Results and discussion

Table 2 presents the AUC scores obtained by CART and SVM on each imbalanced dataset after performing undersampling. The best scores are marked

in bold. We can initially observe that undersampling improved the classification performance on all cases. Our proposed method achieves the best AUC scores in 13 out of 20 datasets for CART, and 11 out of 20 in the case of SVM. In 7 out of 20 datasets, EVUS performed better compared to the other methods for both classifiers. This can imply that the choice of the classifier did not affect much the end results. Furthermore, EVUS performed relatively better in cases when there are many borderline points in the dataset, i.e., when the overlapping between the classes is strong. By contrast, one can say that EVUS is not much of use in the cases of small number of borderline majority points.

Table 2: AUC results for KEEL datasets using CART and SVM.

	CART							*SVM*						
Datasets	BL	RUS	CC	CNN	OSS	NM	EVUS	BL	RUS	CC	CNN	OSS	NM	EVUS
wisconsin	0.935	0.945	0.939	0.910	0.914	0.936	**0.964**	0.971	0.972	0.971	0.968	0.971	0.967	**0.977**
glass0	0.753	0.758	0.747	0.762	0.767	0.673	**0.770**	0.500	**0.670**	0.656	0.500	0.500	0.485	0.667
vehicle3	0.708	0.720	0.709	0.675	0.672	0.649	**0.789**	0.500	0.606	**0.662**	0.648	0.517	0.639	0.619
ecoli1	0.826	0.842	0.848	0.852	0.876	0.804	**0.886**	0.861	0.902	0.889	**0.904**	0.875	0.838	**0.904**
yeast3	0.808	**0.897**	0.869	0.824	0.811	0.768	0.882	0.841	0.854	**0.923**	0.890	0.865	0.846	0.891
ecoli-0-6-7 vs 3-5	0.829	0.776	0.821	0.841	0.799	0.758	**0.852**	0.839	0.804	0.837	0.868	0.845	0.726	**0.886**
yeast-0-3-5-9 vs 7-8	0.659	0.614	0.544	0.616	**0.665**	0.546	0.663	0.597	0.728	0.676	0.597	0.607	0.538	**0.645**
ecoli-0-2-6-7 vs 3-5	0.822	0.790	0.797	0.825	0.809	0.748	**0.829**	0.839	0.827	0.862	**0.896**	0.842	0.670	0.867
ecoli-0-1-4-7 vs 2-3-5-6	0.777	0.756	0.814	0.822	0.819	0.735	**0.852**	0.840	0.869	0.902	0.850	0.815	0.674	**0.906**
glass-0-1-4-6 vs 2	0.571	0.680	**0.685**	0.590	0.560	0.601	0.596	0.500	**0.630**	0.585	0.500	0.500	0.473	0.600
glass4	0.838	0.850	0.716	0.743	0.698	0.875	**0.898**	0.500	0.788	0.843	0.500	0.500	0.573	**0.895**
yeast-2 vs 8	0.714	0.678	0.553	0.683	0.762	0.718	**0.764**	0.774	0.772	0.772	0.774	0.774	**0.816**	0.782
winequality-red-4	0.540	**0.595**	0.568	0.593	0.579	0.456	0.588	0.500	0.524	0.546	0.500	0.500	0.421	**0.556**
winequality-red-8 vs 6	0.612	0.695	0.713	0.759	0.637	0.454	**0.784**	0.500	0.662	0.546	0.500	0.500	0.356	**0.566**
kr-vs-k-zero vs eight	0.822	0.962	0.778	**1.000**	**1.000**	0.960	**1.000**	0.791	0.884	**0.959**	0.807	0.791	0.796	0.931
winequality-white-3-9 vs 5	0.630	0.644	0.558	0.595	0.612	0.542	**0.648**	0.500	**0.636**	0.550	0.500	0.500	0.526	0.585
poker-8-9 vs 6	0.527	0.543	0.630	0.563	0.543	**0.661**	0.573	0.500	0.525	0.497	0.500	0.500	0.531	**0.629**
poker-8 vs 6	0.549	0.511	**0.635**	0.496	0.547	0.541	0.548	0.500	0.522	0.515	0.500	0.500	**0.536**	0.525
kddcup-rootkit-imap vs back	0.992	0.997	0.561	**1.000**	0.911	0.560	**1.000**	0.997	**1.000**	0.998	0.550	0.973	0.999	**1.000**
abalone19	0.529	0.673	**0.690**	0.553	0.528	0.522	0.629	0.500	0.623	0.626	0.500	0.500	0.591	**0.665**

To assess the significance of the comparisons, Table 3 presents the statistical analysis made by Wilcoxon's signed ranks test. $R+$ represents the sum of ranks in favor of EVUS, $R-$, the sum of ranks in favor of the other compared methods, and p-values are calculated for each comparison. As shown in Table 3, all p-values are lower than 0.10.

Table 3: Wilcoxon's signed ranks test results comparing the AUC scores for both CART and SVM.

	CART			*SVM*		
Comparisons	$R+$	$R-$	p-value	$R+$	$R-$	p-value
EVUS vs RUS	173.5	36.5	0.009436	133.5	76.5	0.063069
EVUS vs CC	164.0	46.0	0.013321	154.0	56.0	0.03479
EVUS vs CNN	169.0	41.0	0.00015	177.0	33.0	0.000518
EVUS vs OSS	187.5	22.5	0.000107	210.0	0.0	<0.000001
EVUS vs NM	200.0	10.0	0.000041	198.0	12.0	0.000067

Thus, one can say that EVUS outperformed the compared techniques at a significance level of $\alpha = 0.10$.

5 Conclusions

Throughout this paper, we have proposed a new method called Evidential Undersampling (EVUS), based on the evidence theory. Majority samples are selected for elimination based on soft evidential labels, which provide us with more information about the point's location. This resulted in AUC improvements over well-known undersampling methods. The main motivation behind our proposed algorithm was to improve the visibility of the minority class regions, to avoid the misclassification of minority objects.

For future work, we intend to explore heuristic methods to further optimize the parameters β and t. This will provide a more adaptive behavior of the parameters based on the amount of overlap and noise present in the majority class.

References

1. Alaba, P.A., Popoola, S.I., Olatomiwa, L., Akanle, M.B., Ohunakin, O.S., Adetiba, E., Alex, O.D., Atayero, A.A., Daud, W.M.A.W.: Towards a more efficient and cost-sensitive extreme learning machine: A state-of-the-art review of recent trend. Neurocomputing **350**, 70–90 (2019)
2. 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)
3. Angiulli, F.: Fast condensed nearest neighbor rule. In: Proceedings of the 22nd International Conference on Machine Learning. p. 25–32. ICML'05, Association for Computing Machinery, New York, NY, USA (2005)
4. 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)
5. Chawla, N.V.: C4.5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure. In: Proceedings of the ICML'03 Workshop on Class Imbalances (2003)
6. 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)
7. Colorni, A., Dorigo, M., Maniezzo, V.: Distributed optimization by ant colonies. In: Proceedings of the first European conference on artificial life. vol. 142, pp. 134–142. Paris, France (1991)
8. Dempster, A.P.: A generalization of bayesian inference. Journal of the Royal Statistical Society: Series B (Methodological) **30**(2), 205–232 (1968)
9. Feng, Y., Zhou, M., Tong, X.: Imbalanced classification: an objective-oriented review. arXiv preprint arXiv:2002.04592 (2020)
10. García, S., Herrera, F.: Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. Evolutionary computation **17**(3), 275–306 (2009)

11. Grina, F., Elouedi, Z., Lefevre, E.: A preprocessing approach for class-imbalanced data using smote and belief function theory. In: Proceedings of International Conference on Intelligent Data Engineering and Automated Learning. pp. 3–11. Springer (2020)
12. 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)
13. Hart, P.: The condensed nearest neighbor rule. IEEE transactions on information theory **14**(3), 515–516 (1968)
14. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering **21**(9), 1263–1284 (2009)
15. Ivan, T.: Two modification of cnn. IEEE transactions on Systems, Man and Communications, SMC **6**, 769–772 (1976)
16. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: Analysis and implementation. IEEE transactions on pattern analysis and machine intelligence **24**(7), 881–892 (2002)
17. Kessentini, Y., Burger, T., Paquet, T.: A Dempster–Shafer theory based combination of handwriting recognition systems with multiple rejection strategies. Pattern recognition **48**(2), 534–544 (2015)
18. Korkmaz, S.: Deep learning-based imbalanced data classification for drug discovery. Journal of Chemical Information and Modeling **60**(9), 4180–4190 (2020)
19. Koziarski, M., Woźniak, M., Krawczyk, B.: Combined Cleaning and Resampling Algorithm for Multi-Class Imbalanced Data with Label Noise. arXiv (2020)
20. Kubat, M., Matwin, S.: Addressing the curse of imbalanced training sets: One-sided selection. In: Proceedings of the Fourteenth International Conference on Machine Learning. ICML'97, vol. 97, pp. 179–186 (1997)
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. Lin, W.C., Tsai, C.F., Hu, Y.H., Jhang, J.S.: Clustering-based undersampling in class-imbalanced data. Information Sciences **409-410**, 17–26 (2017)
23. 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)
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. Mani, I., Zhang, I.: knn approach to unbalanced data distributions: a case study involving information extraction. In: Proceedings of workshop on learning from imbalanced datasets. vol. 126 (2003)
27. 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)
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. Ribeiro, V.H.A., Reynoso-Meza, G.: Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets. Expert Systems with Applications **147** (2020)
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. Tong, Z., Xu, P., Denœux, T.: ConvNet and Dempster-Shafer theory for object recognition. In: International Conference on Scalable Uncertainty Management. pp. 368–381. Springer (2019)
35. 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)
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)
38. Xue, J.H., Hall, P.: Why does rebalancing class-unbalanced data improve auc for linear discriminant analysis? IEEE transactions on pattern analysis and machine intelligence **37**(5), 1109–1112 (2014)
39. Yen, S.J., Lee, Y.S.: Cluster-based under-sampling approaches for imbalanced data distributions. Expert Systems with Applications **36**(3), 5718–5727 (2009)
40. Yu, H., Ni, J., Zhao, J.: Acosampling: An ant colony optimization-based undersampling method for classifying imbalanced dna microarray data. Neurocomputing **101**, 309–318 (2013)