Pattern Recognition 118 (2021) 107965

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

A review of methods for imbalanced multi-label classification

Adane Nega Tarekegn[a,d,*], Mario Giacobini[b], Krzysztof Michalak[c]

[a] Modelling and Data Science Program, Department of Mathematics, University of Turin, Italy
[b] Data Analysis and Modeling Unit, Department of Veterinary Sciences, University of Turin, Italy
[c] Department of Information Technologies, Wroclaw University of Economics, Poland
[d] Faculty of Computing, Bahir Institute of Technology - Bahir Dar University, Ethiopia

ARTICLE INFO

Article history:
Received 29 January 2020
Revised 18 March 2021
Accepted 26 March 2021
Available online 6 May 2021

Keywords:
Imbalanced Data
Multi-label Classification
Imbalanced Classification
Machine learning
Imbalanced Approaches
Review on Imbalanced Classification

ABSTRACT

Multi-Label Classification (MLC) is an extension of the standard single-label classification where each data instance is associated with several labels simultaneously. MLC has gained much importance in recent years due to its wide range of application domains. However, the class imbalance problem has become an inherent characteristic of many multi-label datasets, where the samples and their corresponding labels are non-uniformly distributed over the data space. The imbalanced problem in MLC imposes challenges to multi-label data analytics which can be viewed from three perspectives: imbalance within labels, among labels, and label-sets. In this paper, we provide a review of the approaches for handling the imbalance problem in multi-label data by collecting the existing research work. As the first systematic study of approaches addressing an imbalanced problem in MLC, this paper provides a comprehensive survey of the state-of-the-art methods for imbalanced MLC, including the characteristics of imbalanced multi-label datasets, evaluation measures and comparative analysis of the proposed methods. The study also discusses important results reported so far in the literature and highlights some of their strengths and limitations to guide future research.

Contents

* Corresponding author.
E-mail addresses: adanenega.tarekegn@unito.it (A.N. Tarekegn), mario.giacobini@unito.it (M. Giacobini), krzysztof.michalak@ue.wroc.pl (K. Michalak).

https://doi.org/10.1016/j.patcog.2021.107965
0031-3203/

1. Introduction

Classification is one of the most important machine learning topics [1]. The goal of classification is to train a computational model using a set of the labelled samples, obtaining a model that is able to correctly classify new unlabeled samples. Traditional single-label classification is one of the most well-established machine learning paradigms. It provides fast and accurate predictions and is successfully applied in many application domains [2-4]. Binary and multi-class classifications are subcategories of single-label classification that concern learning from a set of samples that are associated with a single label. Unlike traditional classification, multi-label classification (MLC) assigns a set of relevant labels to an instance simultaneously [5,6]. Recently, MLC has gained much importance and attracted research attention with a wide range of applications, including medical diagnosis, music categorization, emotion recognition, and image/video annotation [7,8]. In all these cases, the task is to assign a label set to each unseen instance [9]. For example, in bioinformatics, one gene sequence can be associated with a set of multiple molecular functions [10]. In-text categorization, a new article can cover multiple aspects of an event, thus being assigned to a set of multiple topics [11].

There are two well-known approaches for solving the MLC task: problem transformation and algorithm adaptation. The former transforms the MLC task into one or more single-label classification [12], or label ranking (LR) [13] tasks, while the latter aims to adapt the traditional machine learning algorithms to handle multi-label dataset (MLD) directly [14]. The three most often used problem transformation methods are Binary Relevance (BR) [15], Label Power-set (LP) [16], and Classifier Chains (CC) [17]. BR transforms the multi-label problem into a set of independent binary problems. Then, each binary problem is processed by using a traditional classifier. LP considers each unique set of labels as class identifier, transforming the original MLD into a multi-class dataset. After using it to train a regular classifier, the predicted classes are back-transformed into the subsets of labels. Both BR and LP are the foundation for many multi-label ensemble-based methods. CC resolves the BR limitations by taking into account the label correlation task.

The second approach, algorithm adaptation, focuses on introducing MLC-specific changes in classification algorithms, i.e. the single-label classification is revisited in order to be adapted to an MLD. Several adaptations of traditional classifiers have been proposed in the literature, such as Multi-Label k Nearest Neighbors (MLKNN) [18], multi-class multi-label perceptron (MMP) [19], and Ranking Support Vector Machine (Rank-SVM) [20]. An extensive review of MLC methods is provided in [21].

The main challenge of MLC is the imbalanced nature of the MLD, where the samples and their respective labels are non-uniformly distributed over the data space. The problem transformation and adaptation approaches applied to the MLC task are not effective in handling the imbalance problem in an MLC. An imbalanced dataset, in general, becomes a significant challenge in many real-world applications, such as fraud detection [22], risk management [23] and medical diagnosis [24,25]. For example, in a disease diagnostic problem where the cases of the disease are usually rare as compared to the healthy members of the population, the main interest of the task is to detect people with diseases. Hence, an effective classification model is the one that could provide proper labelling of the rare patterns. The imbalanced class distribution has been extensively studied in the context of single-label classification using the commonly existing approaches, such as resampling methods [26]. However, the existing methods cannot be directly applied as a solution to the imbalanced problem in an MLC due to imbalance between labels and label-sets. The imbalance problem becomes more complex for MLDs with a higher number of labels.

In this paper, a literature survey was performed in order to identify a broad range of approaches for addressing the imbalanced problem in MLC. The contributions of this article are threefold: (1) to the best of the authors' knowledge it is the first survey paper focused on the role of imbalance techniques in an MLC task. It presents the characteristics of an imbalanced MLD, a comprehensive survey of different approaches for imbalanced MLC and a summary of evaluation measures; (2) This article presents a comparative analysis of existing approaches and investigates the pros and cons of each approach; (3) The results presented here provide guidance for choosing appropriate techniques and developing better approaches for handling an imbalanced MLC in further studies in this area.

The rest of the paper is organized as follows. Section 2 presents the research methodology and statistical trends. Section 3 describes the characterization of imbalanced MLDs, including its taxonomy and imbalanced level measures. Section 4 is the main section of this survey which discusses various approaches for addressing the imbalanced problem in MLC. Section 5 contains a short description of datasets and tools. Section 6 describes various metrics for the evaluation of MLC model. Section 7 presents a comparative analysis of solutions with advantages and limitations. Finally, future research directions and conclusions are provided in sections 8 and 9, respectively.

2. Methods and statistical trends

In order to ensure as an objective selection of literature sources as possible, a well-defined search methodology for collecting source articles was adopted in our work. This methodology is presented in detail in this section.

2.1. Data sources and search strategy

For this systematic review, research articles related to imbalanced MLC were searched for, in order to compile the published papers from 2006 up to 2019. First, well-known library databases which covered different research fields were used as a source of information for searching and collecting the literature: DBLP, IEEExplore, Springer, ACM Digital Library, Elsevier, Google Scholar, etc. Boolean operators were used for searching for terms with similar meanings and restricting the research. Predetermined search keywords that included a combination of query phrases, such as 'imbalanced multi-label classification' or 'addressing imbalanced problem' or 'multi-label dataset or 'multi-label prediction', were included. We also attempted to search for articles from other sources (such as peer review journals and conferences).

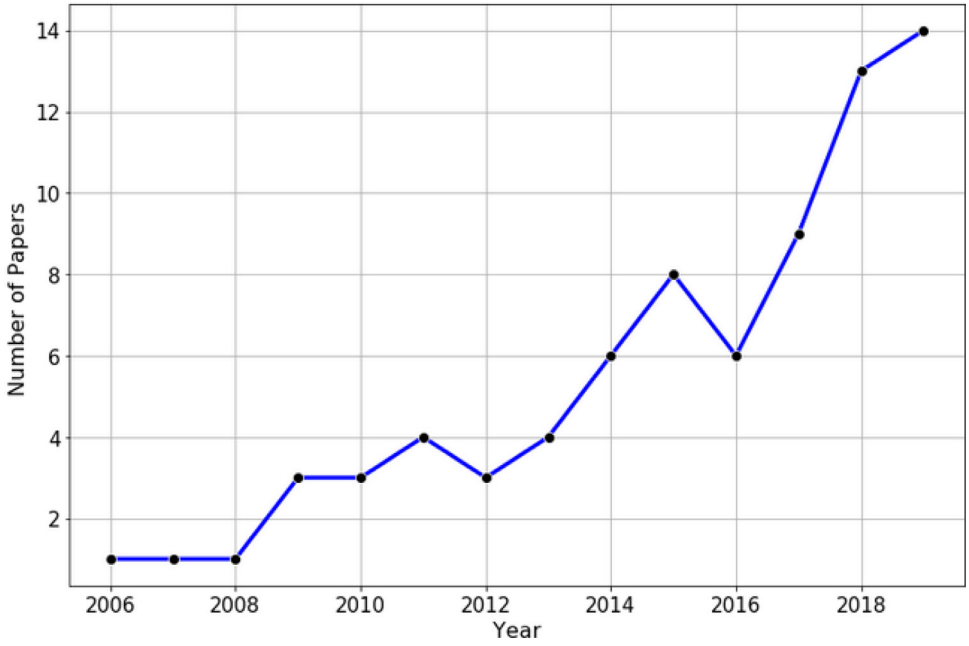

Fig. 1. Publishing trends for Imbalanced Multi-label Classification.

2.2. Selection of studies

The main focus of this paper is to review methods for handling the imbalanced problem in MLC. The following eligibility criteria, which had to be jointly satisfied, were used to select the relevant publications: (1) the study is based on imbalanced MLDs; (2) the work adopts or proposes methods for addressing imbalanced MLC; (3) experimental results evaluate MLC algorithms using multi-label measures; (4) the publication is a full-text article written in English. Articles which provide an MLD-based evaluation of proposed approaches for addressing imbalanced MLC are selected for review without restrictions on the dates of publication. Published works with duplicated title, abstract or content were manually removed, keeping only one copy of the publication. Generally, at the initial stage of searching, 392 publications were collected and identified, from which 86 were duplicates and 219 were discarded on the basis of title and abstract. Finally, by reviewing the full text of each paper, 74 papers were found to be relevant to this study.

2.3. Statistical trends

Fig. 1 presents the publication trends of imbalanced multi-label learning by plotting the number of publications from 2006 to 2019. The number of publications has shown stable growth for the years between 2012 and 2015 and 2016 and 2019 in comparison to the other periods. The number of publications was lowered in the year 2016 compared to 2015 and later showed an increase in the number of publications in the subsequent years. More recently, the number of published works on imbalanced MLC is much higher than the previous years. This suggests that the imbalanced MLC has remained a valuable research topic which has gained wide attention from researchers.

3. Characteristics of imbalanced multi-label datasets

This section discusses imbalance problems in multi-label datasets (MLDs) and characterization measures used for examining the characteristics of such datasets. In this and the following sections, we will use the following notation:

- $M = \{(x_i, Y_i), i = 1, ..., m\}$ – an MLD consisting of $m = \mid M \mid$ multi-label examples,
- $L = \{\lambda_j : j = 1,..., q\}$ – the set of all labels in the given multi-label classification problem,
- q – the number of labels, $q = |L|$
- x_i – the attribute vector of the ith sample in M (with $i = 1, ..., m$),
- $Y_i \subseteq L$ – the *actual* label-set for the ith sample in M (with $i = 1, ..., m$),
- $Z_i \subseteq L$ – the *predicted* label-set for the ith sample in M (with $i = 1, ..., m$),
- $ri(\lambda)$ – the rank predicted by a label ranking (LR) method for the label $\lambda \in L$. The most relevant label receives the highest rank, which is 1, and the least relevant one receives the lowest rank, which is q.

3.1. Imbalance problems in MLD

In any classification task, the presence of imbalanced data [27,28] is a common and challenging problem which affects the learning process of a classification model. In particular, imbalance learning is a well-known and inherent characteristic of many MLDs which affects the learning process of many classification algorithms. The imbalance problem in an MLD can be viewed from three perspectives: imbalance within labels, imbalance between labels, and imbalance among the label-sets. In the case of imbalance within labels, each label usually contains an extremely high number of negative samples and a very small number of positive samples [29-31]. In the imbalance between labels, the frequency of individual labels in the MLD is considered where the number of 1's (positive class) in one label may be higher than the number of 1's in the other label [32,33]. Since every instance of an MLD is associated with several outputs or labels, it is common that some of them are majority ones while others are minority labels, i.e., some labels have many more positive samples than others. The third type of label imbalance that usually occurs in MLD is the sparse frequency of label-sets [34]. If a full label-set is taken into account, the proportion of positive to negative samples for each class may be associated with the most common label-sets. In MLDs, due to the label sparseness, there are usually more frequent label-sets and

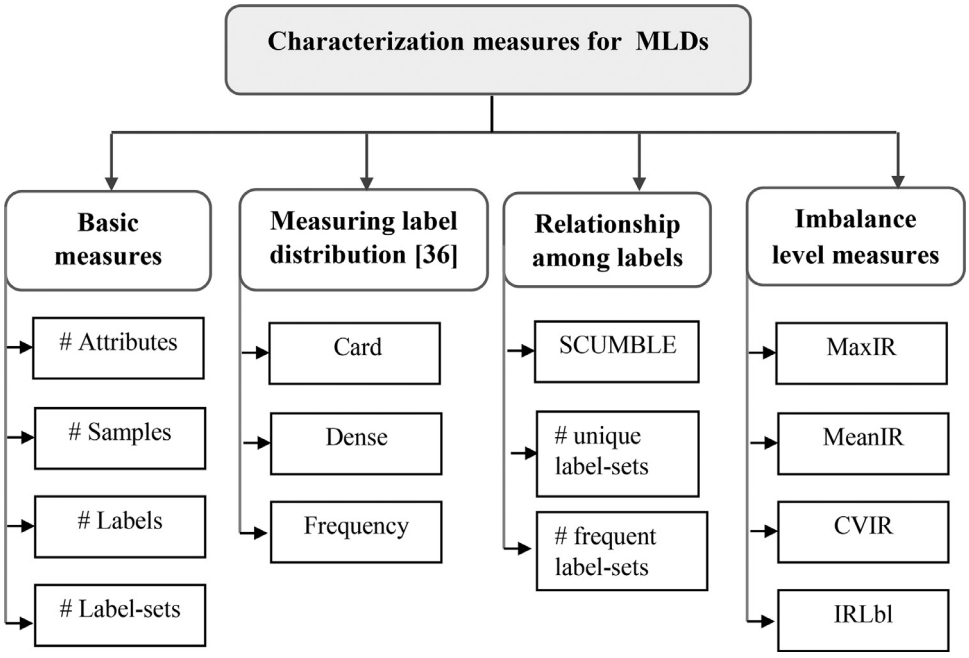

Fig. 2. Taxonomy of characterization measures for MLDs [35,36].

unique label-sets. This also implies that some of the label-sets may be considered majority and the remaining label-sets may be considered minority cases at the same time.

3.2. Characterization measures

Before building a classification model to solve a specific problem, we usually examine the characteristics of the dataset being studied to gain an understanding of the relationships between variables and to determine an appropriate model for it. When faced with an MLD, we must also examine the relationships between the labels, the concurrence level among labels, and the imbalance level of the data. The most basic information which can be obtained from an MLD includes the number of samples, attributes, labels and label-sets. The commonly used characterization measures of an MLD [35] include label distribution measures, imbalance level measures, and concurrence level measures (Fig. 2).

Imbalance level measures: Most MLDs are imbalanced, in which some of the labels are very frequent, while others are quite rare. Therefore, it is important to define the level of imbalance in MLD, considering all the labels. Four different measures are proposed in the literature to assess label imbalance [37]: Imbalance ratio per label (IRLbl), Mean imbalance ratio (MeanIR), Maximum IRbl (MaxIR) and Coefficient of variation of IRLbl (CVIR).

Imbalance ratio per label (IRLbl): Let M be an MLD with a set of labels L and Y_i be the label-set of the ith instance. IRLbl is calculated for the label λ as the ratio between the majority label and the label λ, where IRLbl is 1 for the most frequent label and a greater value for the rest. The larger the value of IRLbl, the higher the imbalance level for the concerned label.

$$IRLbl(\lambda) = \frac{\max_{\lambda' \in L}\left(\sum_{i=1}^{m} h(\lambda', Y_i)\right)}{\sum_{i=1}^{m} h(\lambda, Y_i)}, \quad h(\lambda, Y_i) = \begin{cases} 1 & \lambda \in Y_i \\ 0 & \lambda \notin Y_i \end{cases} \tag{1}$$

Mean imbalance ratio (MeanIR): It is the mean imbalance ratio among all labels in an MLD.

$$MeanIR = \frac{1}{q} \sum_{\lambda \in L} IRLbl(\lambda) \tag{2}$$

Maximum imbalance ratio (MaxIR): The ratio of the most common label against the rarest one.

$$MaxIR = \max_{\lambda \in L} (IRLbl(\lambda)) \tag{3}$$

Coefficient of variation of IRLbl (CVIR): CVIR measures the variation of IRLbl, i.e., the similarity of the level of imbalance between all labels. It indicates if labels experience a similar level of imbalance or, there are large differences among them. The higher the CVIR value, the higher would be this difference:

$$CVIR = \frac{IRLbl\sigma}{MeanIR}, \; IRLbl\sigma = \sqrt{\sum_{\lambda \in L} \frac{(IRLbl(\lambda) - MeanIR)^2}{q-1}} \tag{4}$$

Concurrence level measures: The number of different label-sets, as well as the amount of them being unique label-sets (appearing only once in MLD), give us an indication of how sparsely the labels are distributed. The label-sets by themselves allow knowing how the labels in L are related. SCUMBLE [38] is proposed to assess the concurrence among very frequent and rare labels. A small score will denote an MLD with not much concurrence among imbalanced labels, whereas a large one would evidence the opposite case.

$$SCUMBLE(M) = \frac{1}{m} \sum_{i=1}^{m} \left[1 - \frac{1}{\overline{IRLbl_i}} \left(\prod_{\lambda \in L} IRLbl_{i\lambda} \right)^{(1/q)} \right] \tag{5}$$

where $\overline{IRLbl_i}$ represents the average imbalance level of the labels appearing in the ith sample, and $IRLbl_{i\lambda}$ is equal to $IRLbl(\lambda)$ if $\lambda \in Y_i$ and otherwise $IRLbl_{i\lambda} = 0$.

4. Approaches for imbalanced multi-label classification

The imbalanced approaches proposed for MLC can be divided into four categories: resampling methods, classifier adaptation, ensemble approaches and cost-sensitive methods. Fig. 3 summarizes the categorization of these approaches with descriptions in the next sections.

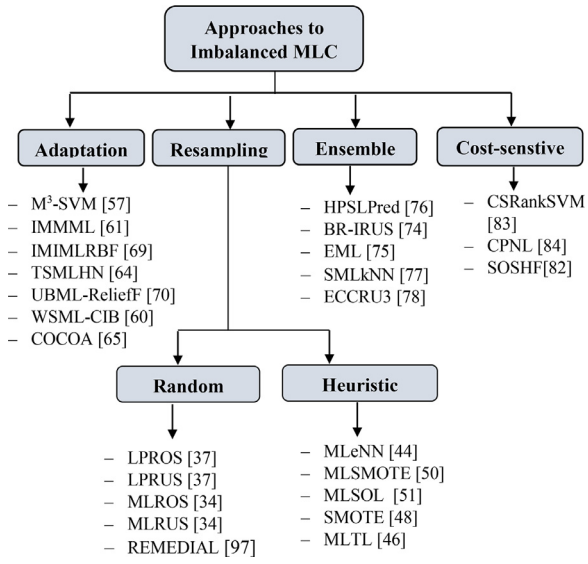

Fig. 3. Categorization of methods proposed in the literature to address imbalanced MLC.

4.1. Resampling methods

The resampling approaches are the most commonly used techniques to handle imbalanced data. These approaches are based on the preprocessing of the MLDs. They aim to produce new, more balanced versions of MLDs, and they belong to the classifier-independent group. Resampling methods are based on undersampling [39], which removes samples associated with the majority label and oversampling [40], which generates new samples associated with the minority label, or both actions at the same time. These methods can also be grouped into two categories: random methods and heuristic methods, according to the way in which the samples are added or removed. The former randomly choose the samples to be deleted or produced associated with a specific label. The latter can be based on disparate heuristics to search for the proper instances, as well as to generate new ones. These two resampling approaches can be adapted to deal with MLDs, as discussed in the following sections.

4.1.1. Multi-label random resampling

The random resampling approach applied to MLC uses different methods than the ones used in single-label classification, since the existing resampling methods cannot be directly used in MLC. Random resampling methods applicable to MLC can be based on the LP transformation, BR methods, imbalance measures, etc. LP-RUS and LP-ROS are two examples of resampling methods proposed in [37] based on LP transformation. The LP transformation method transforms the MLD into a multi-class dataset, processing each different combination of labels (label-set) as a class. LP-RUS randomly removes instances assigned with the most frequent label-set, and the processing stops when the number of samples in the MLD is reduced by an indicated percentage. LP-ROS is a multi-label random oversampling method that works by cloning random samples of minority label-sets until the size of the MLD increases by the prespecified percentage. Although LP-based resampling has its own advantages to solve the imbalance problem, it is limited by the labels sparseness in the MLDs. In other words, there are MLDs with as many distinct label combinations as instances. This implies that all label-sets would be considered to be both majority and minority cases at the same time. Thus LP-ROS and LP-RUS could hardly fix the imbalance problem in such cases. An alternative approach to tackle this limitation would be evaluating the individual imbalance level of each label. ML-RUS and ML-ROS are examples of such approaches based on the frequency of individual labels, instead of the full label-sets, isolating the instances with one or more minority labels [34]. The main aim of ML-RUS is to delete samples with majority labels and of ML-ROS to clone samples with minority labels. These two methods rely on IRLbl and MeanIR measures. Labels whose IRLbl is greater than MeanIR are considered to be minority labels, while labels whose IRLbl is smaller than MeanIR can be considered to be majority labels. Both ML-ROS and ML-RUS resample the dataset, which improves the classification performance. One main limitation with these ML-based methods is that some of the minority samples selected by ML-ROS may contain the most frequent labels, due to the joint appearance of minority and majority labels. Therefore, the oversampling will include both the majority and minority labels. As a remedy to this problem, REMEDIAL is proposed in [33]. REMEDIAL method tackles the imbalanced problem by decoupling the majority and minority labels, of which the imbalance level is assessed by SCUMBLE. REMEDIAL could be either a standalone sampling method or can be combined with other resampling techniques. For example in [41] REMEDIAL was combined with MLSMOTE. Other strategies, such as best first oversampling [42], and imbalance in hierarchical MLDs [43] have been utilized to address the imbalance problems of MLC.

4.1.2. Multi-label heuristic resampling

In this approach, the instances to be deleted or cloned are heuristically selected, instead of being randomly chosen. Random resampling methods can result in a loss of potentially useful information during undersampling, which in turn causes overfitting in the course of oversampling. To overcome these limitations, heuristic approaches can be employed as an alternative for selecting the right samples in the process of undersampling and oversampling. MLeNN is one of the heuristic multi-label undersampling approaches proposed by Charte et al. [44]. It is built upon the Edited Nearest-Neighbor (ENN) rule [45] and depends on the MeanIR and IRLbl measures to assess the imbalance level in MLDs. MLeNN is used to make a careful selection of instances to remove from the majority samples in a heuristic way. The MLeNN method performs better than LP-RUS. MLTL is a similar heuristic-based approach recently proposed in [46]. This method adopts the classic Tomek Link algorithm [47] to address the imbalance, which can be used as an undersampling or cleaning technique.

Generally, heuristic-based undersampling methods, unlike random undersampling, try to eliminate the least significant instances of the majority class and thus minimize the risk of losing important information. However, these methods have also some drawbacks: (1) they do not allow determining the number of removed samples which usually depends on the nature of the data; (2) they are difficult to apply when the minority and majority labels jointly appear in the same instances.

Heuristic-based synthetic instance generation has also been explored to handle imbalanced MLDs. A proposal in [48] is based on the original SMOTE algorithm [49] together with three transformation strategies. The first strategy uses a binary relevance method to transform instances into positive and negative to apply SMOTE. The second approach transforms instances in which the minority label appears in isolation into positive and the remaining into negative. The third strategy considers all samples in which the minority label appears and applies SMOTE several times. In the paper [48], it was observed that the third method improved the results, whereas the other two produced a general degradation of performance. An extension of SMOTE, called MLSMOTE, applied to MLDs, was proposed in [50]. MLSMOTE considers a list of minority labels using the instances in which these labels appear as seeds to generate new instances. First, the nearest neighbors of the seed instances are found, and then the features of the synthetic instances are ob-

tained by an interpolation technique. MLSMOTE takes into account several minority labels to produce synthetic instances instead of only one label, which is an advantage since most MLDs have multiple minority labels.

Another recently proposed approach is MLSOL [51]. This method focuses on analyzing imbalance by looking at the local characteristics of minority samples, rather than the imbalance of the whole dataset. MLSOL first calculates the weight vector for seed instance selection and a type matrix for synthetic instance generation based on the local label distribution. Once the seed instance is selected based on the weight vector, the reference instance is randomly chosen from the k nearest neighbors of the seed instance. An ensemble framework is incorporated into MLSOL to improve its robustness. The use of weighted sampling for seed instance selection and its ensemble version allows MLSOL to create more diverse models and to achieve better performance with greater error correction than MLSMOTE. In MLSMOTE, the labels of the synthetic instance are fixed, while in MLSOL, the labels of the new instance change according to its location, which avoids the introduction of noise. Other works study induction based undersampling [52] and reverse-nearest neighbourhood-based oversampling [53].

Generally, the resampling methods, both random and heuristic, are popular approaches for dealing with imbalanced data. However, since random oversampling usually involves exact copies of samples to increase the size of the data space, it may lead to overfitting [54,55], and also requires more time during the training phase. Oversampling doesn't introduce new data, so it can not address the fundamental 'lack of data' issue. As a result, oversampling may not always be effective at improving the detection of minority samples [55,56].

4.2. Classifier adaptation

Apart from resampling methods, adapting the existing machine learning algorithms is another way of facing the imbalance problem. Adaptation methods could be categorized as dedicated algorithms that directly learn the imbalance distribution from the classes in the datasets. Some multi-label methods adapted to deal with imbalanced MLC have been proposed in the literature. In [57], a min-max modular network with SVM was proposed to address the imbalanced problem of MLDs. It works by decomposing a multi-label imbalanced classification problem into a series of small two-class subproblems. In the learning process, each subproblem can be tackled by one of the standard classification algorithms, and then the outputs of the classifiers are combined by using minimization and maximization principles [58] to generate solutions to the original problem. This method works according to the principles of the Min-Max Modular network and presents different decomposition strategies to improve the performance of these networks.

Another proposal based on adaptation methods is presented in [59]. It uses an enrichment process in neural network training to address the multi-label and imbalanced data problems, such as semantic scene classification, robotic state recognition, and other real-world applications. The enrichment process manages the training data using three steps: the first step is an initialization, which uses a clustering method to group similar instances and gets a balanced representation to initialize the neural network. In the second step, the network is iteratively trained, as usual, while data samples are added and removed from the training set, according to their prevalence. The final step checks if the enrichment process has to be repeated for a predefined number of iterations or it has reached the stop condition. This way, the overall balance of the neural network used as a classifier is improved.

Recently, an adaptation approach was proposed in [60] to address the imbalance in MLC. It is based on an asymmetric stage-wise loss function to adjust the loss cost of positive and negative samples dynamically. In [61], imbalanced multi-modal multi-label learning (IMMML) was proposed. It was designed to tackle the imbalance problem in the subcellular localization prediction of the human proteins with multiple sites. The algorithm is based on a Gaussian process model, combined with latent functions on the feature space and covariance matrices to obtain correlations among labels. The imbalance problem is solved, giving each label a weighting coefficient linked to the likelihood of labels on each sample. IMMML is designed as a specific solution to a definite problem, hardly applicable in a different context.

The proposal in [62], Imbalanced multi-instance multi-label radial basis function neural networks (IMIMLRBF), is an extension of MIMLRBF [63]. IMIMLRBF is a multi-instance and multi-label classification algorithm based on radial basis neural networks. The adaptation works in two ways. First, the number of units in the hidden layer, with MIMLRBF being constant, is computed according to the number of samples of each label. Then, the weights associated with the links between the hidden and output layers are adjusted, biasing them depending on the label frequencies. In [64], an approach based on the multi-label hyper network was proposed to address the imbalance problem in MLC. In this algorithm, labels of an MLD are separated into two groups based on their imbalance ratios. These two groups are common labels and imbalanced labels. The algorithm works in two steps. In the first step, a multi-label hyper network is trained, and it produces preliminary predictions. In the second step, the correlations between imbalanced labels and common labels are used for refining the predictions obtained in the first step, thereby improving the classification performance.

Zhang et al. [65] proposed the class-imbalance aware algorithm named cross-coupling aggregation (COCOA). For each class in the dataset, COCOA combines the predictive results of a binary-class imbalance classifier corresponding to the current label and the predictive results of some multi-class imbalance learners. The final decision for each class label is obtained by aggregating the outputs of binary and multi-class learners. This approach has also been applied for a decision support system in medical diagnosis with imbalanced clinical data [66]. Pouyanfar et al. [67] propose recent work, entitled "multi-label multimodal deep learning framework for imbalanced data classification" to address challenges in multi-media data classification. The proposed framework handles the imbalanced problem in MLC by assigning a specific weight to each class automatically during the classification task. Other models based on a neural network were proposed in [68,69]. Apart from the above-mentioned methods, the RelieF feature selection algorithm [70], concept drift and KNN based approach [71] have been employed to address the imbalance in MLC.

4.3. Ensemble methods

Ensemble methods combine several base models in order to produce one optimal predictive model [72]. The use of sets of classifiers as ensembles has proven to be effective in single-label classification. A similar approach has been used in MLC for improving predictive performance and solving the imbalanced problem. The ensemble of multi-label classifiers trains several multi-label classifiers. Thus, all the trained classifiers are different and can provide diverse multi-label predictions. There are several ways of joining the outputs of these classifiers [73].

An inverse random undersampling (BR-IRUS) method is proposed in [74]. BR-IRUS is implemented on an ensemble of binary classifiers which are trained for individual labels using a subset of the original data. The subset of the instances contains all samples in which the minority label is present, along with a small por-

tion of the remaining samples. This way, each individual classifier solves a prediction task. Joining the predictions given by the classifiers associated with a label, a more defined boundary around the minority label space is generated.

In [75], a heterogeneous ensemble of multi-label learners is proposed by combining state-of-the-art multi-label methods. This method simultaneously tackles both the sample imbalance and label correlation problems. The ensemble is composed of five classifiers. All of them are trained using different algorithms on the same data. Several methods for joining the individual predictions are tested, along with different thresholding and weighting schemes with adjustments made through cross-validation.

The authors in [76] proposed an ensemble classifier called HPSLpred with an imbalanced source of human protein subcellular location prediction. HPSLpred integrates 12 kinds of basic classifiers to address the imbalanced problem. The authors in [77] used a two-stage stack-like ensemble of MLkNN classifier to exploit label associations in MLC. The algorithm shows an improvement in comparison to MLkNN without stacking. ECCRU3 [78] extends the ECC resilient to class imbalance by coupling undersampling and improving the exploitation of majority samples. Furthermore, other ensemble classification algorithms have been employed in MLC, such as ensemble of multi-label classifiers [79], bagging and adaptive boosting [80].

4.4. Cost-sensitive approaches

Cost-sensitive methods use different cost metrics to describe the costs of any particular misclassified sample, aiming to minimize the total cost. Most commonly, these methods are applied to imbalanced learning by associating high misclassifying cost to the minority classes [81]. In traditional classification, the objective is to minimize the misclassification rate, and thus most classifiers assume that the misclassification costs are equal. A more general setting is the cost-sensitive classification where the costs caused by different kinds of errors are not assumed to be equal. Cost-sensitive approaches can be incorporated both at the data and algorithmic level, by considering higher costs for the misclassification of minority samples with respect to majority samples. In contrast to traditional classification, there are very few studies on cost-sensitive learning studies in MLC. The reason for this can be due to the fact that cost-sensitive learning strategies are difficult with regard to the assignment of an effective cost matrix [26].

Some cost-sensitive approaches have been migrated to the multi-label scenario to explore a class-imbalance problem, among them: SOSHF [82] transforms the multi-label learning task to an imbalanced single label classification type via cost-sensitive clustering, and, subsequently, the oblique structured Hellinger decision trees address the new task. In [83], cost-sensitive ranking support vector machine for MLD is attempted, which assigns a different misclassification cost for each label-set to effectively tackle the problem of imbalance in MLC. Another cost-sensitive multi-label learning is proposed in [84]. This work extends BR to consider the exploitation of the label correlations and exploration of the class-imbalance simultaneously. A cost-sensitive loss is utilized to tackle the class-imbalance problem.

5. Datasets and software tools

To evaluate the proposed methods of imbalanced MLC, most authors used publicly available benchmark MLDs with different formats (text, audio, images, etc.). Twenty-six MLDs in ARFF file format, along with their descriptions and statistics, can be found in the online MULAN repository [85]. The MULAN repository is the most used resource by many authors of articles that concern the MLC task. Other MLD repositories include the MEKA repository [86] and the R ultimate MLD repository [87]. Software tools associated with each repository are made available in order to help analyze MLDs and perform MLC. These include MULAN [88], MEKA [86], mldr package in R [35] and multilearn library in Python [89].

6. Model evaluation and performance metrics

Various metrics have been proposed in the literature to evaluate the classification performance of MLC models [90]. Unlike the traditional classification, which produces a single class as output being either a correct or wrong prediction, the output of any multi-label classifier consists of a label-set predicted for each instance. The evaluation of MLC requires different measures with respect to the ground truth of multi-label prediction results. The measures can be broadly categorized into three groups: example-based [91], label-based [92], and ranking-based measures [93]. Example-based measures are computed individually for each sample, then averaged to obtain the final value. Label-based measures are computed for each label, instead of per instance. The ranking-based metrics evaluate the ranking of labels with respect to the original MLDs.

6.1. Example-based measures

Hamming loss (HL) [90] is the most common performance measure in MLC, computed as the symmetric difference between the predicted and true labels and divided by the total number of labels in the MLD. The smaller the value of the Hamming Loss, the better the performance:

$$\mathrm{HL} = \frac{1}{m}\sum_{i=1}^{m}\frac{|Y_i \triangle Z_i|}{q}, \tag{6}$$

where Δ denotes the symmetric difference of the two sets and corresponds to the XOR operation in Boolean logic. HL measures the fraction of labels that are misclassified.

Subset Accuracy (SA) [15] evaluates the percentage of correctly predicted labels among all predicted and true labels. This is a very strict evaluation measure as it requires the predicted set of labels to be an exact match of the true set of labels.

$$\mathrm{SA} = \frac{1}{m}\sum_{i=1}^{m} I(Z_i = Y_i), \tag{7}$$

where I(true) = 1 and I(false) = 0.

Accuracy [91] is the ratio of predicted correct labels with respect to the total number (predicted and actual) of labels for each instance.

$$\mathrm{Accuracy} = \frac{1}{m}\sum_{i=1}^{m}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \tag{8}$$

Precision [94] computed as indicated in Eq. (9) is the proportion of predicted correct outputs to the total number of predicted outputs, averaged over all instances.

$$\mathrm{Precision} = \frac{1}{m}\sum_{i=1}^{m}\frac{|Y_i \cap Z_i|}{|Z_i|} \tag{9}$$

Recall measures the proportion of predicted correct labels to the total number of true labels, averaged over all instances.

$$\mathrm{Recall} = \frac{1}{m}\sum_{i=1}^{m}\frac{|Y_i \cap Z_i|}{|Y_i|} \tag{10}$$

F-measure represents the harmonic mean of Recall and Precision, providing a balanced assessment between precision and recall. It is a weighted measure of how many relevant labels are predicted and how many of the predicted labels are relevant.

$$\mathrm{F-measure} = 2*\frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{11}$$

As in single-label multi-class classification, the higher the value of accuracy, precision, recall and F-measure, the better the performance of the learning algorithm.

6.2. Label-based measures

Label based measures evaluate each label separately and then average over all labels. Therefore, any known measure, used for evaluation of a binary classifier (e.g. accuracy, precision, recall, F-measure, ROC, etc.), can be used here. Label-based measures are calculated for all labels by using two averaging operations, called macro averaging where any of the measures can be computed on individual class labels first and then averaged over all classes and micro-averaging, the measures can be computed globally over all instances and all class labels [15]. Let EM denote one of the evaluation metrics, $FP\lambda$ the number of False Positives, $TP\lambda$ the number of True Positives, $FN\lambda$ the number of False Negatives, and $TN\lambda$ the number of True Negatives. Then, the macro and micro averaged measures can be calculated as follows [95]:

$$EM_{macro} = \frac{1}{q}\sum_{\lambda=1}^{q} EM(TP\lambda,\ FP\lambda,\ TN\lambda,\ FN\lambda) \tag{12}$$

$$EM_{micro} = \mathrm{EM}\left(\sum_{\lambda=1}^{q} TP\lambda,\ \sum_{\lambda=1}^{q} FP\lambda,\ \sum_{\lambda=1}^{q} TN\lambda,\ \sum_{\lambda=1}^{q} FN\lambda\right) \tag{13}$$

6.3. Ranking –based measures

One Error measures how many times the best-ranked label given by the model is not part of the true label-set of the sample [96]. The smaller the value of one error, the better the performance:

$$\mathrm{OneError} = \frac{1}{m}\sum_{i=1}^{m}\delta(\operatorname{argmin} ri(\lambda)),\ \delta(\lambda) = \begin{cases} 1 & \text{if } \lambda \notin Y_i \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

Coverage is the metric that evaluates how far on average, a learning algorithm needs to go down in the ordered list of predicted labels to cover all the true labels of an instance. Clearly, the smaller the value of coverage, the better the performance.

$$\mathrm{Coverage} = \frac{1}{m}\sum\nolimits_{i=1}^{m} \max(\mathrm{ri}(\lambda)) - 1, \lambda \in Y_i \tag{15}$$

Ranking loss (RL) measures how many times a relevant label appears ranked lower than a non-relevant label. The smaller the value of RL the better the performance:

$$\mathrm{RL} = \frac{1}{\mathrm{m}}\ \sum\nolimits_{\mathrm{i}=1}^{\mathrm{m}} \frac{1}{|\ \mathrm{Y_i}\ |\ |\ \bar{\mathrm{Y}}_\mathrm{i}|}|\left\{(\lambda_a, \lambda_b) : ri(\lambda_a) > ri(\lambda_b),\ \ (\lambda_a, \lambda_b) \in \mathrm{Y_i} x \bar{\mathrm{Y}}_\mathrm{i}\right\}|, \tag{16}$$

where $\bar{\mathrm{Y}}_\mathrm{i}$ is the complementary set of $\mathrm{Y_i}$ with respect to L.

Average precision (AvgPrec) evaluates the average fraction of labels ranked above a particular label $\lambda \epsilon \mathrm{Y_i}$ which actually are in $\mathrm{Y_i}$. The higher the average precision, the better the performance.

$$\mathrm{AvgPrec} = \frac{1}{\mathrm{m}}\ \sum\nolimits_{\mathrm{i}=1}^{\mathrm{m}} \frac{1}{|\ \mathrm{Y_i}\ |} \sum_{\lambda \in Y_i} \frac{\left|\left\{\lambda' \in Y_i : ri(\lambda') \le ri(\lambda)\right\}\right|}{ri(\lambda)} \tag{17}$$

7. Comparative analysis

In this section, we present the comparative analysis of the different methods proposed in the literature for addressing an imbalance problem in MLC. Table 1 depicts the advantages and disadvantages of classifier adaptation, resampling, ensemble and cost-sensitive methods. These methods are effective alternatives for imbalanced MLC task. However, there exist various constraints with respect to addressing the imbalance. Adaptation method makes the model insensitive to the imbalanced sample distribution by modifying the base classifier. This method is inefficient when the label space is too large and requires extensive knowledge about the base classifier and the problem domain. Resampling methods proposed for imbalanced MLC are advantageous as they are classifier independent and do not require any specific multi-label classifier to preprocess MLDs. Thus, a preprocessed MLD can be used as input to any of the MLC algorithms.

However, the large differences in imbalance levels between labels and the high level of concurrence among imbalanced labels would greatly influence the behavior of resampling approaches and, as a result, only certain MLDs with the lowest concurrence level can be most benefitted from the resampling. Ensemble approaches have a problem of computational complexity since more than one classifier has to be trained and combined to obtain a final prediction result. In Table 2, a detailed comparison of various methods from the different approaches using various parameters is presented. The comparison criteria include the MLC approaches used, balancing method, advantages and limitations of each proposed approach. The authors in [46] used MLC algorithms to compare the state of the art multi-label resampling approaches using different imbalanced MLDs with varying level of imbalance. Table 3 presents the experimental result of resampling approaches on six MLDs using micro F-score as an evaluation metric and RAkEL as MLC classifier. The experimental results in Table 3 indicate that LPROS achieved better results on three datasets (emotions, scene and yeast), while MLROS and MLTL have shown better results on CAL500 and Medical datasets, respectively.

8. Future research directions

Existing works have proposed various techniques to tackle the imbalance issues in MLC. However, several challenges remain, and imbalanced classification of MLD still requires significant development. The following are some of the possible future research directions to deal with imbalanced MLC.

1. The success of currently available multi-label resampling algorithms is highly influenced by certain characteristics of the tackled problem. (1) Joint occurrence of minority and majority labels in the same instance. The potential existence of samples associated with rare and frequent labels in an MLD could make the resampling strategies ineffective. A recent study in [41] has attempted resampling by decoupling imbalanced labels, but it has limitations for some resampling methods and MLC algorithms due to that the decoupled labels are located in the same instance. More sophisticated approaches are needed, considering label-set based relocation and defining thresholds for decoupling that can be able to work with any of the available MLC methods; (2) MLDs with a large number of imbalanced labels pose a scalability challenge. Some methods have been proposed by Wang [71] to solve this problem. This approach could reduce Hamming Loss but did not completely eliminate it. Some approaches, such as embedding [101], may help to address such challenges.
2. In MLC, there is a need for imbalance–aware classifiers that do not require resampling strategies. It seems promising to use the existing MLC methods (such as hierarchical MLC or classifier chains) and combine them with the imbalance–aware solutions that are available in the multi-class classification domain. An ideal goal would be the development of such multi-label classifiers that display similar performance to canonical methods on balanced multi-label problems while being at the same time robust to the presence of imbalance.
3. Ensemble methods are well-known to tackle both imbalance and label correlation problems. These methods work by em-

Table 1
Advantage and disadvantage of different categories of imbalance-handling methods.

Approach	Advantages	Constraints/Disadvantages
Resampling methods	- Applicable to any MLC classifier - Classifier independent	- High level of concurrence between imbalanced labels and a large number of unique label-sets - May introduce noisy data
Classifier adaptation	- Effective in a certain context - MLD does not change	- Requires extensive knowledge of the specific classifier and problem domains. - Algorithm dependent solutions
Ensemble methods	- Decrease variance and improve prediction - Reduce overfitting	- Computational complexity - No clear criteria for selecting the type and number of MLC classifiers
Cost-sensitive	- Computationally efficient	- Real cost values are unknown in most applications domains

Table 2
Comparison of specific methods proposed for addressing imbalanced MLC.

Article	ApproachesUsed	Balancing Method	Advantages	Limitations
(Charte et al. 2013) [37]	LP-based Transformation [16]	Random resampling	Helps to reduce imbalance among the label-sets	Label sparseness in the MLD, hardly fix the imbalance problem
(Giraldo-Forero et al. 2013) [48]	Binary relevance [90]	Heuristic over sampling	Easy to apply SMOTE for a class-imbalance problem	It doesn't consider imbalance between labels/label-sets
(K. Chen et al.,2006) [57]	Decomposition strategy [98]	Classifier adaptation	Subproblems can be balanced	One label may happen more frequently than other
(Charte et al. 2015) [34]	imbalance measure of individual labels	Random resampling	Reduces highly imbalanced labels	The joint appearance of the majority and minority labels affect one another
(M. Tahir et al.,2012) [74]	Binary relevance (BR) method	Ensemble approach	Reduces class-imbalance problem	Doesn't consider imbalance between labels/label-sets
(Pereira et al,2019) [46]	LP (label-set) based	Heuristic over undersampling	Defined threshold for Hamming distance to remove majority label	Difficult to apply to highly concurrent imbalanced labels
(F. Luo et al.,2019) [60]	asymmetric stage-wise functions	Classifier adaptation	Accuracy on minority samples can be improved	More applicable to only missing (unlabelled) labels [99]
(M. A. Tahir et al.,2012) [75]	LP, BR and CLR based transformation	Ensemble approach	Tackles both class imbalance and label correlation	Computationally intensive and base classifiers may be problem-specific
(P. Cao et al.,2017) [83]	problem transformation and algorithm adaptation	Cost sensitive learning	Reduces imbalance in the label-set space without changing the original data	It doesn't consider the imbalance problem between individual labels
(Charte et al.,2019) [41]	BR and LP transformation	Heuristic and random resampling	Solves the limitation of multi-label oversampling	Limited to high label concurrence problem under certain conditions
(Zhang et al.,2015) [65]	Binary and multi-class learner	Classifier adaptation	Considers label correlation while solving imbalance	Limited to an imbalance within labels, rather than between labels/label-sets
(Ding et al,2018) [100]	BR and classifier chains	Cost-sensitive	Use of penalty function to balancing	Does not consider imbalance among labels
(Charte et al, 2014) [44]	LP (label-set) based	Heuristic resampling	Makes right selection of samples to remove	Difficult to apply to highly concurrent imbalanced labels
(F. Charte et al., 2019) [41]	Label decoupling	Hybrid approach	Improves concurrent imbalanced MLC	Limited to certain circumstances, not a general solution

Table 3
Comparison of resampling methods for MLC with imbalanced datasets [46].

Approaches	Datasets					
	CAL500	Emotions	Enron	Medical	Scene	Yeast
None	0.3354	0.621	0.5496	0.8132	0.6237	0.5812
LPROS [37]	0.4924	**0.6814**	0.6306	0.8761	**0.7617**	**0.6721**
LPRUS [37]	0.3751	0.5838	0.5158	0.7853	0.6339	0.5823
MLROS [34]	**0.5413**	0.6395	0.6694	0.8354	0.6500	0.6671
MLRUS [34]	0.3255	0.5846	0.5259	0.8345	0.6919	0.5677
REMEDIAL [97]	0.2951	0.325	0.1135	0.637	0.5648	0.456
REMEDIAL-HwR-ROS [41]	0.2503	0.5111	0.2822	0.5064	0.6361	0.3929
REMEDIAL-HwR-HUS [41]	0.1293	0.349	**0.6841**	0.7841	0.7176	0.3849
REMEDIAL-HwR-SMT [41]	0.1542	0.3056	0.1851	0.4114	0.3888	0.3772
MLeNN [44]	0.3466	0.621	0.6489	0.8774	0.6415	0.5846
MLSMOTE [50]	0.3839	0.4265	0.6125	0.8546	0.4532	0.5794
MLTL [46]	0.3720	0.6409	0.6499	**0.8798**	0.7502	0.6348

ploying the nontrainable average combining rule. However, since MLC algorithms are computationally intensive and MLDs are highly imbalanced, it opens new research challenges on how to use other combination techniques efficiently such as trainable combiners (fuzzy integral) [102] or class indifferent combiners (decision templates and Dempster-Shafer combination) [103]. Another issue that needs further investigation is how to select the base classifiers in MLC since different combinations of base classifiers may perform differently for the specific problem domain. Moreover, there are no clear indicators of how large should the constructed ensemble be when applied to MLC tasks.

4. Another interesting strategy can be the use of hybrid methods, concentrating on combining previously mentioned approaches

to take advantage of their strong points and reduce their weaknesses. It is recommended to combine one of the resampling methods with another one or with adaptation methods investigating the potential to improve the results in these cases. Label correlations can also be combined with resampling techniques in order to improve the learning performance of extremely imbalanced labels [64].

5. Another interesting direction is to investigate the possibilities of using cost-sensitive learning solutions. RAkEL [104] is the most popular method, transforming a multi-label problem with a large number of label-sets into smaller subsets. Hence, it seems straightforward to balance label-set distribution by automatically generating a misclassification cost vector in accordance with the label-set distribution.

9. Conclusions

This paper, to the best of our knowledge, presents the first survey of approaches to the class imbalance problem in multi-label data classification (MLC) which includes the characteristics of the data, problem descriptions, solutions and limitations of the approaches for solving imbalanced problems. In this study, numerous articles related to imbalanced MLC published between 2006 and 2019 were collected and reviewed. Various methods and techniques that have been proposed to overcome the difficulties found in imbalanced MLC can be grouped into four categories: resampling methods, classifier adaptations, ensemble methods and cost-sensitive learning approaches. These approaches have their own limitations, even though some of them have shown good performance in handling imbalanced classes, labels and label-sets in MLC. For example, methods which are proposed for handling imbalance problem between labels cannot be applied to handle the imbalance problem among label-sets. We also found out that research in imbalanced MLC is very limited and that the majority of the existing works addressing the imbalance problem focus on single-label classification. Despite a growing demand for multi-label classification in different domains, developing a comprehensive framework for handling an imbalanced problem in an MLD is still understudied. As a result, this paper concludes with a discussion on the challenges of imbalanced MLC and some future research directions that are worthy of further study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.C. Aggarwal, Data Classification, Chapman and Hall/CRC, 2014. https://doi.org/10.1201/b17320.

[2] M.M.R. Khan, R.B. Arif, A.B. Siddique, M.R. Oishe, Study and observation of the variation of accuracies of KNN, SVM, LMNN, ENN algorithms on eleven different datasets from UCI machine learning repository, 4th International Conference on ICEEiCT, 2018, doi:10.1109/CEEICT.2018.8628041.

[3] A.L. Wang, B.X. Chen, C.G. Wang, D.D. Hua, Non-intrusive load monitoring algorithm based on features of V–I trajectory, Electr. Power Syst. Res. (2018), doi:10.1016/j.epsr.2017.12.012.

[4] A. Tarekegn, F. Ricceri, G. Costa, E. Ferracin, M. Giacobini, Detection of frailty using genetic programming, in: 2020: pp. 228–243. https://doi.org/10.1007/978-3-030-44094-7_15.

[5] G. Tsoumakas, I. Katakis, Multi-label classification, Int. J. Data Warehous. Min. 3 (2007) 1–13, doi:10.4018/jdwm.2007070101.

[6] Z. Ahmadi, S. Kramer, A label compression method for online multi-label classification, Pattern Recognit. Lett. 111 (2018) 64–71, doi:10.1016/j.patrec.2018.04.015.

[7] Y. Zhang, Y. Wang, X.Y. Liu, S. Mi, M.L. Zhang, Large-scale multi-label classification using unknown streaming images: Large-scale multi-label classification using unknown streaming images, Pattern Recognit. (2020), doi:10.1016/j.patcog.2019.107100.

[8] T.T. Nguyen, M.T. Dang, A.V. Luong, A.W.C. Liew, T. Liang, J. McCall, Multi-label classification via incremental clustering on an evolving data stream, Pattern Recognit. (2019), doi:10.1016/j.patcog.2019.06.001.

[9] M.L. Zhang, Z.H. Zhou, ML-KNN: a lazy learning approach to multi-label learning, Pattern Recognit. (2007), doi:10.1016/j.patcog.2006.12.019.

[10] G. Yu, C. Domeniconi, H. Rangwala, G. Zhang, Z. Yu, Transductive multi-label ensemble classification for protein function prediction, in: Proceedings of the 18th ACM SIGKDD International Conference on KDD '12, New York, New York, USA, ACM Press, 2012, p. 1077, doi:10.1145/2339530.2339700.

[11] S.C. Dharmadhikari, A novel multi label text classification model using semi supervised learning, Int. J. Data Min. Knowl. Manag. Process 2 (2012) 11–20, doi:10.5121/ijdkp.2012.2402.

[12] G. Tsoumakas, I. Vlahavas, Random k-labelsets: an ensemble method for multilabel classification, in: Machine Learning: ECML 2007, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 406–417, doi:10.1007/978-3-540-74958-5_38.

[13] J. Fürnkranz, E. Hüllermeier, E. Loza Mencía, K. Brinker, Multilabel classification via calibrated label ranking, Mach. Learn. 73 (2008) 133–153, doi:10.1007/s10994-008-5064-8.

[14] M.L. Zhang, Z.H. Zhou, A review on multi-label learning algorithms, IEEE Trans. Knowl. Data Eng. (2014), doi:10.1109/TKDE.2013.39.

[15] G. Tsoumakas, I. Katakis, I. Vlahavas, Mining multi-label data, in: Data Mining and Knowledge Discovery Handbook, Springer US, Boston, MA, 2009, pp. 667–685, doi:10.1007/978-0-387-09823-4_34.

[16] M.R. Boutell, J. Luo, X. Shen, Learning multi-label scene classification, Pattern Recognit. 37 (2004) 1757–1771, doi:10.1016/j.patcog.2004.03.009.

[17] J. Read, B. Pfahringer, G. Holmes, Classifier chains for multi-label classification, Mach. Learn. 85 (2011) 333–359, doi:10.1007/s10994-011-5256-5.

[18] Min-Ling Zhang, Zhi-Hua Zhou, A k-nearest neighbor based algorithm for multi-label classification, in: 2005. https://doi.org/10.1109/grc.2005.1547385.

[19] E.L. Mencía, J. Fürnkranz, Pairwise learning of multilabel classifications with perceptrons, in: Proceedings of the International Joint Conference on Neural Networks, 2008, doi:10.1109/IJCNN.2008.4634206.

[20] A. Elisseeff, J. Weston, A kernel method for multi-labelled classification, Advances in Neural Information Processing Systems, 14, The MIT Press, 2002, doi:10.7551/mitpress/1120.003.0092.

[21] G. Tsoumakas, I. Katakis, I. Vlahavas, A review of multi-label classification methods, in: Proceedings of the 2nd ADBIS Workshop on Data Mining and Knowledge Discovery (ADMKD 2006), 2006.

[22] C.A. Catania, F. Bromberg, C.G. Garino, An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection, Expert Systems with Applications, 2012, doi:10.1016/j.eswa.2011.08.068.

[23] Y.M. Huang, C.M. Hung, H.C. Jiau, Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem, Nonlinear Anal.: Real World Appl. (2006), doi:10.1016/j.nonrwa.2005.04.006.

[24] A. Tarekegn, F. Ricceri, G. Costa, E. Ferracin, M. Giacobini, Predictive modeling for frailty conditions in elderly people: machine learning approaches, JMIR Med. Inform. (2020), doi:10.2196/16678.

[25] A. Jain, S. Ratnoo, D. Kumar, Addressing class imbalance problem in medical diagnosis: a genetic algorithm approach, in: 2017 International Conference on ICICIC, IEEE, 2017, pp. 1–8, doi:10.1109/ICOMICON.2017.8279150.

[26] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, G. Bing, Learning from class-imbalanced data: review of methods and applications, Expert Syst. App. 73 (2017) 220–239, doi:10.1016/j.eswa.2016.12.035.

[27] N.V. Chawla, N. Japkowicz, A. Kotcz, Special issue on learning from imbalanced data sets, ACM SIGKDD Expl. Newslett. (2004), doi:10.1145/1007730.1007733.

[28] Haibo He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (2009) 1263–1284, doi:10.1109/TKDE.2008.239.

[29] Y. Sun, A.K.C. Wong, M.S. Kamel, Classification of imbalanced data: a review, Int. J. Pattern Recognit. Artif. Intell. (2009), doi:10.1142/S0218001409007326.

[30] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, Y. Zhou, A novel ensemble method for classifying imbalanced data, Pattern Recognit. (2015), doi:10.1016/j.patcog.2014.11.014.

[31] W.W.Y. Ng, G. Zeng, J. Zhang, D.S. Yeung, W. Pedrycz, Dual autoencoders features for imbalance classification problem, Pattern Recognit. (2016), doi:10.1016/j.patcog.2016.06.013.

[32] M. Fang, Y. Xiao, C. Wang, J. Xie, Multi-label classification: dealing with imbalance by combining labels, in: Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2014, doi:10.1109/ICTAI.2014.42.

[33] F. Charte, A.J. Rivera, M.J. del Jesus, F. Herrera, Dealing with difficult minority labels in imbalanced mutilabel data sets, Neurocomputing (2019), doi:10.1016/j.neucom.2016.08.158.

[34] F. Charte, A.J. Rivera, M.J. del Jesus, F. Herrera, Addressing imbalance in multilabel classification: Measures and random resampling algorithms, Neurocomputing (2015), doi:10.1016/j.neucom.2014.08.091.

[35] F. Charte, D. Charte, Working with multilabel datasets in R: the mldr package, R Journal (2015), doi:10.32614/rj-2015-027.

[36] F.C. Bernardini, R.B. da Silva, R.M. Rodovalho, E.B.M. Meza, Cardinality and density measures and their influence to multi-label learning methods, Learn. Nonlinear Models 12 (2014) 53–71, doi:10.21528/LNLM-vol12-no1-art4.

[37] F. Charte, A. Rivera, M.J. del Jesus, F. Herrera, A First Approach to Deal with Imbalance in Multi-label Datasets, International Conference on Hybrid Artificial Intelligence Systems, Springer, Berlin, Heidelberg 8073 (2013) Lecture Notes in Computer Science, doi:10.1007/978-3-642-40846-5_16.

[38] F. Charte, A. Rivera, M.J. del Jesus, F. Herrera, Concurrence among Imbalanced Labels and Its Influence on Multilabel Resampling Algorithms, International Conference on Hybrid Artificial Intelligence Systems, Springer, Cham 8480 (2014) Lecture Notes in Computer Science, doi:10.1007/978-3-319-07617-1_10.

[39] X.Y. Liu, J. Wu, Z.H. Zhou, Exploratory undersampling for class-imbalance learning, IEEE Trans. Syst., Man, Cybern. (2009) Part B: Cybernetics., doi:10.1109/TSMCB.2008.2007853.

[40] F.J. Castellanos, J.J. Valero-Mas, J. Calvo-Zaragoza, J.R. Rico-Juan, Oversampling imbalanced data in the string space, Pattern Recognit. Lett. (2018), doi:10.1016/j.patrec.2018.01.003.

[41] F. Charte, A.J. Rivera, M.J. del Jesus, F. Herrera, Tackling multilabel imbalance through label decoupling and data resampling hybridization, Neurocomputing 326–327 (2019) 110–122, doi:10.1016/j.neucom.2017.01.118.

[42] X. Ai, J. Wu, V.S. Sheng, Y. Yao, P. Zhao, Z. Cui, Best first over-sampling for multilabel classification, in: Proceedings of the 24th ACM International on CIKM '15, New York, New York, USA, ACM Press, 2015, pp. 1803–1806, doi:10.1145/2806416.2806634.

[43] R.Miranda Pereira, Y. Maldonado, E. Gomes Da Costa, C.N. Silla, Dealing with imbalanceness in hierarchical multi-label datasets using multi-label resampling techniques, in: Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2018, doi:10.1109/ICTAI.2018.00128.

[44] F. Charte, A.J. Rivera, M.J. del Jesus, F. Herrera, MLeNN: A First Approach to Heuristic Multilabel Undersampling, 2014, pp. 1–9. Lecture Notes in Computer Science, doi:10.1007/978-3-319-10840-7_1.

[45] D.L. Wilson, Asymptotic properties of nearest neighbor rules using edited data, IEEE Trans. Syst., Man, Cybern. SMC-2 (1972) 408–421, doi:10.1109/TSMC.1972.4309137.

[46] R.M. Pereira, Y.M.G. Costa, C.N. Silla Jr., MLTL: A multi-label approach for the Tomek Link undersampling algorithm, Neurocomputing 383 (2020) 95–105, doi:10.1016/j.neucom.2019.11.076.

[47] I. Tomek, Two modifications of CNN, IEEE Trans. Syst., Man, Cybern. SMC-6 (1976) 769–772, doi:10.1109/TSMC.1976.4309452.

[48] A.F. Giraldo-Forero, J.A. Jaramillo-Garzón, J.F. Ruiz-Muñoz, C.G. Castellanos-Domínguez, Managing Imbalanced Data Sets in Multi-label Problems: A Case Study with the SMOTE Algorithm, 2013, pp. 334–342. Lecture Notes in Computer Science, doi:10.1007/978-3-642-41822-8_42.

[49] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, doi:10.1613/jair.953.

[50] F. Charte, A.J. Rivera, M.J. del Jesus, F. Herrera, MLSMOTE: approaching imbalanced multilabel learning through synthetic instance generation, Knowl.-Based Syst. 89 (2015) 385–397, doi:10.1016/j.knosys.2015.07.019.

[51] B. Liu, G. Tsoumakas, Synthetic oversampling of multi-label data based on local label distribution, (2019). http://arxiv.org/abs/1905.00609.

[52] S. Dendamrongvit, P. Vateekul, M. Kubat, Irrelevant attributes and imbalanced classes in multi-label text-categorization domains, Intell. Data Anal. 15 (2011) 843–859, doi:10.3233/IDA-2011-0499.

[53] P. Sadhukhan, S. Palit, Reverse-nearest neighborhood based oversampling for imbalanced, multi-label datasets, Pattern Recognit. Lett. 125 (2019) 813–820, doi:10.1016/j.patrec.2019.08.009.

[54] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches, IEEE Trans. Syst., Man, Cybern., Part C 42 (2012) 463–484, doi:10.1109/TSMCC.2011.2161285.

[55] J. Burez, D. Van den Poel, Handling class imbalance in customer churn prediction, Expert Syst. Appl. 36 (2009) 4626–4636, doi:10.1016/j.eswa.2008.05.027.

[56] C. Drummond, R.C. Holte, C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling, Workshop on Learning from Imbalanced Datasets II, 2003 https://doi.org/10.1.1.68.6858.

[57] K. Chen, B.L. Lu, J.T. Kwok, Efficient classification of multi-label and imbalanced data using min-max modular classifiers, in: IEEE International Conference on Neural Networks - Conference Proceedings, 2006, doi:10.1109/ijcnn.2006.246893.

[58] B.L. Lu, M. Ito, Task decomposition and module combination based on class relations: a modular neural network for pattern classification, IEEE Trans. Neural Netw. (1999), doi:10.1109/72.788664.

[59] G. Tepvorachai, C. Papachristou, Multi-label imbalanced data enrichment process in neural net classifier training, in: 2008 IEEE International Joint Conference on Neural Networks, IEEE, 2008, pp. 1301–1307, doi:10.1109/IJCNN.2008.4633966.

[60] F.F. Luo, W.Z. Guo, G.L., Addressing imbalance in weakly supervised multi-label learning, IEEE Access (2019), doi:10.1109/ACCESS.2019.2906409.

[61] J. He, H. Gu, W. Liu, Imbalanced multi-modal multi-label learning for subcellular localization prediction of human proteins with both single and multiple sites, PLoS One 7 (2012) e37155, doi:10.1371/journal.pone.0037155.

[62] M.L. Zhang, Ml-rbf: RBF Neural Networks for Multi-Label Learning, Neural Processing Letters, 2009, doi:10.1007/s11063-009-9095-3.

[63] M.L. Zhang, Z.J. Wa, MIMLRBF: RBF neural networks for multi-instance multi-label learning, Neurocomputing (2009), doi:10.1016/j.neucom.2009.07.008.

[64] K.W. Sun, C.H. Lee, Addressing class-imbalance in multi-label learning via two-stage multi-label hypernetwork, Neurocomputing (2017), doi:10.1016/j.neucom.2017.05.049.

[65] M.L. Zhang, Y.K. Li, X.Y. Liu, Towards class-imbalance aware multi-label learning, IJCAI International Joint Conference on Artificial Intelligence, 2015.

[66] H. Han, M. Huang, Y. Zhang, J. Liu, Decision support system for medical diagnosis utilizing imbalanced clinical data, Appl. Sci. 8 (2018) 1597, doi:10.3390/app8091597.

[67] S. Pouyanfar, T. Wang, S.-C. Chen, A multi-label multimodal deep learning framework for imbalanced data classification, in: 2019 IEEE Conference on MIPR, IEEE, 2019, pp. 199–204, doi:10.1109/MIPR.2019.00043.

[68] K. Sozykin, S. Protasov, A. Khan, R. Hussain, J. Lee, Multi-label class-imbalanced action recognition in hockey videos via 3D convolutional neural networks, in: Proceedings - 2018 IEEE/ACIS 19th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD, 2018, p. 2018, doi:10.1109/SNPD.2018.8441034.

[69] C. Li, G. Shi, Improvement of learning algorithm for the multi-instance multi-label RBF neural networks trained with imbalanced samples, J. Inf. Sci. Eng. (2013).

[70] Y. Xie, D. Li, D. Zhang, H. Shuang, An improved multi-label relief feature selection algorithm for unbalanced datasets, Advances in Intelligent Systems and Computing, 2018, doi:10.1007/978-3-319-69096-4_21.

[71] E.S. Xioufis, M. Spiliopoulou, G. Tsoumakas, I. Vlahavas, Dealing with concept drift and class imbalance in multi-label stream classification, IJCAI International Joint Conference on AI, 2011, doi:10.5591/978-1-57735-516-8/IJCAI11-266.

[72] Eva L.Gibaja M.Moyano, Krzysztof J.Cios, Sebastián Ventura, Review of ensembles of multi-label classifiers: models, experimental study and prospects, Inf. Fusion 44 (2018) 33–45 November.

[73] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. (1998), doi:10.1109/34.667881.

[74] M.A. Tahir, J. Kittler, F. Yan, Inverse random under sampling for class imbalance problem and its application to multi-label classification, Pattern Recognit. 45 (2012) 3738–3750, doi:10.1016/j.patcog.2012.03.014.

[75] M.A. Tahir, J. Kittler, A. Bouridane, Multilabel classification using heterogeneous ensemble of multi-label classifiers, Pattern Recognit. Lett. 33 (2012) 513–523, doi:10.1016/j.patrec.2011.10.019.

[76] S. Wan, Y. Duan, Q. Zou, HPSLPred: an ensemble multi-label classifier for human protein subcellular location prediction with imbalanced source, PROTEOMICS 17 (2017) 1700262, doi:10.1002/pmic.201700262.

[77] B.M.N. Arjun Pakrashi, Stacked-MLkNN: a stacking based improvement to multi-label k-nearest neighbours, The 1st International Workshop on Learning with Imbalanced Domains: Theory and Applications, 2017.

[78] G.T. Bin Liu, Making classifier chains resilient to class imbalance, in: Proceedings of Machine Learning Research, Beijing, 2018 2018.

[79] M.A. Tahir, J. Kittler, K. Mikolajczyk, F. Yan, Improving multilabel classification performance by using ensemble of multi-label classifiers, in: 2010: pp. 11–21. https://doi.org/10.1007/978-3-642-12127-2_2.

[80] G.I. Winata, M.L. Khodra, Handling imbalanced dataset in multi-label text categorization using bagging and adaptive boosting, in: 2015 International Conference on Electrical Engineering and Informatics (ICEEI), IEEE, 2015, pp. 500–505, doi:10.1109/ICEEI.2015.7352552.

[81] Y. Sun, M.S. Kamel, A.K.C. Wong, Y. Wang, Cost-sensitive boosting for classification of imbalanced data, Pattern Recognit. 40 (2007) 3358–3378, doi:10.1016/j.patcog.2007.04.009.

[82] Z.A. Daniels, D.N. Metaxas, Addressing imbalance in multi-label classification using structured hellinger forests, in: 31st AAAI Conference on AI AAAI, 2017, p. 2017.

[83] P. Cao, X. Liu, D. Zhao, O. Zaiane, Cost sensitive ranking support vector machine for multi-label data learning, in: Advances in Intelligent Systems and Computing, 2017, pp. 244–255, doi:10.1007/978-3-319-52941-7_25.

[84] G. Wu, Y. Tian, D. Liu, Cost-sensitive multi-label learning with positive and negative label pairwise correlations, Neural Netw. 108 (2018) 411–423, doi:10.1016/j.neunet.2018.09.003.

[85] J. V. Tsoumakas, G., Xioufis, E.S., Vilcek, MULAN multi-label dataset repository, (n.d.). http://mulan.sourceforge.net/datasets-mlc.html.

[86] J. Read, P. Reutemann, B. Pfahringer, G. Holmes, MEKA: a multi-label/multi-target extension to WEKA, J. Mach. Learn. Res. 17 (2016).

[87] F. Charte, D. Charte, A. Rivera, M.J. del Jesus, F. Herrera, R Ultimate Multilabel Dataset Repository, 2016, pp. 487–499. Lecture Notes in Computer Science, doi:10.1007/978-3-319-32034-2_41.

[88] G. Tsoumakas, E. Spyromitros-Xioufis, J. Vilcek, I. Vlahavas, MULAN: a java library for multi-label learning, J. Mach. Learn. Res. (2011).

[89] P. Szymański, T. Kajdanowicz, Scikit-multilearn: Scikit-based Python environment for performing multi-label classification, J. Mach. Learn. Res. (2019).

[90] G. Tsoumakas, I. Katakis, Multi-label classification: An overview, Int. J. Data Warehous. (2007), doi:10.4018/jdwm.2007070101.

[91] S. Godbole, S. Sarawagi, Discriminative Methods for Multi-labeled Classification, 2004, pp. 22–30. Lecture Notes in Computer Science, doi:10.1007/978-3-540-24775-3_5.

[92] O. Maimon, L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook, Springer US, Boston, MA, 2010, doi:10.1007/978-0-387-09823-4.

[93] F. Herrera, F. Charte, A.J. Rivera, M.J. Del Jesus, Multilabel classification: Problem analysis, metrics and techniques, 2016, doi:10.1007/978-3-319-41111-8.

[94] E. Gibaja, S. Ventura, A tutorial on multilabel learning, ACM Comput. Surv. 47 (2015) 1–38, doi:10.1145/2716262.

[95] R.B. Pereira, A. Plastino, B. Zadrozny, L.H.C. Merschmann, Correlation analysis of performance measures for multi-label classification, Inf. Process. Manag. (2018), doi:10.1016/j.ipm.2018.01.002.

[96] G. Madjarov, D. Kocev, D. Gjorgjevikj, S. Džeroski, An extensive experimental comparison of methods for multi-label learning, Pattern Recognit. (2012), doi:10.1016/j.patcog.2012.03.004.
[97] F. Charte, A. Rivera, M.J. Del Jesus, F. Herrera, Resampling multilabel datasets by decoupling highly imbalanced labels, Lect. Notes Artif. Intell. (2015), doi:10.1007/978-3-319-19644-2_41.
[98] W. Chmielnicki, K. Stąpor, Using the one–versus–rest strategy with samples balancing to improve pairwise coupling classification, Int. J. Appl. Math. Comput. Sci. 26 (2016) 191–201, doi:10.1515/amcs-2016-0013.
[99] B. Wu, S. Lyu, B.G. Hu, Q. Ji, Multi-label learning with missing labels for image annotation and facial action unit recognition, Pattern Recognit. (2015), doi:10.1016/j.patcog.2015.01.022.
[100] M. Ding, Y. Yang, Multi-label imbalanced classification based on assessments of cost and value, Appl. Intell. (2018), doi:10.1007/s10489-018-1156-8.
[101] V. Kumar, A.K. Pujari, V. Padmanabhan, V.R. Kagita, Group preserving label embedding for multi-label classification, Pattern Recognit. (2019), doi:10.1016/j.patcog.2019.01.009.
[102] I. Dimou, M. Zervakis, On the analogy of classifier ensembles with primary classifiers: statistical performance and optimality, J. Pattern Recognit. Res. (2013), doi:10.13176/11.497.
[103] M.R. Ahmadzadeh, M. Petrou, Use of Dempster-Shafer theory to combine classifiers which use different class boundaries, Pattern Anal. Appl. (2003), doi:10.1007/s10044-002-0176-4.
[104] G. Tsoumakas, I. Katakis, I. Vlahavas, Random k-labelsets for multilabel classification, IEEE Trans. Knowl. Data Eng. (2011), doi:10.1109/TKDE.2010.164.

Adane Tarekegn received M.Sc. degree in Computer Science at the University of Gondar, Ethiopia, 2015. Currently, he is a PhD student in modelling and data science at the University of Torino, Italy. His research interests include machine learning, data analytics, pattern recognition, and related topics.

Mario Giacobini is Associate Professor in Informatics at the Department of Veterinary Sciences of the University of Torino, Italy, where he leads the Data Analysis and Modeling Unit. His research interests concentrate on the development of data science approaches to the study and modeling of biological phenomena.

Krzysztof Michalak received the Ph.D. in computer science in 2010 and currently is an assistant professor at the Department of Information Technologies of the Wroclaw University of Economics, Poland. His interests include graph-based optimization, metaheuristic algorithms, machine learning, knowledge-based optimization and real-world applications such as epidemics control and systemic risk mitigation.