Boosting the Performance of Evolutionary Multi-Instance Classification with Transfer Learning

Nadia Omri[1], Maha Elarbi[1], Slim Bechikh[1,2], Carlos A. Coello Coello[3]
[1]*SMART Lab, Computer Science Department, ISG, University of Tunis, Tunisia*
[2]*COL Lab, School of Computer Science, University of Nottingham, UK*
[3]*Cinvestav (Evolutionary Computation Group), Departamento de Computación, Mexico City 07360, Mexico*
Corresponding author email: nadiaomri37@gmail.com

***Abstract*—Multiple Instance Learning (MIL) has proven effective for learning from weakly labeled data, where labels are provided at the bag level rather than at the instance level. Among existing MIL approaches, MILEIS (Multiple-Instance Learning with Evolutionary Instance Selection) is an evolutionary instance selection–based framework that reduces instance ambiguity by mapping bags into a discriminative feature space using optimized instance prototypes. However, many existing MIL approaches including MILEIS, rely solely on target task data and do not exploit knowledge from related tasks, which can limit their performance in data-scarce scenarios. To address this limitation, transfer learning enables the reuse of knowledge from source tasks to improve target task learning. In this paper, we propose TL-MILEIS, a transfer learning extension of MILEIS that incorporates a task similarity evaluation mechanism to ensure safe knowledge transfer. The proposed approach first performs instance selection following the MILEIS paradigm, and then evaluates the similarity between source and target tasks to determine whether knowledge transfer is beneficial. When task relevance is established, discriminative information from the source task is leveraged to improve target task learning; otherwise, standard MILEIS is applied. Experimental results on multiple benchmark datasets show that TL-MILEIS consistently outperforms classical MIL methods, evolutionary MIL approaches, and existing multiple instance transfer learning techniques, highlighting the effectiveness of incorporating transfer learning into evolutionary MIL frameworks.**

***Index Terms*—Multiple instance classification, Transfer learning, Evolutionary optimization.**

I. Introduction

Over the past few decades, supervised learning has become a dominant paradigm for training predictive models. However, in many real-world applications, including text categorization, the availability of fully annotated data remains limited, making it important for machine learning algorithms to learn effectively from weak or insufficient labeling information [1]. Multiple Instance Learning (MIL) addresses this limitation by relaxing the requirement for instance-level labels and learning instead from labeled bags of instances. In the context of text categorization, documents can be naturally represented as bags of sentences or paragraphs where only document-level labels are available while the relevance of individual segments remains unknown. MIL is well suited for such weakly labeled text data, as it enables the identification of informative textual patterns without requiring fine-grained annotations.

In the MIL framework, data instances are organized into groups called bags, where each of them may contain a variable number of instances. Labels are assigned only at the bag level, while the labels of individual instances remain unknown during both training and testing. Given weakly labeled bags, MIL aims to learn a function that predicts bag labels, where key instances responsible for the label are called witnesses [1].

Despite significant progress in the field of MIL, the existing methods continue to face remarkable challenges, particularly when confronted with noisy data, scarce training samples, or poorly representative instances. Moreover, MIL classifiers are typically trained in isolation for a specific task, which limits their ability to generalize or to be efficiently adapted to related tasks, often resulting in suboptimal performance when data are limited. To address these issues, transfer learning has emerged as a promising strategy for enhancing MIL by enabling knowledge reuse across tasks. Transfer learning is a machine learning paradigm in which a model trained on a source task is adapted to a different but related target task, allowing previously learned representations to guide learning in the target domain. Rather than training a model from scratch, this approach exploits shared information between tasks, leading to improved robustness and performance.

In this paper, we address the problem of transfer learning for multiple instance learning on text categorization problems. The main contributions of this paper are as follows:

1) We develop a new transfer learning–based evolutionary approach, called TL-MILEIS, which extends the MILEIS algorithm by enabling the effective transfer of knowledge from a related source task to a target task. By leveraging informative patterns learned from previous tasks, TL-MILEIS, improves generalization, and enhances classification performance in limited and weakly supervised MIL scenarios.
2) We propose a new framework that evaluates the similarity between the source and target tasks before conducting the classification process to avoid negative transfer. If both tasks are found to be similar, we propose a novel evolutionary multiple-instance transfer learning classifier to effectively transfer knowledge from the source task to the target task.
3) We present experimental findings from various real-world MIL datasets in order to test the efficiency and

performance of our proposed transfer learning approach in terms of classification accuracy compared to classical and transfer learning methods.

The remainder of this paper is organized as follows. Section II provides a short review of the previous related work in this area. Section III describes our proposed transfer learning based evolutionary approach for Multiple Instance Classification (MIC) including the different motivations behind our proposal. Section IV presents the experimental results obtained in this study. Finally, Section V concludes the paper and provides some possible paths for future research.

II. Previous Related work

Several methods have been proposed in the literature to deal with the MIC problem: (1) classical Machine Learning methods, (2) evolutionary approaches [15], and (3) transfer learning based approaches. Representative approaches of each of these types of methods are provided next. Additionally, we also describe our proposed baseline evolutionary framework, called MILEIS.

A. Classical ML approaches for MIC

The Axis-Parallel Hyper-Rectangle (APR) algorithm proposed by Dietterich et al. [3], aims to construct a hyper-rectangle enclosing at least one instance from each positive bag while excluding all instances from negative bags. Probabilistic approaches were later introduced, such as the Diverse Density (DD) framework and its extension, the Expectation–Maximization Diverse Density (EM-DD) method by Zhang et al. [5], which iteratively identifies the most representative instances responsible for positive bag labels. Margin-based methods have been widely adopted as well. Andrews et al. [6] proposed mi-SVM and MI-SVM, which extend support vector machines to instance-level and bag-level MIL settings by jointly inferring latent instance labels and decision boundaries. Feature-embedding approaches were later proposed such as MILES [10] that maps bags into an instance-based feature space and employs sparse SVMs for classification. Sun et al. [7] proposed a weakly supervised Convolutional Neural Networks (CNN) framework called MILCNN. This framework was originally suggested for the data augmentation problem. It combines the CNN with a particular MIL loss function derived with regards to the bag in order to evaluate data augmentation generated images. More recently, classical kernel-based methods have been further enhanced through efficient optimization strategies, such as the Semiproximal SVM (SPSVM) proposed by Avolio et al. [8], which combines accuracy, computational efficiency, and nonlinear kernel extensions for binary MIL problems.

B. Evolutionary approaches for MIC

Genetic Programming (GP) has been adapted to MIL settings. In this context, Zafra et al. [9] proposed Grammar-Guided Genetic Programming (G3P), where each individual encodes an interpretable IF–THEN rule generated from a context-free grammar. The IF part represents logical combinations of instance-level feature constraints, while the THEN part predicts the bag label, enabling direct bag-level classification. Instance selection has also been addressed using evolutionary strategies. Zhang et al. [10] introduced MILEIS, which combines a Gaussian kernel–based density estimator for initial instance selection with an evolutionary process to identify informative Instance Prototypes (IPs). MILEIS achieves classification accuracy and serves as the baseline search engine for the framework proposed here, as detailed in Subsection II-D.

C. Transfer Learning-based approaches

Early work by Thrun [11] highlighted the importance of previously learned knowledge for improving generalization, laying the foundation for transfer learning research. Subsequent studies explored different transfer mechanisms, including shared latent structures across tasks [14], cross-language classification via information-theoretic principles, heterogeneous transfer learning across distinct feature spaces, and cross-domain text classification through probabilistic latent semantic analysis. Many transfer learning approaches rely on assumptions such as shared parameters or related data distributions between source and target tasks. However, these assumptions often require prior knowledge and may limit their applicability in real-world settings. Within the Multiple Instance Learning framework, Zhang et al. [13] introduced a Multiple Instance Transfer Learning (MITL) approach, proposing a discriminative model that enables effective knowledge transfer across domains while accounting for the bag–instance structure. To further reduce cross-domain bias, Tian et al. [4] proposed a sparse coding–based method that maps source and target data into a common feature space. More recently, Liu et al. [14] developed SMITL, a selective transfer learning approach for MIC in text categorization, which evaluates task relatedness before transferring knowledge and constructs a source-aware classifier to enhance target performance.

D. Baseline Approach: MILEIS

MILEIS [10] is an evolutionary instance-selection-based MIL approach that transforms the MIL problem into a standard single instance learning task through instance-based feature mapping. Rather than relying solely on the original instances in the training bags, MILEIS introduces an evolutionary search process to identify a set of IPs that better represent the underlying concept of positive bags. A Gaussian kernel density estimator (KDE) is first employed to select a pool of promising candidate instances from the training data, which are then refined through genetic operations (i.e., mutation, crossover, and selection), allowing the algorithm to explore a broader and more flexible prototype space. Each optimized IP defines a feature that measures the similarity between a bag and the prototype, embedding bags into a new feature space where conventional classifiers can be applied. Although MILEIS effectively mitigates instance ambiguity and enhances classification accuracy, its optimization relies exclusively on target task data, limiting its performance under scarce, noisy, or weakly representative training conditions and preventing the model from leveraging useful patterns from related tasks.

III. Our Proposed Transfer learning-based Evolutionary Approach for MIC (TL-MILEIS)

A. Motivation and Core Idea

Even though current MIL techniques are effective, efficiency is highly dependent on both the quality and quantity of labeled training data. While labels remain weak, noisy, or ambiguous at the instance level, it is often costly or difficult to obtain sufficiently large and representative training sets in real-world MIL scenarios. Since the learned representations are based on constrained and possibly biased data distributions, models trained exclusively on the target domain often show poor generalization and unstable instance selection. Transfer learning offers a solid solution to these challenges by enabling the reuse of knowledge acquired from related source domains, where richer or more reliable information may be available. Although MILEIS effectively reduces information loss by employing evolutionary search, it still relies exclusively on target training data, and its performance may degrade under limited samples or high noise. Transfer learning is particularly useful when the target domain has very few labeled data. Unfortunately, its effectiveness is not always guaranteed. When two tasks are unrelated, transferred knowledge may hurt target performance, a phenomenon known as *negative transfer*. It is therefore crucial to assess the similarity of source and target tasks before conducting the classification process.

In order to address all of the previously mentioned research gaps, combining the MILEIS approach with transfer learning can lead to better classification accuracy. Therefore, we have proposed a new transfer learning-based evolutionary approach (i.e., TL-MILEIS). Figure 1 illustrates the main scheme of our proposed approach. First, we identify the IPs from the source and the target tasks, and then we calculate the similarity between both tasks. After that, If the source task and the target task are identified as related, we construct a classifier to transfer knowledge from the source task to the target task. Otherwise, the knowledge transfer will degrade the accuracy of the classifier. In this case, we will not perform the transfer learning step, and the approach becomes the original MILEIS method to classify the target bags.

B. Initial instance selection

The KDE picks the most negative instance from each negative bag and the most positive instance from each positive bag using the likelihood value of each instance [10]:

$$f(x) = \frac{1}{Z\sum_i n_i} \sum_{yi=-1} \sum_{j=1}^{n} \exp(-\gamma \|x - x_{i,j}^{-}\|) \qquad (1)$$

where x_i^-, j denotes the j^{th} instance in the i^{th} negative bag, x is an instance, Z is a constant normalizing factor that can be disregarded in our approach, $yi = -1$ denotes the label of a negative bag, n_i is the number of instances of the i^{th} bag, and γ is a scale parameter used to limit the amount of influence for training instances.

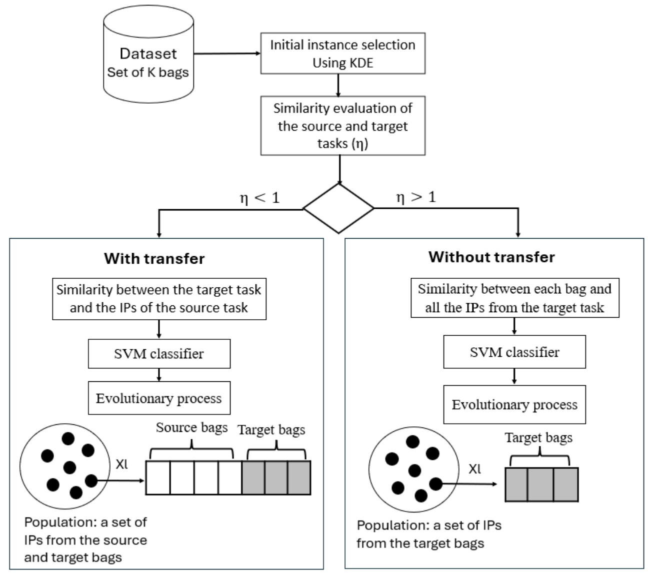

Fig. 1. The workflow of of our proposed approach, called TL-MILEIS.

C. Similarity evaluation

The main task of multiple instance transfer learning is to transfer knowledge from a source task to a target task. However, the two tasks may not be related in reality, and transfer may be ineffective or may even hurt the target task. To avoid this, this step consists on evaluating the similarity between the source and the target tasks.

We first extract positive features from the positive bags for the source and target tasks respectively. First of all, for the source task, we put the instances from the positive bags into $S_s^+ = [x_1, . . ., xS_s^+]$ and store the instances of the negative bags into $S_s^- = [x_1, . . ., xS_s^-]$. Similarly, for the target task, we have S_t^+ and S_t^- to store positive bag instances and negative bag instances, respectively. The set of positive features for S_s^+ is denoted by VP_s. We extract the positive features which appear frequently in positive bags S_s^+ while occurring less frequently in negative bags S_s^-.

Feature strength is utilized to indicate the importance of a feature in classification. In general, a feature with a higher feature strength indicates that it is more effective in terms of classification. We do not have any prior knowledge about the feature distribution between the positive examples and the negative examples. For S_s^+, we calculate the feature strength of a particular feature f_k by measuring the differences of the normalized examples frequency between S_s^+ and S_s^- [14]:

$$H_s(f_k) = \frac{n_{S_s^+}(f_k) - min_{S_s^+}}{max_{S_s^+} - min_{S_s^+}} - \frac{n_{S_s^-}(f_k) - min_{S_s^-}}{max_{S_s^-} - min_{S_s^-}} \qquad (2)$$

Here, $n_{S_s^+}(f_k)$ and $n_{S_s^-}(f_k)$ represent the number of examples that contain f_k in S_s^+ and S_s^-, respectively, $max_{S_s^+}$ and $min_{S_s^+}$ are the maximum and minimum values of $n_{S_s^+}(f_k)$

for $f_k \in V_s$ where V_s is the feature set of the source task. $max_{S_s^-}$ and $min_{S_s^-}$ are the maximum and minimum values of $n_{S_s^-}(f_k)$ for $f_k \in V_s$.

In the multi-instance textual learning, the positive features should appear more frequently in the positive bags and should appear less in the negative bags. Therefore, positive features should have a relatively large value of $H_s(f_k)$ in equation (2). Thus, the positive features are identified by extracting f_k such that $H_s(f_k) > \theta_s$ [14], where:

$$\theta_s = \frac{1}{V_s} \sum_{f_k \in V_s} H_s(f_k) \tag{3}$$

After that, we extract positive features for the source task and put them into VP_s. We then let:

$$G_s = \frac{1}{VP_s} \sum_{f_k \in VP_s} H_s(f_k) \tag{4}$$

Similarly, for the target task, we calculate feature strength $H_t(f_k)$, positive feature set VP_t and G_t [14].

We compare the similarity of the positive features between two tasks to identify whether they are related or not. Specifically, for the sets VP_s and VP_t, the feature strength of feature f_k ($f_k \in VP_s \cup VP_t$) between S_s^+ and S_t^+ , denoted as $H_c(f_k)$, is calculated in the same way as equation (2):

$$H_c(f_k) = \frac{n_{S_s^+}(f_k) - min_{S_s^+}}{max_{S_s^+} - min_{S_s^+}} - \frac{n_{S_t^+}(f_k) - min_{S_t^+}}{max_{S_t^+} - min_{S_t^+}} \tag{5}$$

Let

$$G_c = \frac{1}{VP_s \cup VP_t} \sum_{f_k \in VP_s \cup VP_t} |H_c(f_k)| \tag{6}$$

We then calculate [14]:

$$\eta = \frac{G_c}{min(G_s, G_t)} \tag{7}$$

The smaller the η, the more likely both tasks are related [14]. If η is larger than 1, then both tasks are not related.

D. TL-MILEIS Transfer learning decision

If the source task and the target task are identified as related ($\eta < 1$), we then construct a classifier to transfer knowledge from the source task to the target task. Otherwise, ($\eta > 1$), it is more likely that the classifier performance will not be improved. In this case, we won't conduct the transfer learning step, and the approach reduces to the original MILEIS method.

1) With Transfer Learning : In this step, where the source task and the target task are related, we calculate the similarity measure between each bag from the target task B_{ti} and all the IPs from the source task that are selected in the initial instance selection step. The similarity measure is computed based on the Hausdorff distance as follows [10]:

$$s(B_i, x) = \exp(-\lambda\ d(B_i, x)) = \max_{x_i \in B_i} \exp(-\lambda \|x_i - x\|^2) \tag{8}$$

where λ is a scaling parameter that allows amplifying the bag-to-instance distance [10]. The higher the λ value, the higher the influence of the distance magnitude, and vice versa. The goal of scaling is to ensure: (a) each positive bag has a high similarity to at least one IP and (b) each negative bag has a low similarity with respect to all IPs. This makes the classifier more capable of distinguishing between positive and negative bags, $d(B_i, x)$ denotes the distance between an instance x and its nearest neighbor in B_i, and $d(B_i, x) = \min_{x_i \in B_i} \|x_i - x\|^2$.

We then train a classifier that can be applied to the bag features. To this end, we use the ELM classifier as in the original MILEIS paper. After acquiring the classifier result for bag-level features, we can verify the chosen IPs and update them as needed in order to obtain the optimal set of IPs; hence, an evolutionary process is adopted. The solution is encoded as a vector where the size of the vector is the number of training bags in both source and target tasks. Regarding the crossover operator, we applied uniform crossover since it enables us to vary all chromosomal segments, where each gene in the offspring has an equal chance of being chosen from either parent using a randomly generated binary mask. For the mutation operator, we applied a two-steps uniform mutation operator, where we randomly select a gene (i.e., an IP from the vector) and then modify its value with another instance from the corresponding bag. Finally, the fitness function is obtained by the precision and recall values generated using a defined classifier, formulated as follows:

$$Gmean = \sqrt{(precision \times recall)} \tag{9}$$

After the computation of the fitness function value of all the solutions, the solution with the maximum $Gmean$ value will be selected as the optimum solution.

2) Without transfer learning (MILEIS): In this step, we calculate the similarity measure between each bag B_i and all the IPs from the target task selected in the initial instance selection step. The similarity measure is computed based on the Hausdorff distance. The similarity matrix combines the bag-level features of the target bags. Each line i of the similarity matrix contains the similarity measure values between the i^{th} bag from the target task and each instance prototype selected from the target task. It is important to note that the number of selected IPs from each bag is equal to the number of bags in the dataset. As for the classifier in this step, we use the ELM classifier for the bag features. Finally, in the evolutionary step, we apply the same genetic operators as in the transfer learning approach. However, a key distinction emerges in the way solutions are represented where we consider the target bags.

IV. EXPERIMENTAL STUDY

A. Baseline and performance metrics

In this section, we investigate the performance of our proposed TL-MILEIS against five baseline methods: The classical ML approaches mi-SVM [6] and SubMIL [12], the evolutionary-based approach MILEIS [10], and the transfer learning-based approaches MITL [13] and SMITL [14].

Classification accuracy is used as the performance measure throughout our experiments.

B. Datasets and settings

To evaluate our framework, we conducted experiments on three datasets. The 20 Newsgroups corpus contains 20 subcategories with 1000 samples each. Reuters-21578 contains news wire articles organized into five top categories with different sub-categories each. Web-KB contains 8,282 web pages across seven categories. Since these datasets were not originally designed for multiple instance learning or transfer learning, they were reorganized accordingly following [14]. Task similarity between source and target domains was also computed to determine whether transfer learning was applicable. For each target dataset, 10% of the samples were used for training and the remainder for testing, reflecting a realistic transfer learning scenario with limited labeled target data. All source data were incorporated into training. The procedure was repeated ten times and results are reported as average classification accuracy over the twelve datasets.

As for the parameters settings, TL-MILEIS adopted the same evolutionary parameters as those used in the original MILEIS framework, including population size, crossover rate, mutation rate, and ELM classifier, in order to ensure a fair comparison. To evaluate the statistical significance of the performance differences between pairs of algorithms, we employed the Wilcoxon signed-rank test at $\alpha = 0.05$.

C. Experimental results and discussion

Among the evaluated source–target task pairs, the first four as well as the ninth to twelfth pairs shared positive classes belonging to the same top-level category and are consequently considered related tasks. For these pairs, TL-MILEIS correctly identified task relatedness and activated the transfer learning mechanism. Table I summarizes the comparative performance of mi-SVM, SubMIL, MILEIS, MITL, SMITL, and our proposed TL-MILEIS. The results show that TL-MILEIS consistently achieved the highest accuracy across all related task pairs. In addition, MITL also outperformed mi-SVM and MILEIS, highlighting the benefit of incorporating transfer knowledge. In contrast, mi-SVM, SubMIL, and MILEIS exhibited inferior performance, as they rely exclusively on the limited target data and do not exploit information from the source domain. These results empirically demonstrate the effectiveness of transfer learning in enhancing multiple instance classification under data-scarce conditions.

For the fifth, sixth, seventh, eighth, eleventh, and twelfth source–target task pairs, the tasks belong to different top-level categories and are therefore considered unrelated. Transferring knowledge in such cases may degrade target performance, referred to as negative transfer (cf. Table II). The task similarity criterion in TL-MILEIS correctly identifies these pairs as unrelated and avoids performing the transfer learning step, reducing to a standard MILEIS algorithm and achieving identical performance. In contrast, MITL does not explicitly assess task relevance and applies transfer learning randomly, meaning that information from unrelated source tasks does not contribute positively to target learning. Under these conditions, MILEIS shows better performance than MITL, largely due to its evolutionary instance selection mechanism, which remains effective even in the absence of transferable knowledge.

Figure 2 illustrates the performance variation of six MIL methods: mi-SVM, SubMIL, MILEIS, MITL, SMITL, and our proposed TL-MILEIS evaluated over 31 independent runs on the first dataset pair. Each boxplot summarizes the variability of the results in terms of median values providing vision into both accuracy and stability. As shown in the figure, classical MIL methods such as mi-SVM and SubMIL exhibit lower median performance and higher variability compared to evolutionary and transfer learning–based approaches. MILEIS showed remarkable improvements, reflecting the benefits of using an evolutionary process. Remarkably, SMITL and MITL achieve better performance values which shows the advantage of knowledge transfer. However, TL-MILEIS achieves the highest median performance and consistently outperforms all

TABLE I

ACCURACY OBTAINED BY MI-SVM, SUBMIL, MILEIS, MITL, SMITL, AND TL-MILEIS ON THE TARGET TASK FOR THE RELATED TASKS. THE NOTATION (+) AND (-), RESPECTIVELY, INDICATES THE PRESENCE OR ABSENCE OF STATISTICAL SIGNIFICANCE.

Datasets	**mi-SVM**	**SubMIL**	**MILEIS**	**MITL**	**SMITL**	**TL-MILEIS**
1	0.617 (+++++)	0.581 (++++)	0.623 (+++)	0.662 (-+)	0.666 (+)	0.682
2	0.619 (+++++)	0.630 (-+++)	0.633 (+++)	0.652 (++)	0.687 (+)	0.713
3	0.811 (+++++)	0.768 (++++)	0.830 (-++)	0.831 (++)	0.855 (+)	0.887
4	0.580 (+++++)	0.569 (++++)	0.640 (-++)	0.643 (++)	0.662 (+)	0.698
9	0.733 (+++++)	0.740 (++++)	0.759 (+++)	0.768 (++)	0.810 (+)	0.853
10	0.750 (+++++)	0.762 (++++)	0.773 (+++)	0.821 (++)	0.873 (-)	0.879
12	0.630 (+++++)	0.641 (++++)	0.688 (+++)	0.713 (++)	0.798 (-)	0.801
13	0.705 (+++++)	0.719 (++++)	0.733 (+++)	0.837 (++)	0.840 (+)	0.868

TABLE II

ACCURACY OF MI-SVM, SUBMIL, MILEIS, MITL, SMITL, AND TL-MILEIS ON THE TARGET TASK FOR THE NON-RELATED TASKS. THE NOTATION (+) AND (-), RESPECTIVELY, INDICATES THE PRESENCE OR ABSENCE OF STATISTICAL SIGNIFICANCE.

Datasets	**mi-SVM**	**SubMIL**	**MILEIS**	**MITL**	**SMITL**	**TL-MILEIS**
5	0.711 (+++++)	0.699 (++++)	0.754 (++-)	0.620 (++)	0.712 (+)	0.754
6	0.767 (-++++)	0.774 (++++)	0.820 (++-)	0.717 (++)	0.784 (+)	0.820
7	0.680 (-++++)	0.685 (++++)	0.712 (++-)	0.584 (++)	0.704 (+)	0.712
8	0.780 (+++++)	0.771 (++++)	0.799 (++-)	0.712 (++)	0.788 (-)	0.790
11	0.769 (+++++)	0.781 (++++)	0.805 (++-)	0.732 (++)	0.789 (+)	0.805
12	0.720 (+++++)	0.737 (++++)	0.796 (++-)	0.695 (++)	0.774 (+)	0.796

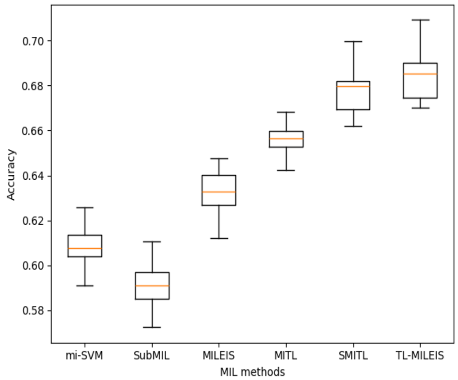

Fig. 2. Performance variation of MIL Methods on the first dataset pair over 31 Independent Runs.

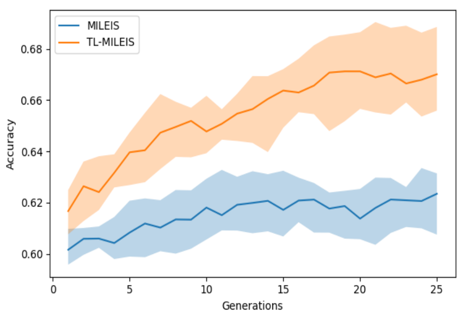

Fig. 3. Evolution of MILEIS and TL-MILEIS on the first dataset pair over 25 generations.

baselines, highlighting the effectiveness of incorporating safe transfer learning into the evolutionary MIL framework.

Figure 3 illustrates the evolutionary behavior of MILEIS and TL-MILEIS over 25 generations on the first dataset pair. Each curve represents the median classification performance over 31 independent runs, with shaded areas denoting the interquartile range (Q1–Q3). TL-MILEIS consistently maintains higher median performance throughout the evolutionary process and exhibits a narrower interquartile range, indicating improved stability and robustness compared to MILEIS. Both methods exhibit gradual performance improvements rather than abrupt gains, which is consistent with evolutionary optimization under noisy and weakly labeled MIL settings. However, TL-MILEIS benefits from knowledge transferred from a related source task, providing superior performance across generations. These results further suggest that safe transfer learning helps to regularize the evolutionary process, reducing the effects of noise and limited training data.

V. Conclusions and Future Work

In this work, we introduced TL-MILEIS, a transfer learning evolutionary MIL approach designed to enhance multiple instance classification under weak supervision and limited training data. Unlike conventional MIL approaches that operate exclusively on the target task, TL-MILEIS explicitly exploits knowledge from related source tasks by incorporating a task similarity evaluation mechanism prior to transfer. This strategy allows the model to selectively reuse informative patterns from other domains while avoiding negative transfer when tasks are unrelated. Through extensive experimental evaluation, TL-MILEIS showed a superior classification performance and improved robustness compared to classical MIL methods, evolutionary MIL approaches, and existing multiple instance transfer learning techniques. The results confirm that guiding instance selection and evolutionary representation with transferred knowledge significantly enhances learning effectiveness in challenging MIL scenarios.

As part of our future work, we plan to extend TL-MILEIS beyond text categorization to other multiple instance learning domains, such as image classification, medical image analysis, audio event detection, and bioinformatics applications. In addition, an important research direction involves exploring TL-MILEIS in the context of multi-source transfer learning, where knowledge is simultaneously transferred from multiple related source tasks rather than a single source [16].

References

[1] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, “Multiple instance learning: A survey of problem characteristics and applications,” *Pattern Recognition*, vol. 77, pp. 329–353, 2018.

[2] M. Ilse, J. Tomczak, and M. Welling, “Attention-based deep multiple instance learning,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2127–2136.

[3] T. G. Dietterich, R. H. Lathrop, and T. Lozano-P’erez, “Solving the multiple instance problem with axis-parallel rectangles,” *Artificial Intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.

[4] M. Gholizade, H. Soltanizadeh, M. Rahmanimanesh, and S. S. Sana, “A review of recent advances and strategies in transfer learning,” *Int. J. Syst. Assur. Eng. Manag.*, pp. 1–40, 2025.

[5] Q. Zhang and S. Goldman, “EM-DD: An improved multiple-instance learning technique,” *Adv. Neural Inf. Process. Syst.*, vol. 14, 2001.

[6] S. Andrews, I. Tsochantaridis, and T. Hofmann, “Support vector machines for multiple-instance learning,” *Adv. Neural Inf. Process. Syst.*, vol. 15, 2002.

[7] M. Sun, T. X. Han, M.-C. Liu, and A. Khodayari-Rostamabad, “Multiple instance learning convolutional neural networks for object recognition,” in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, 2016, pp. 3270–3275.

[8] M. Avolio and A. Fuduli, “The semiproximal SVM approach for multiple instance learning: a kernel-based computational study,” *Optim. Lett.*, vol. 18, no. 2, pp. 635–649, 2024.

[9] A. Zafra and S. Ventura, “G3P-MI: A genetic programming algorithm for multiple instance learning,” *Inf. Sci.*, vol. 180, no. 23, pp. 4496–4513, 2010.

[10] Y. Zhang, J. Wu, C. Zhou, P. Zhang, and Z. Cai, “Multiple-instance learning with evolutionary instance selection,” in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2016, pp. 229–241.

[11] S. Thrun, “Is learning the n-th thing any easier than learning the first?,” *Adv. Neural Inf. Process. Syst.*, vol. 8, 1995.

[12] J. Yuan, X. Huang, H. Liu, B. Li, and W. Xiong, “SubMIL: Discriminative subspaces for multi-instance learning,” *Neurocomputing*, vol. 173, pp. 1768–1774, 2016.

[13] D. Zhang and L. Si, “Multiple instance transfer learning,” in *Proc. 2009 IEEE Int. Conf. Data Mining Workshops*, 2009, pp. 406–411.

[14] B. Liu, Y. Xiao, and Z. Hao, “A selective multiple instance transfer learning method for text categorization problems,” *Knowl.-Based Syst.*, vol. 141, pp. 178–187, 2018. arning approach,” *Citeseer*, 2000.

[15] R. Said, S. Bechikh, A. Louati, A. Aldaej, and L. B. Said, “Solving combinatorial multi-objective bi-level optimization problems using multiple populations and migration schemes,” *IEEE Access*, vol. 8, pp. 141674–141695, 2020.

[16] M. Gholizade, H. Soltanizadeh, M. Rahmanimanesh, and S. S. Sana, “A review of recent advances and strategies in transfer learning,” *Int. J. Syst. Assur. Eng. Manag.*, pp. 1–40, 2025.