Learning from imbalanced data using an evidential undersampling-based ensemble

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

[1] Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, 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. In many real-world binary classification problems, one class tends to be heavily underrepresented when it consists of far fewer observations than the other class. This results in creating a biased model with undesirable performance. Different techniques, such as undersampling, have been proposed to fix this issue. Ensemble methods have also been proven to be a good strategy to improve the performance of the resulting model in the case of class imbalance. In this paper, we propose an evidential undersampling-based ensemble approach. To alleviate the issue of losing important data, our undersampling technique assigns soft evidential labels to each majority instance, which are later used to discard only the unwanted observations, such as noisy and ambiguous examples. Finally, to improve the final results, the proposed undersampling approach is incorporated into an evidential classifier fusion-based ensemble. The comparative study against well-known ensemble methods reveal that our method is efficient according to the G-Mean and F-Score measures.

Keywords: Imbalanced classification · Ensemble learning · Undersampling · Evidence theory

1 Introduction

Imbalanced classification is a common issue in modern machine learning problems. In binary classification, it is a scenario that occurs when a class, refereed to as the minority class, is highly under-represented in the dataset, while the other class represents the majority [14]. Due to the naturally-skewed class distributions, class imbalance has been widely observed in many real-world applications, such as medical diagnosis [15], network intrusion detection [10], language translation [20], and fraud detection [23]. From a practical point of view, the minority class usually yields higher interests. For example, failing to detect a fraudulent transaction can be crucial to a banking organization.

In addition to the skewed class distribution, the complexity of the data is an important factor for classification models. Other related data imperfections

include data uncertainty, i.e., class overlapping (ambiguity) and noise. The data uncertainty issue was proven to increase the difficulty for classifiers to yield good performance on imbalanced datasets [35].

In order to deal with the poor performance on imbalanced data, many strategies have been developed to deal with this issue. The proposed methods are a variety of re-sampling, classifier modifications, cost-sensitive learning, or ensemble approaches [12]. Data re-sampling is one of the most simple yet efficient strategies to deal with imbalanced classification. Theses methods typically aim at re-balancing the data at the preprocessing level. This gives it the advantage of versatility, for the reason that it is classifier-independent. More recently, ensemble learning is incorporated and combined with different strategies, such as resampling. Ensembles are used to improve a single classifier by combining several base classifiers (also called weak learners) that outperform every independent one.

However, most of these methods have been observed to suffer from limitations, such as the presence of other data imperfections (i.e. high class overlapping, noise and outliers).

In this paper, we propose an evidential undersampling-based ensemble, in which we incorporate an evidential undersampling method into an ensemble learning framework. Instead of randomly undersampling the data, our presented approach uses evidence theory [6, 31], which was recently used for imbalanced classification [11, 25], in order to represent majority observations with soft evidential memberships. Consequently, this gives us an idea on the location of each majority instance. 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). The intuition behind this is to improve the visibility of the minority class, since it usually is the main learning interest in most class-imbalanced problems. The issue with this undersampling solution, is that it is rather difficult to know the exact amount of ambiguity or class overlapping present in the data. The resulting undersampled subset is heavily controlled by our assumption of how much ambiguity is present. To fix this issue, we integrate this evidential method into the process of a bagging ensemble. The goal is to train multiple base classifiers using different subsets created by our version of undersampling. Finally, we use a classifier fusion approach based on the evidence theory.

The remainder of this paper will be divided as follows. The next section presents related works in resampling and ensemble learning. The theory of evidence will be recalled in Section 3. Section 4 details each step of our contribution, i.e., the evidential mechanism used for undersampling and the classifier combination method. Experimental evaluation and discussion are conducted in Section 5. Our paper ends with a conclusion and an outlook on future work in Section 6.

2 Resampling and ensemble methods for imbalanced classification

In this work, we focus on binary imbalanced classification, which is the most widely studied problem in imbalanced learning. In this section, we mainly review existing works relative to resampling and ensemble learning.

2.1 Resampling

Resampling methods focus on modifying the training set to balance the class distribution. It can be categorized into 3 groups: oversampling, undersampling, or hybrid methods. Oversampling techniques aim at creating synthetic minority examples to get rid of the class imbalance. The most traditional version is to randomly replicate existing minority data. To avoid overfitting, Synthetic Minority Oversampling Technique (SMOTE) was suggested [3]. SMOTE creates minority instances by interpolating between existing observations that lie together. Nonetheless, many works [5, 7, 13] have proposed other versions of SMOTE, since SMOTE can cause potential amplification of noise, and overlap already present in the data.

Undersampling is another form of resampling, which eliminates examples from the majority class to re-balance the dataset. Similarly to oversampling, the traditional way of undersampling is randomly selecting majority instances to discard, which may potentially remove meaningful information from the dataset. Henceforth, other methods were presented for safer undersampling. Commonly, filtering techniques, such as the Edited Nearest Neighbors (ENN) [39] and Tomek Links (TL) [16], are occasionally used for undersampling imbalanced data. More recently, other mechanisms have been used, such as clustering [21], evolutionary algorithms [19], and evidence theory [11].

Combining oversampling and undersampling is also a solid solution to imbalanced learning. It usually consists of combining a SMOTE-like method with an undersampling approach [18, 28].

2.2 Ensemble learning in imbalanced classification

The main idea behind ensembles is to improve a single classifier by combining the results of multiple classifiers that outperform every independent one. This paper focuses on resampling-based ensembles, which combines ensemble learning with resampling techniques to tackles class imbalance. Most works considered the use of bagging, boosting, or a combination of the two.

Bagging builds ensembles using the concept of *independent learning*. This strategy trains the base classifiers independently from each other, and uses data re-sampling to introduce diversity into the predictions of the models. While boosting learns of the misclassification of previous iterations by adapting the importance of misclassified objects in future iterations. Random undersampling is popularly used with ensembles [36]. SMOTEBagging and UnderBagging

were suggested in [37]. The former integrates SMOTE's oversampling into the bagging algorithm, with an adaptive way of computing the resampling rate, while Underbagging does the same using random undersampling. In order to optimize the model performance, a hybrid resampling technique was combined with bagging [17].

Boosting-based ensembles have also been proposed for the class imbalance issue. Similar to bagging-based ensembles, these methods merge data re-sampling techniques into boosting algorithms, more specifically the AdaBoost algorithm [9]. SMOTEBoost [4] performs SMOTE during each boosting iteration in order to generate minority objects. RUSBoost [30] is also similar to SMOTEBoost, but it eliminates instances from the majority class by random undersampling in each iteration. Evolutionary algorithms were also used to create a boosting-based algorithm [19]. SMOTEWB [29] is another boosting ensemble, which combines SMOTE with a noise detection method, into a boosting framework.

Some methods have used hybrid approaches involving both boosting and bagging, such as EasyEnsemble and BalanceCascade [22].

3 Evidence theory

The theory of evidence [6,31,33], also called belief function theory or Dempster-Shafer theory (DST), is a flexible and well-founded framework to represent and combine uncertain information. The frame of discernment denotes a finite set of M exclusive possible propositions, 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 (also referred to as *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 $m(A)$ quantifies the amount of belief allocated to a event 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.

To combine several *bba*s, *Dempster*'s rule [6] is a popular choice. Let m_1 and m_2 two BBAs defined on the same frame of discernment Ω, their combination based on *Dempster*'s rule gives the following *bba*:

$$m_1 \oplus m_2(A) = \begin{cases} \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)} & for\ A \neq \emptyset\ and\ A \in 2^{\Omega}. \\ 0 & for\ A = \emptyset. \end{cases} \tag{4}$$

4 Evidential undersampling-based ensemble learning

An evidential undersampling method [11] is incorporated into a bagging ensemble, to create an Ensemble-based Evidential Undersampling (E-EVUS). The main idea is to create diverse undersampled subsets using different assumptions of ambiguity. This will add diversity to the resulting model, by combining various decision boundaries created by each base learner.

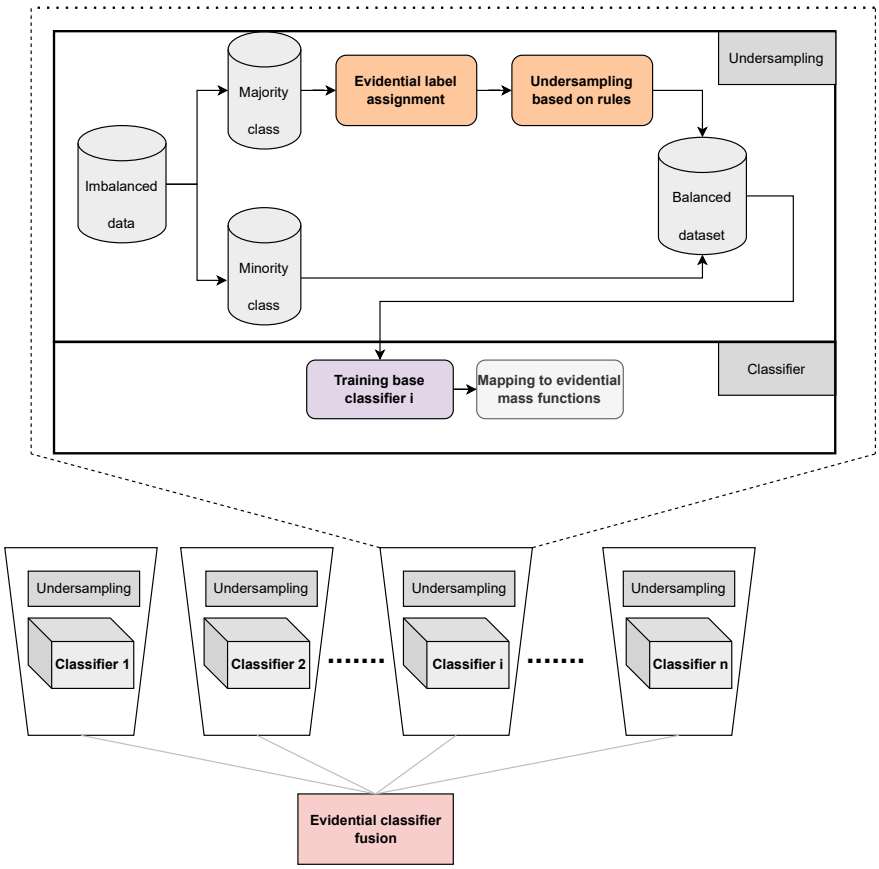


Fig. 1: Evidential Undersampling-based Ensemble learning (E-EVUS)

Our idea is detailed and illustrated in Fig 1. Before training each base classifier, the evidential undersampling process starts by assigning a soft label structure to each majority example. The amount of ambiguity present is randomly selected for each classifier. Then, the selection of instances to eliminate is made based on the location of each instance. The idea is to avoid the loss of important majority data, by only discarding unwanted observations, such as class overlapping, noise and outliers. After performing undersampling, each subset will be used to train a base model. In this paper, we use support vector machines [34] as weak classifiers. However, it is important to note that one can use any other classifier. Finally, we accomplish classifier fusion using an evidential combination, to create the final learning model.

4.1 Evidential label assignment

We recall the method used in [11] to create soft labels, which is based on the credal classification rule (CCR) [24]. It consists of firstly determining the center for each class by simply computing 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 calculated 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 U represents the meta-class, ω_i are the classes involved in U, and C_i is the corresponding center.

The evidential distribution 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.

Let x_s be an observation from the majority class. Each class center represents a piece of evidence to the membership of the majority instance. The mass values for x_s should depend on $d(x_s, C)$, i.e., the distance between x_s and the corresponding class's center. The greater the distance, the lower the mass value. Henceforth, if x_s is more close to a specific class center, it means that x_s belongs very likely to the respective class. Subsequently, the initial (unnormalized) masse values should be represented by decreasing distance based functions. To deal with anisotropic datasets, the Mahalanobis distance is used in this work as recommended by [24].

The unnormalized masses are calculated as follows:

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

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

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

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 $0 < \beta < 1$. The latter parameter is what gives us the ability to control the amount of overlap in the

data, thus, allowing for diversity. In this ensemble framework, the value of β is randomly selected for each base classifier. 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 both 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.

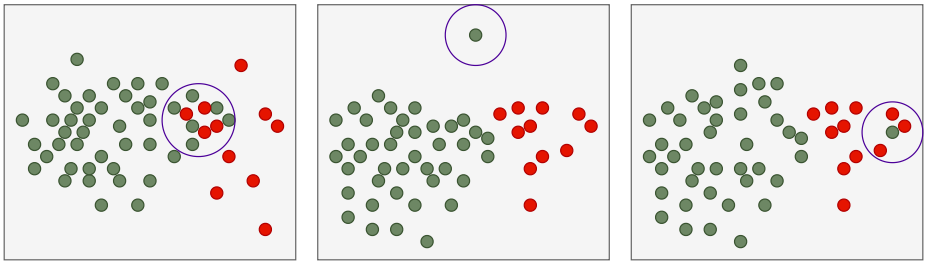
Finally, the calculated unnormalized masses are normalized as follows:

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

4.2 Undersampling

After assigning *bbas*, 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 examples are usually present in regions where there is heavy overlap between classes as seen in Figure 2a. This situation could be described by 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 U are considered as part of an overlapping region, and are automatically discarded. The mass value assigned to U is heavily influenced by the randomly selected parameter β. Henceforth, the higher value of β will result in fewer objects committed to the ambiguous region. As for majority objects whose highest mass is not committed to U (i.e. not in overlapping region), the instance is necessarily committed to one of the singletons in Ω ($\{\omega_1\}$, $\{\omega_2\}$, or $\{\omega_0\}$). In this situation, we use the *plausibility* function defined in eq. (3) 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:** Majority observations should normally have the maximum plausibility committed to ω_1 which measures the membership value towards the majority class. By contrast, having Pl_{max} committed to ω_2 signify that they are located in a minority region, as illustrated in Figure 2c. 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 2b. 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 ω_0 are considered as outliers and removed from the dataset.



(a) Ambiguous samples in an overlapping area.

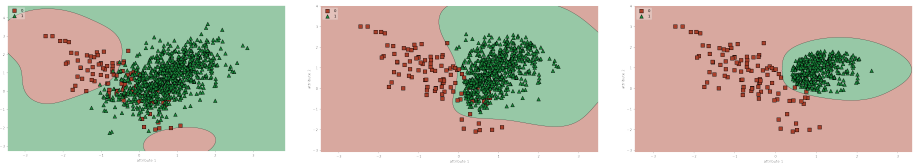
(b) An outlier far from both classes.

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

Fig. 2: 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 Base classifier learning and combination

Our previously presented method achieved good performance in imbalanced classification tasks because it aims at improving the visibility of the minority class, by eliminating the unwanted examples [11]. However, the performance is highly influenced by the selected value for the parameter β, which controls the amount to eliminate from the ambiguous region. To tackle this issue, our evidential undersampling method is included into a bagging ensemble. For each iteration, a different value of the parameter β is randomly selected. As a result, very different subsets are created, as seen in Fig 3. The figure shows the results of undersampling performed on a real binary imbalanced example, before training a SVM classifier. As illustrated, the undersampled subsets can yield very diverse decision boundaries, depending on different ambiguity assumptions.



(a) SVM's decision boundaries on original distribution without resampling.

(b) SVM's decision boundaries on undersampled data using our approach with $\beta = 0.7$.

(c) SVM's decision boundaries on undersampled data using our approach with $\beta = 0.2$.

Fig. 3: Comparing the resulted decision boundaries by SVM after performing our evidential undersampling on different amounts of overlap.

Then, each subset is used to train a base classifier. In this paper, we use the Support Vector Machine (SVM) classifier. It is a popular choice used in many imbalanced learning problems [12]. Each classifier is trained independently in bagging ensembles. Thus, we make the assumption that each model's output is an independent piece of evidence. Henceforth, we can use the Dempster's rule of combination presented in Eq. 4, as suggested in [27]. In our case, the output of each base classifier should be represented by mass functions. For this purpose, we propose to convert SVM's output into probability distributions using Platt scaling [26], before using the inverse pignistic transform [8]. As a result, a mass function is created for each base learner. Thus, the Dempster rule of combination can be applied to create a final combined mass function. Finally, the decision is made by choosing the singleton with the maximum plausibility $Pl_{max} = max_{\omega \in \Omega} Pl(\{\omega\})$.

5 Experimental study

In this section, we will firstly detail 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. We selected 14 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.

Reference methods and parameters. We compared our proposed method (E-EVUS) against well-known ensemble-based methods: EasyEnsemble (EASY) [22], RUSBoost [30], and RUSBagging [37]. For each ensemble, we use the base classifier suggested in the respective paper. In our case, we use the SVM classifier, as previously discussed. The following parameters were considered for our proposal: α was set to 1 as recommended in [24], the outlier parameter t for $m(\{\omega_0\})$ was fixed to 2 to obtain good results in average, and the number of base classifiers to train is set to 10. The other methods were set according to the suggested settings by the authors.

Metrics and evaluation strategy To appropriately assess the methods in imbalanced scenarios, we use the G-Mean (GM) [2] and the F1-score, which are popular measures for evaluating classifiers in imbalanced learning. Following the confusion matrix described in Table 2, the evaluation metrics used in this paper are mathematically formulated as follows:

$$Precision = \frac{TP}{TP - FP} \tag{10}$$

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

Datasets	Imbalance ratios (IR)	Features	Samples
wisconsin	1.860	9	683
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
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
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
poker-8-9 vs 6	58.400	10	1485
poker-8 vs 6	85.880	10	1477
abalone-20 vs 8-9-10	72.690	8	1916

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$\text{G-Mean} = \sqrt{sensitivity \times specificity} \tag{13}$$

$$F1 - score = \frac{2 \times sensitivity \times precision}{sensitivity + precision} \tag{14}$$

Table 2: Confusion matrix.

	Predictive Positive (P)	Predictive Negative (N)
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

In order to ensure the fairness of the observed results, we adopt a 5-fold stratified cross validation to eliminate inconsistencies. Finally, statistical comparisons were carried out using the Wilcoxon's signed rank tests [38] to further evaluate the significance of the results.

5.2 Results and discussion

The measured scores G-Mean and F-score are reported in Table 3. The best scores are presented in bold. According to the presented experimental results, we can remark that our approach achieved relatively well for both assessment measures. In fact, the results show that our method has the best scores on G-Mean (10 out of 14) and F1-score (9 out of 14). The two chosen metrics consider the accuracy of both classes, since, as defined in Eq. 13, G-Mean is the square root of the product between the negative accuracy (i.e., specificity), and the positive accuracy (i.e., sensitivity). Meanwhile, the F-measure is based on precision and sensitivity. Therefore, we can initially argue that our proposal E-EVUS indeed improves the learning on the minority class while keeping the accuracy for the majority one.

Table 3: G-Mean and F1-score results for KEEL datasets using different ensemble techniques.

	G-Mean				*F-Measure*			
Datasets	EASY	RUSBagging	RUSBOOST	E-EVUS	EASY	RUSBagging	RUSBOOST	E-EVUS
wisconsin	0.962	0.971	0.945	**0.975**	0.973	**0.976**	0.962	**0.976**
vehicle3	0.736	0.768	0.719	**0.770**	0.817	**0.823**	0.822	0.822
ecoli1	0.782	0.823	0.810	**0.874**	0.905	**0.921**	0.919	0.918
yeast3	0.916	**0.922**	0.801	0.864	0.950	**0.953**	0.951	**0.953**
ecoli-0-6-7 vs 3-5	0.772	0.835	0.619	**0.879**	0.931	0.924	0.924	**0.944**
ecoli-0-2-6-7 vs 3-5	0.847	0.831	0.785	**0.874**	0.935	0.943	**0.950**	**0.950**
ecoli-0-1-4-7 vs 2-3-5-6	0.847	**0.853**	0.718	0.768	0.923	0.931	**0.956**	0.935
glass-0-1-4-6 vs 2	0.580	0.610	0.267	**0.655**	0.818	0.805	0.830	**0.872**
glass4	0.820	0.798	0.782	**0.852**	**0.971**	0.938	0.927	0.965
winequality-red-4	**0.664**	0.632	0.436	0.658	0.776	0.846	0.727	**0.890**
winequality-red-8 vs 6	0.740	0.674	0.368	**0.770**	0.875	0.854	0.893	**0.956**
poker-8-9 vs 6	0.377	0.611	0.177	**0.675**	0.510	0.872	0.881	**0.932**
poker-8 vs 6	0.508	0.639	0.308	**0.717**	0.526	0.755	0.800	**0.882**
abalone19	**0.753**	0.704	0.269	0.723	0.803	0.856	**0.957**	0.934

In the observed results, E-EVUS performed significantly better than the reference methods in highly uncertain datasets, i.e., where there are significant class overlapping, such as *poker-8-9 vs 6*, and *winequality-red-8 vs 6*. This is likely due to the fact that our method succeeded at eliminating the difficult and uncertain majority samples, which allowed for better learning of more difficult minority data examples.

In order to evaluate the significance of the comparisons, Table 4 presents the statistical analysis made by Wilcoxon's signed ranks test. $R+$ represents the sum of ranks in favor of E-EVUS, while $R-$ reflects the sum of ranks in favor of the other reference methods, and p-values are computer for each comparison. As shown in Table 4, almost all p-values are lower than 0.5. Thus, one can say that our method significantly outperformed the other techniques, for both selected metrics, with a signficance level of $\alpha = 0.05$.

Table 4: Wilcoxon's signed ranks test results comparing the G-Mean and F1-score metrics against the compared approaches.

	G-Mean			*F1-score*		
Comparisons	$R+$	$R-$	p-value	$R+$	$R-$	p-value
E-EVUS vs EASY	81.5	23.5	0.078491	101.0	4.0	0.0008544
E-EVUS vs RUSBagging	84.0	21.0	0.049438	103.0	2.0	0.004741
E-EVUS vs RUSBoost	105.0	0.0	0.000122	93.0	12.0	0.0341704

6 Conclusion

In this paper, we propose an evidential undersampling-based ensemble (E-EVUS), in which we use evidence theory to better learn from imbalanced datasets. The goal is to improve the visibility of the minority class by removing unwanted examples, such as noisy and overlapped observations. This technique is incorporated into a bagging ensemble framework, in order to diversify the created subsets. Therefore, it is more likely to improve the final decision boundary of the classifier.

In addition, our experimental study demonstrates that integrating evidential undersampling into ensemble learning, could result to diversity of base models, which facilitates the learning performance. Further investigations can include the use of hybrid resampling into our ensemble learning method.

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. 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)
4. Chawla, N.V., Lazarevic, A., Hall, L.O., Bowyer, K.W.: Smoteboost: Improving prediction of the minority class in boosting. In: European conference on principles of data mining and knowledge discovery. pp. 107–119. Springer (2003)
5. Dablain, D., Krawczyk, B., Chawla, N.V.: Deepsmote: Fusing deep learning and smote for imbalanced data. IEEE Transactions on Neural Networks and Learning Systems (2022)
6. Dempster, A.P.: A generalization of bayesian inference. Journal of the Royal Statistical Society: Series B (Methodological) **30**(2), 205–232 (1968)
7. 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)

8. Dubois, D., Prade, H., Smets, P.: A definition of subjective possibility. International journal of approximate reasoning **48**(2), 352–364 (2008)
9. Freund, Y., Schapire, R.E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences **55**(1), 119–139 (1997)
10. Fu, Y., Du, Y., Cao, Z., Li, Q., Xiang, W.: A deep learning model for network intrusion detection with imbalanced data. Electronics **11**(6), 898 (2022)
11. Grina, F., Elouedi, Z., Lefevre, E.: Evidential undersampling approach for imbalanced datasets with class-overlapping and noise. In: International Conference on Modeling Decisions for Artificial Intelligence. pp. 181–192. Springer (2021)
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. 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)
14. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering **21**(9), 1263–1284 (2009)
15. Huynh, T., Nibali, A., He, Z.: Semi-supervised learning for medical image classification using imbalanced training data. Computer Methods and Programs in Biomedicine p. 106628 (2022)
16. Ivan, T.: Two modification of cnn. IEEE transactions on Systems, Man and Communications, SMC **6**, 769–772 (1976)
17. Jung, I., Ji, J., Cho, C.: Emsm: Ensemble mixed sampling method for classifying imbalanced intrusion detection data. Electronics **11**(9), 1346 (2022)
18. 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)
19. Krawczyk, B., Galar, M., Jeleń, Ł., Herrera, F.: Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. Applied Soft Computing **38**, 714–726 (2016)
20. Li, X., Gong, H.: Robust optimization for multilingual translation with imbalanced data. Advances in Neural Information Processing Systems **34** (2021)
21. 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)
22. Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **39**(2), 539–550 (2008)
23. Liu, Y., Ao, X., Qin, Z., Chi, J., Feng, J., Yang, H., He, Q.: Pick and choose: a gnn-based imbalanced learning approach for fraud detection. In: Proceedings of the Web Conference 2021. pp. 3168–3177 (2021)
24. 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)
25. Niu, J., Liu, Z.: Imbalance data classification based on belief function theory. In: International Conference on Belief Functions. pp. 96–104. Springer (2021)
26. Platt, J.: Probabilistic outputs for svms and comparisons to regularized likehood methods. In: Advances in Large Margin Classifiers. MIT Press (1999)
27. Quost, B., Masson, M.H., Denœux, T.: Classifier fusion in the dempster–shafer framework using optimized t-norm based combination rules. International Journal of Approximate Reasoning **52**(3), 353–374 (2011)

28. 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**, 184–203 (2015)
29. Sağlam, F., Cengiz, M.A.: A novel smote-based resampling technique trough noise detection and the boosting procedure. Expert Systems with Applications **200**, 117023 (2022)
30. Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A.: Rusboost: A hybrid approach to alleviating class imbalance. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **40**(1), 185–197 (2009)
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. Vapnik, V.: The nature of statistical learning theory. Springer science & business media (2013)
35. Vuttipittayamongkol, P., Elyan, E., Petrovski, A.: On the class overlap problem in imbalanced data classification. Knowledge-based systems **212**, 106631 (2021)
36. Wallace, B.C., Small, K., Brodley, C.E., Trikalinos, T.A.: Class imbalance, redux. In: 2011 IEEE 11th international conference on data mining. pp. 754–763. IEEE (2011)
37. Wang, S., Yao, X.: Diversity analysis on imbalanced data sets by using ensemble models. In: 2009 IEEE Symposium on Computational Intelligence and Data Mining. pp. 324–331. IEEE (2009)
38. Wilcoxon, F.: Individual comparisons by ranking methods. In: Breakthroughs in statistics, pp. 196–202. Springer (1992)
39. Wilson, D.L.: Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics (3), 408–421 (1972)