Expert Systems With Applications 250 (2024) 123893

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Review

Exploring Multiple Instance Learning (MIL): A brief survey

Muhammad Waqas [a,b,*], Syed Umaid Ahmed [b], Muhammad Atif Tahir [b], Jia Wu [a], Rizwan Qureshi [c]

[a] Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[b] FAST School of Computing, National University of Computer Emerging Science (FAST-NUCES), Karachi, Pakistan
[c] Center of Regenerative Medicine, Hong Kong Institute of Science and Innovation, Chinese Academy of Science, Hong Kong Science Park, Hong Kong

ARTICLE INFO

Keywords:
Multiple Instance Learning (MIL)
Multi-Instance Learning(MIL)
SUpervised MIL
Unsupervised MIL
Bag and Instance Classification
Review
MIL Applications

ABSTRACT

Multiple Instance Learning (MIL) is a learning paradigm, where training instances are arranged in sets, called bags, and only bag-level labels are available during training. This learning paradigm has been successfully applied in various real-world scenarios, including medical image analysis, object detection, image classification, drug activity prediction, and many others. This survey paper presents a comprehensive analysis of MIL, highlighting its significance, recent advancements, methodologies, applications, and evolving trends across diverse domains. The survey begins by explaining the core principles that form the basis of MIL and how it differs from traditional learning approaches. This sets the foundation for comprehending the distinct challenges and techniques of solving MIL problems. Next, we discuss how supervised learning algorithms are tailored to support MIL and combine this discussion with a review of seminal MIL algorithms as well as the latest innovations that incorporate neural networks, deep learning architectures, and attention techniques. This comprehensive analysis helps to understand the strengths, limitations, and adaptability of these methods across diverse data modalities, complexities, and applications. In summary, this survey paper provides an essential resource for researchers, practitioners, and enthusiasts seeking a comprehensive understanding of Multiple Instance Learning. It covers foundational concepts, traditional methods, recent advancements, and future directions. By providing a holistic view of MIL's dynamic landscape, this paper aims to inspire further innovation and exploration in this ever-evolving field.

1. Introduction

In the last few decades, supervised learning has gained popularity as a method for training predictive models (Nasteski, 2017). This learning approach is designed to create a mapping between input instances (represented as feature vectors) and outputs (represented as labels), based on a labeled training dataset (Saravanan & Sujatha, 2018). The training dataset includes input–output pairs, where each instance is labeled with a specific class. Foulds and Frank (2010). In practice, the instances in the data are always represented as labeled feature vectors in standard supervised formulation (Ahmed, Shuja, & Tahir, 2023). However, real-world data can often take the form of structured objects such as graphs and sets, and labeling these objects can be difficult or even impossible due to time constraints or monetary expenses (Herrera et al., 2016a; Maron & Ratan, 1998). In such scenarios, it becomes crucial for machine learning algorithms to learn from limited labeling information to make the most of the available data (Herrera et al., 2016a).

Multiple Instance Learning (MIL) is a technique that overcomes the limitations of traditional supervised learning when individual instance labels are unavailable (Carbonneau, Cheplygina, Granger, & Gagnon, 2018). It is advantageous when there is weakly labeled data and helps extract meaningful patterns or relationships (Gao et al., 2023; Kamoona, Gostar, Bab-Hadiashar, & Hoseinnezhad, 2023). In MIL settings, data instances are not individually labeled. Instead, the instances are grouped into bags, where each bag can contain a varying number of instances. Labels are only provided based on the bag, while the labels of the individual instances are not available (for both training and testing instances). With these weakly labeled bags, the objective is to learn a mapping function that can map test bags or test instances to labels. Furthermore, the relevant instances in the bags provide vital information regarding the subject of interest and are referred to as witnesses (Carbonneau et al., 2018).

For example, In the problem of drug activity prediction (Dietterich, Lathrop, & Lozano-Pérez, 1997), the objective is to predict the activity

* Corresponding author at: Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.
E-mail addresses: mwaqas@mdanderson.org, waqas.sheikh@nu.edu.pk (M. Waqas), k226020@nu.edu.pk (S.U. Ahmed), atif.tahi@nu.edu.pk (M.A. Tahir), JWu11@mdanderson.org (J. Wu), Rizwan@crmh-cas.org.hk (R. Qureshi).

https://doi.org/10.1016/j.eswa.2024.123893
Received 13 December 2023; Received in revised form 28 March 2024; Accepted 1 April 2024
Available online 8 April 2024
0957-4174/

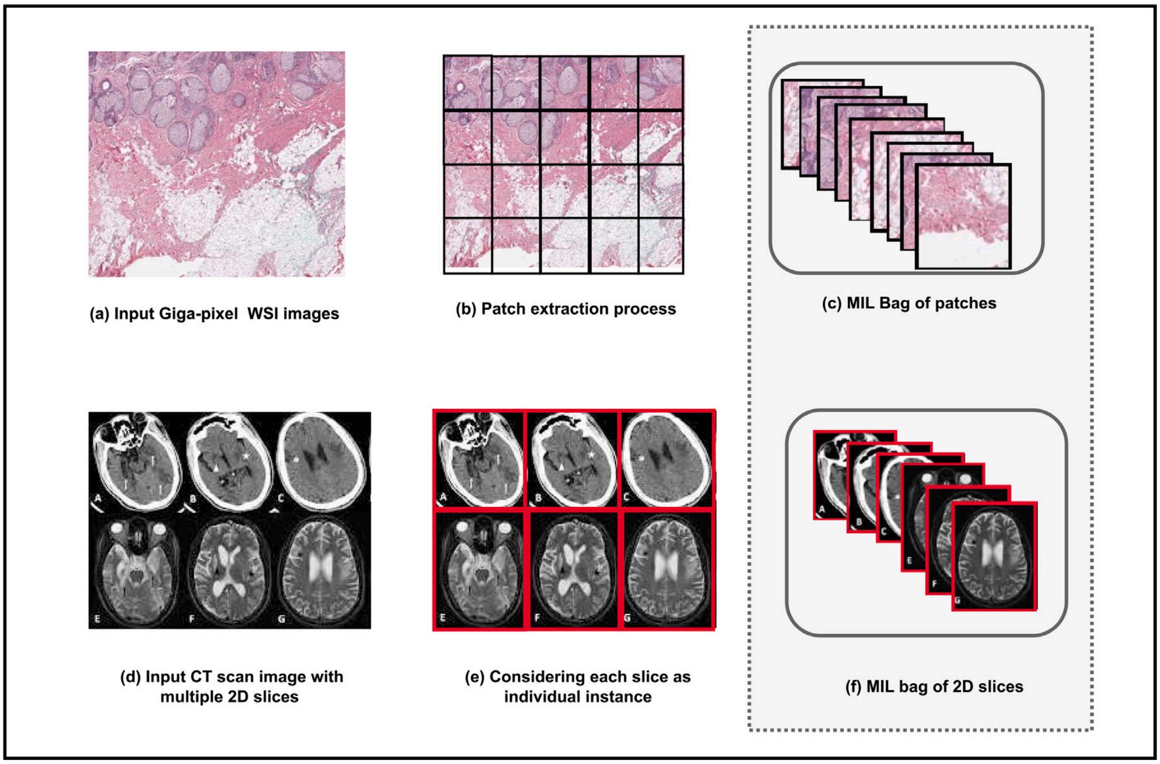

Fig. 1. In complex data learning problems, individual pixels/instances are difficult to label. A group of pixels/instances are extracted and labeled as patch/bag, and the problem is formulated as multi-instance learning.

of molecular binding to a receptor. This binding process depends on the molecular structures, considered feature vectors. Due to the flexible bonds, a molecule can assume several conformations when dissolved in a solution. If at least one low-energy conformation is present, the molecule is considered active and inactive otherwise (Dou et al., 2023; Maser et al., 2023). Therefore, a molecule can represent a collection of conformations with corresponding active or inactive labels. This problem led to the development of the Multiple Instance Learning (MIL) paradigm (Dietterich et al., 1997). Its goal is to learn from bags of instances, each containing many instances, but only a subset is relevant to the task. This learning paradigm eliminates the costs of labeling complex or difficult datasets (Waqas, Tahir and Khan, 2023).

MIL has also significant application in medical image classification and analysis (Fuhrman et al., 2023; Stegmüller, Bozorgtabar, Spahr, & Thiran, 2023), especially in pathology slides or radiological images where it is not feasible to label each pixel, patch, or slice due to the complexity involved and the risk of errors (Yang, Tu, Lei, & Long, 2023). With MIL, learning from bags of image patches or regions becomes possible, enabling the identification of patterns that are indicative of diseases without requiring pixel-level annotations (Manzari, Ahmadabadi, Kashiani, Shokouhi, & Ayatollahi, 2023; Yu et al., 2023). Fig. 1 shows the transformation of a complex learning process from Gigabyte WSI and 3D CT scan image classification problem to MIL.

MIL is also applied to several other real-world applications such as image and video classification (Feng, Xiong, Li, Lang, & Huang, 2014; Maron & Ratan, 1998; Wang et al., 2012) object detection (Manandhar, Morton, Collins, & Torrione, 2012), medical image analysis (Dundar, Krishnapuram, Rao, & Fung, 2006) and web mining tasks (Salton, 1971). Fig. 2 illustrates the difference between traditional supervised learning and multiple-instance learning. Additionally, the relevant instances in the bags provide vital information regarding the subject of interest and are referred to as witnesses. In the context of binary classification, a bag is labeled as positive if it contains one or more examples that are classified as positive. While the negative bags solely consist of negative examples (Zhou, 2018). In binary classification, positive bags typically contain one or more positive instances, but only a small fraction of them are relevant to the task at hand.

This review of Multiple Instance Learning (MIL) is motivated by its importance, methodological diversity, wide-ranging applications, and challenges. MIL encompasses a range of methodologies, primarily involving the customization of supervised learning algorithms. This paper offers valuable insights into their strengths, limitations, and applicability across various domains, including healthcare, anomaly detection, image analysis, and many others (Carbonneau et al., 2018).

This paper aims to comprehensively discuss Multiple Instance Learning (MIL), covering state-of-the-art machine learning and deep learning techniques. The major contribution of the paper and differentiating factors are as follows:

- The paper covers both classical MIL techniques and the latest advancements in deep neural network techniques for MIL. We divide the MIL studies into two major categories: the extension of existing supervised learning algorithms for MIL and the state-of-the-art neural network-based techniques. These algorithms are discussed with their assumptions and characteristics.
- Different instance selection and bag-to-vector transformation techniques are discussed along with their applications in several real-world scenarios.
- The practical applications of multiple instance learning (MIL) in different domains, such as drug activity prediction and image analysis are presented. Furthermore, we provide details and characteristics of state-of-the-art publicly available datasets from various domains, including traditional MIL, medical image classification, object detection, and recently proposed synthetic datasets based on the CIFAR-10 and MNIST.
- We present the limitations and future of extensions of several state-of-the-art MIL algorithms, helping researchers identify promising avenues for exploration.

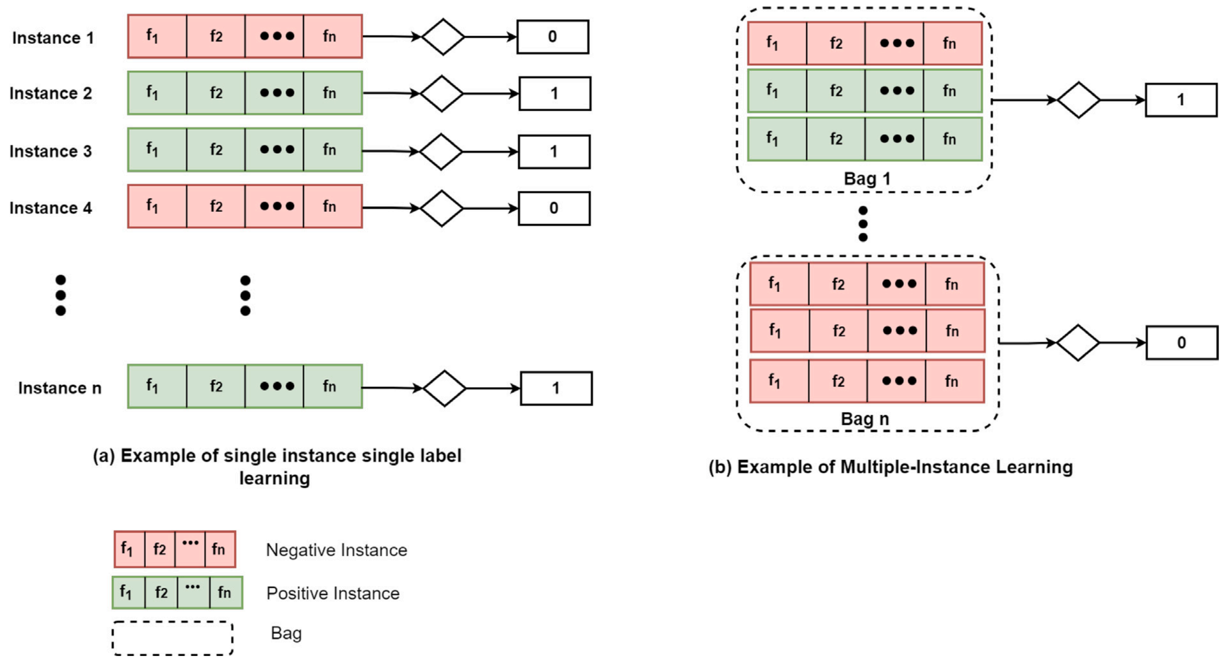

Fig. 2. The example of single instance single label (SISL) and multiple instance learning (MIL).

To the best of our knowledge, this review paper is the first attempt to cover all these different aspects of MIL and provide technical understanding and historical developments into MIL concepts compared to several existing MIL reviews (Amores, 2013; Carbonneau et al., 2018; Fatima, Ali, & Kim, 2023; Quellec, Cazuguel, Cochener, & Lamard, 2017). The rest of the paper is organized as follows:

The MIL classification problems' formulations are discussed in Section 2 and real-world applications of MIL in various domains are discussed in Section 3. The discussion related to conventional machine learning techniques for MIL is presented in Section 4 and instance selection, bag-to-vector transformation techniques, and neural network-based MIL solutions are discussed in Sections 5, 6, and 8, respectively. The publically available datasets related to MIL are presented and discussed in Section 9, followed by open challenges and future directions of the MIL in Section 11. Finally, we conclude the paper in Section 12.

2. MIL problem formulation

In this section, we discuss the granularity of classification and the various assumptions made by MIL algorithms when solving bag classification problems. These assumptions play a vital role in the performance of MIL algorithms.

2.1. MIL classification granularity

Multiple Instance Learning (MIL) problems can be classified into two categories: bag-level or instance-level classification (Carbonneau et al., 2018). In instance-level classification, the model assigns a classification score to each instance and then combines the instance-level information to generate a label for the bag. However, since there are no specific labels for each instance, the accuracy of this method in identifying the bag labels is limited.

There are different approaches to solving the bag-level classification problem. One way is to directly compare the bags using set distance measures (Wang & Zucker, 2000). Another approach is to summarize the contents of the bags into a single feature vector, which transforms the MIL problem into a SISL framework (Wei, Wu, & Zhou, 2016). Alternatively, bag classification can be achieved by categorizing each instance separately and then combining the predictions to determine the label (Ilse, Tomczak, & Welling, 2018; Waqas, Tahir, Khan, 2023; Waqas, Tahir and Qureshi, 2023). Additionally, It is important to note that instance-level and bag-level classifiers differ regarding their misclassification costs since they are trained on different classification methods.

2.2. MIL assumptions for bag-level classification

In MIL, a bag is a hierarchical data object containing sub-objects called instances, while both bags and instances have class labels (Wei & Zhou, 2016). If instance-level labels are given for training bags, the bag-level label can be inferred using supervised learning techniques for instance-level classification. However, the instance-level labels are not provided in MIL applications (Carbonneau et al., 2018). Therefore, The primary goal of the Multiple Instance Learning (MIL) algorithm is to acquire knowledge of a mapping function by utilizing labeled training bags. This acquired knowledge is then utilized to predict the labels of test bags or cases (Zhou, 2004). For binary classification, given a dataset $\mathcal{B} = \left\{\left(\mathrm{B}_1, \mathcal{Y}_1\right), \left(\mathrm{B}_2, \mathcal{Y}_2\right), \left(\mathrm{B}_3, \mathcal{Y}_3\right) \dots \left(B_n, \mathcal{Y}_n\right)\right\}$, where the $B_i = \left\{\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \boldsymbol{x}_{i,3} \dots \boldsymbol{x}_{i,m_i}\right\}$ consists of m_i instances, and $\boldsymbol{x}_{i,j}$ is the j th instance in Bag_i. The instances in Bag_i are associated with a corresponding set of unknown labels $Y_i = \left\{y_{1,i}, y_{2,i}, y_{3,i}, \dots, y_{m,i}\right\} \in \{+1, -1\}$. The label of the Bag_i is given by:

$$\mathcal{Y}_i = \begin{cases} +1 & \text{if } \exists y_{i,j} \in Y_i = +1 \\ -1 & \text{if } \forall y_{i,j} \in Y_i = -1. \end{cases} \tag{1}$$

The MIL assumption in Eq. (1) is also known as the standard MIL assumption. Several MIL methods are based on this assumption (Carbonneau, Granger, Raymond, & Gagnon, 2016; Dietterich et al., 1997, 1997). However, the standard assumption can be revised to address the problem where more than one instance can contribute to the bag label (Carbonneau et al., 2018; Waqas, Tahir, Khan, 2023; Waqas, Tahir, Qureshi, 2023). In this case, several positive instances are necessary to assign a positive label to a bag, known as the non-standard MIL assumption. The non-standard assumption accounts for the contribution of instances in the bag by the interaction, collection, or distribution of the bag's instances. A trivial non-standard MIL assumption is given by:

$$\mathcal{Y}_i = \begin{cases} +1 & \text{if } \mathcal{I} \le \sum_{j=1}^{mi} y_{i,j} \\ -1 & \text{Otherwise,} \end{cases} \tag{2}$$

where $\mathcal{I}$ denotes a minimum number of positive instances in the positive bag.

Furthermore, a collective non-standard assumption of MIL arises when the positive label for the bags is inferred based on several target concepts (Foulds & Frank, 2010). In collective assumption, the positive bag label is triggered if the instances belonging to several concepts are present in the bag. Collective assumption-based methods first define a set of concepts $\mathcal{C}_+$ from positive bags and assign each instance to the concept set. In this case, the bag label is determined by:

$$\mathcal{Y}_i = \begin{cases} +1 & \text{if } \forall c_+ \in \mathcal{C}_+ : \mathcal{I}_{c+} \le \sum_{j=1}^{mi} f_c\left(\mathbf{x}_{i,j}\right) \\ -1 & \text{Otherwise ,} \end{cases} \tag{3}$$

where $\mathcal{I}_{c+}$ denotes the minimum number of target concepts, and $f_c\left(\mathbf{x}_{i,j}\right)$ is instance-wise assignment function which return 1 if $\mathbf{x}_{i,j}$ belongs to concept c_+. A concept can be defined by a single instance or by combining multiple instances. Thus, to incorporate multi-point concepts in MIL an alternative to the instance-wise assignment function is proposed in Yuan, Xu, Zhao, Wen, and Xu (2020), where the given bag is transformed into a vector representation by assigning each instance to the closest concept and converting the MIL problem with collective assumption to the SISL learning problem. Furthermore, the bag structure in MIL can also be perceived as a distribution of instances it contains (Doran & Ray, 2016). In this case, the bag space characterizes all instance space probabilities $P(\mathbf{x}_{i,j} \mid \mathcal{B}_i)$. Hence, the bag classification process involves mapping distribution to bag-level labels.

3. Real-world applications of MIL

Multiple instance learning (MIL) has been successfully applied in various domains, including image classification, object detection, drug discovery, text categorization, remote sensing, sound classification (Briggs, Lakshminarayanan et al., 2012; Choi, Chang, Yang, & Moon, 2024; Hebbar et al., 2021; Wang, Li, & Metze, 2019) and bioinformatics. For example, Molecular activity prediction is performed through MIL by Dietterich et al. (1997). The study aims to predict molecules binding to a musky receptor. As a molecule can adapt to several conformations with varying binding strengths, hence the problem is naturally formulated as MIL. Moreover, the binding strength can be observed for a group of conformations. Following this study, MIL is applied to several biological and drug design problems (Fu et al., 2012; Ray & Craven, 2005b; Srinivasan, Muggleton, King, & Sternberg, 1994). MIL is also applied for image classification task (Chen, Bi, & Wang, 2006; Chen & Wang, 2004; Han & Qi, 2005; Rahmani & Goldman, 2006). The images are represented as bags, and their relevant segments are denoted as instances. MIL algorithms detect local patterns from healthy and unhealthy regions of image or video data collected from patients.

Few of the significant applications include the detection of Tuberculosis (TB) using X-ray (Melendez et al., 2014) and CT scans (Li, Shi, Zhang, Chen, & Zhang, 2015), cancer classification using histopathology images (Su et al., 2022; Yan et al., 2023), diabetic retinopathy screening (Quellec et al., 2012) and identification of vessel diseases (Chen et al., 2015). MIL can be applied in the document, and text categorization (Ray & Craven, 2005b) and clustering (Zhang, Wang, Si, & Li, 2011). The documents in MIL are considered as a bag. However, the short paragraphs and sentences are represented by instances. The objective is to determine whether a document or text references a topic of interest.

3.1. Medical applications of MIL

Medical Imaging and Healthcare procedures rely heavily on complex data to ensure an accurate diagnosis, to effectively design treatment plans, and management of medical conditions (Lee et al., 2017; Prince & Links, 2023). However, obtaining detailed annotations or labels for specific instances within these datasets can be difficult or require significant resources (Nayak & Mishra, 2023). This is where Multiple Instance Learning (MIL) comes in as an invaluable machine learning paradigm. In medical imaging, each bag can represent a medical scan, such as MRI, CT scan, and Whole slide images (WSI), while the instances within the bag might correspond to specific regions or segments within the scan (He et al., 2023; Meng et al., 2023).

In WSI classification, the slides are considered as bags and are available in limited numbers. However, many patches are cropped from each slide to create instances (Liu, Ji, Ye, & Fu, 2024). To tackle a large number of instances, a double-tier MIL framework is developed to deal with a small quantity of bags by using the concept of pseudo bags (Zhang, Meng et al., 2022). It is qualitatively proved that the attention score used for positive region detection under-performs the derived instance probability. The proposed model has achieved state-of-the-art performance over CAMELYON-16 (Bejnordi et al., 2017) and also performed better on the Lung Cancer dataset of TNGA. In another research effort, DeepSMILE (Schirris, Gavves, Nederlof, Horlings, & Teuwen, 2022), a feature variability-aware variant of DeepMIL is used to improve performance and model tumor heterogeneity. It is a Histopathology-specific self-supervised pre-trained feature extractor with VarMIL. It correctly classifies weak labels, and also models intra-tumor heterogeneity by using an aggregation function over the tiles. The genomic tumor is classified by using digitized H&E tissue slides of colorectal cancer tissue (CRC) and breast cancer (BC).

Similarly, a new approach called DAS-MIL is introduced for Whole-Slide Image (WSI) classification in Bontempo, Porrello, Bolelli, Calderara, and Ficarra (2023). DAS-MIL is a multi-scale Multi-Instance Learning (Zhang, Duan et al., 2022) technique that enables information flow across multiple scales in pyramidal structured WSIs through message passing. It also uses a knowledge distillation scheme (Deng et al., 2022) to align latent space representations at different resolutions while maintaining informative diversity. The detailed analysis of the use of Vision Transformer (Dosovitskiy et al., 2020) for medical image analysis and classification in Li, Chen et al. (2023) and Shamshad et al. (2023), These papers introduce attention mechanisms and core components of transformers, followed by a review of transformer architectures specifically designed for medical imaging. They also discuss the limitations of these architectures and address challenges such as diverse learning paradigms, model efficiency, and integration with complementary techniques.

Multiple Instance Learning (MIL) is heavily used to analyze CT (Computed Tomography) scans (Frade et al., 2022; Qi et al., 2021; Waqas, Khan, Anjum, & Tahir, 2020), especially in conditions like COVID-19 from chest CT images (Chen et al., 2022; Meng et al., 2023; Safta & Frigui, 2018). MIL enables the diagnosis of entire scan volumes by considering specific regions within these volumes, eliminating the need for precise annotations for each image segment (Aslani & Jacob, 2023).

Multiple Instance Learning (MIL) has promising MRI applications (Karimi, Dou, Warfield, & Gholipour, 2020). MIL operates by considering the MRI slices or specific regions within the scans as instances within bags or volumes, enabling the classification or detection of abnormalities across the entire MRI volume (Sahiner et al., 2019). This approach proves particularly advantageous when precise pixel-level annotations for each MRI slice are challenging or resource-intensive. MIL has been effectively used for MRI tasks, including lesion detection (Redekop et al., 2022; Vasen et al., 2011), and other disease classification (Tong et al., 2014). MRI analysis aided by MIL enables precise diagnosis without manual annotation, promising efficient medical imaging research and diagnostics (Brendal, 2023).

A recent study (Deng et al., 2024) presents an innovative approach to improving the diagnosis of inflammatory bowel diseases in digital pathology. The proposed method utilizes cross-scale attention scores to create importance maps, which enhance the interpretability and comprehensibility of the CS-MIL model. The study suggests an attention-based early fusion paradigm that could potentially enhance the diagnostic capabilities of inflammatory bowel diseases, making it easier to diagnose and treat. In a recent study (Morales-Álvarez, Schmidt, Hernández-Lobato, & Molina, 2024), a new approach called VGPMIL-PR-I was proposed for identifying prostate cancer using computational pathology. This approach uses a multiple instance learning (MIL) technique and introduces a coupling term, which is influenced by the Ising model, to capture correlations between labels among adjacent patches within whole-slide images (WSIs). Through thorough analysis and visualizations, the study provides valuable insights into the influence of the coupling term. The study in Liu et al. (2024) introduces an adversarial MIL (AdvMIL) framework for conducting survival analysis on gigapixel histological whole-slide images (WSIs). The proposed approach combines adversarial time-to-event modeling (a GAN-based approach) with multiple instance learning (MIL). This approach enables easy upgrades to existing MIL-based methods and improved predictive accuracy while effectively utilizing unlabeled data.

In Zhang, Xu and Liu (2024), a new approach to Multiple Instance Learning (MIL) called "learning from similarity-confidence bags" is proposed. This method uses unbiased estimators to eliminate the need for loss functions or optimizers. It also introduces ways to estimate unknown prior probabilities and provides theoretical guarantees such as an error bound and optimal parameter convergence rate. To prevent overfitting, empirical risk correction techniques are used, ensuring model consistency. For immunotherapy in melanoma treatment and the need for early prognostic biomarkers, the study in Godson et al. (2024) identified immune-related subsets in melanoma patients based on transcriptomic data. The study aims to classify melanoma tumors into immune subtypes using deep learning techniques applied to digitized H&E-stained slides. The study uses multiple MIL frameworks with pre-trained models and self-supervised learning (SSL).

The study in Pérez-Cano et al. (2024) discusses an automated method for diagnosing intracranial hemorrhage (ICH) using computed tomography (CT) scans. The study proposes Multiple Instance Learning (MIL) algorithms and Gaussian Processes (GPs) as classifiers. An end-to-end MIL model is introduced that trains a CNN backbone, an attention mechanism, and a GP classifier for optimal feature extraction. The study demonstrates superior performance compared to existing approaches, especially two-stage GP methods, through experimental validation on two ICH detection datasets. A detailed review specific to digital pathology with a focus on whole slide images is presented in Gadermayr and Tschuchnig (2024). The paper highlights innovative MIL approaches used in digital pathology, remaining challenges, and potential future directions in the field.

For medical image analysis and diagnosis, most of the studies use classification accuracy and AUC as evaluation measures (Waqas et al., 2020; Waqas, Tahir, Khan, 2023; Waqas, Tahir, Qureshi, 2023). Several studies (Provost, Fawcett, Kohavi, et al., 1998; Tax & Duin, 2008) suggested AUC is a suitable evaluation measure compared to accuracy in the case of image classification. Additionally, precision, recall, and F-score are also used in some studies (Ilse et al., 2018; Rymarczyk, Borowa, Tabor, & Zielinski, 2021; Shi, Xing, Xie, Zhang, Cui, & Yang, 2020) to evaluate the model performance.

In summary, Multiple Instance Learning (MIL) is an innovative approach for analyzing vast image datasets in medical imaging (Astorino, Fuduli, Gaudioso, & Vocaturo, 2019; Quellec et al., 2017). It can assess entire image volumes by using specific regions or slices as instances, which simplifies disease diagnosis, anomaly detection, and lesion identification across various imaging modalities (Jiang et al., 2023; Le et al., 2023; Li, Li et al., 2023; Melendez et al., 2014; Quellec, Lamard, Cozic, Coatrieux, & Cazuguel, 2016).

3.2. Object detection and localization in image and video

Multiple Instance Learning (MIL) is an effective technique for object detection. It helps to It can detect objects in images even when the exact location or complete annotations are not available (Wan et al., 2019). The strength of MIL lies in its ability to infer the presence of an object from sets of instances within images (Gao, Wan, Yue, Xu, & Ye, 2022; Huang, Qi, Lu, Zhang, & Ruan, 2017). This eliminates the need for precise object annotations for each instance, making it beneficial for complex scenes or partially occluded objects (Lin, Xu, Yang, & Xu, 2022; Ren, Yang, Zhang and Zhang, 2023). For example, Discrepant Multiple Instance Learning (D-MIL) (Gao et al., 2022) is proposed as an elegant collaboration learning method for object detection through weakly supervised learning. The approach localizes complementary instances by enforcing weakly supervised object detection. Furthermore, object detection is performed with minimum redundancy and maximum completeness using the teachers-student model. The study in Huang et al. (2017) converts saliency detection into a Multiple Instance Learning (MIL) task. Instead of predicting individual superpixels, this technique employs bag-level representations to predict instance labels utilizing the MIL classifier. The paper introduces the MIL optimization mechanism that progressively updates the training bags from simpler to more complex ones. Object Detection has many game-changing applications in the industry. The concept holds significant importance in various domains, including but not limited to autonomous vehicles, surveillance (Ahmed, Affan, Raza and Hashmi, 2022), disaster management (Hanif et al., 2023), and numerous others (Babenko, Yang, & Belongie, 2009).

MIL's capacity to handle scenarios where objects of interest are present but not precisely localized within the image or video frame is particularly advantageous. This application finds utility in many domains, including autonomous driving, surveillance, and medical image analysis. In the context of autonomous driving, the utilization of MIL-based object localization has garnered attention due to its potential to enhance the safety and efficiency of autonomous vehicles. By leveraging MIL techniques, such as Multiple-Instance Support Vector Machines (MI-SVM) and deep learning variants, autonomous vehicles can effectively detect and localize pedestrians, vehicles, and obstacles in real-time from sensor data (Babenko, Yang, & Belongie, 2010; Li & Li, 2020). This precise object localization is pivotal for intelligent decision-making and maneuvering, ultimately contributing to safer navigation in complex urban environments. In this context, MIL assists in the detection and localization of suspicious objects or individuals within video streams, facilitating robust threat assessment and response mechanisms (Khan & Shah, 2006). The MIL framework accommodates the inherent uncertainty in object localization tasks, enabling surveillance systems (Ahmed, Khalid, Affan, Khan, & Ahmad, 2020) to operate effectively in diverse, dynamic, and challenging environments. Additionally, the application of MIL extends to the realm of medical image analysis, where precise object localization is of paramount importance. The diagnosis and treatment planning of various medical conditions often requires the identification and tracking of anomalies within radiological images. By employing MIL, medical practitioners, and researchers can efficiently locate and track regions of interest, enabling accurate diagnosis and personalized treatment strategies (Xu et al., 2014) The utilization of Multiple-Instance Learning for object localization in image and video demonstrates substantial promise across diverse domains. Its adaptability to scenarios where precise localization is challenging and where multiple instances of an object may exist within a single frame underscores its significance in contemporary computer vision applications.

Furthermore, object detection is performed with minimum redundancy and maximum completeness. Therefore, Mean Average Precision (mAP) is used as an evaluation measure (Ahmed, Affan et al., 2022; Gao et al., 2022; Lin et al., 2022; Ren, Yang et al., 2023). Additionally, To assess the accuracy of localization, authors also employ correct localization (CorLoc) as per the PASCAL criterion, where a bounding frame is defined and is deemed positive if it exceeds Intersection over Union (IoU) (Chen et al., 2023; Papandreou, Kokkinos, & Savalle, 2015; Ren, Yang et al., 2023) and localization error (Cinbis, Verbeek, & Schmid, 2016; Zhang, Han, Cheng, & Yang, 2021). Tracking accuracy is used as an evaluation metric in MIL object tracking (Zhang, Meng et al., 2022).

4. Conventional supervised learning-based MIL algorithms

In this section, we present the discussion related to how the conventional machine learning algorithms are customized to support MIL assumptions. The first formulations of the MIL were proposed for drug activity prediction (Dietterich et al., 1997). This work presented an algorithm for predicting molecular activity by learning axis-parallel rectangle (APR). APR involves expanding and shrinking a hyper-rectangle to maximize positive bags' instances and minimize negative ones. The APR provided evidence that other supervised learning algorithms could also be used to solve MIL problems in various other domains. The following subsections introduce extensions of several supervised learning algorithms for MIL.

4.1. Support Vector Machines (SVM) based MIL techniques

Support vector machines (SVM) (Meyer & Wien, 2001) have also been applied to MIL. However, the cost function is customized to accommodate the data in the form of bags. In the case of MIL, SVM optimizes instance-level and bag-level cost functions. For example, the seminal work for the SVM-based MIL technique, the Multi-instance SVM (MI-SVM) (Andrews, Tsochantaridis and Hofmann, 2002) adopts the maximum margin-based approach to estimate the decision boundary to classify the bag while classifying at least one instance from each positive bag as positive and all instance in the negative bag as negative. In the optimization process of MI-SVM, the loss is measured w.r.t bag labels as:

$$\begin{aligned}&\min_{\mathbf{w},b,\{\beta\}} \tfrac{1}{2}\|\mathbf{w}\|^2 + C\textstyle\sum_i \beta_i\\ &\text{s.t.} \underset{1\leqslant i\leqslant n}{\forall} : \mathcal{Y}_i \max_{1\leqslant j\leqslant mi}\left(\left\langle \mathbf{w}, \mathbf{x}_{i,j}\right\rangle + b\right) \geq 1-\beta_i,\end{aligned} \tag{4}$$

where $\beta_i \geq 0$ is the slack variable that allows some instances to be misclassified. In MI-SVM, the maximum function is selected in the constraints. Therefore, it is not convex. The bag classification approach using MI-SVM is shown in Fig. 3(a).

In an alternative approach, the mi-SVM (Andrews, Tsochantaridis et al., 2002) assigns bag-level labels to instances and uses a mixed-integer optimization process. This method performs instance-level classification and treats instance labels as unobserved hidden variables. The objective of mi-SVM is to maximize the margin over unknown instance-level labels as:

$$\begin{aligned}&\min_{\{y_i\}}\min_{\mathbf{w},b,\beta} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \beta_i\\ &\text{s.t.} \underset{1\leqslant i\leqslant n}{\forall}\ \underset{1\leqslant j\leqslant mi}{\forall}\ y_{i,j}\left(\left\langle \mathbf{w}, \mathbf{x}_{i,j}\right\rangle + b\right) \geq 1-\beta_{i,j}, \beta_{i,j} \geq 0.\end{aligned} \tag{5}$$

mi-SVM following the approach pursued in Joachims et al. (1999) transductive inferences. Since mi-SVM enforces the notion that every instance must have a label, the resulting hyperplanes are different for MI-SVM and mi-SVM. The classification process using mi-SVM is shown in Fig. 3(b).

The multiple-instance learning with semi-supervised SVM (MissSVM) (Zhou & Xu, 2007)treats MI datasets as semi-supervised learning problems with unknown labels for positive bag instances (Andrews, Tsochantaridis et al., 2002). In essence, MissSVM's optimization process is similar to MI-SVM (Andrews, Tsochantaridis et al., 2002), with two additional constraints on positive bag instances' by using different slack variables. The additional constraints for negative bag instances and positive bag instances were added to the objective function to prevent positive bag instances from carrying positive or negative labels.

To address non-convex optimization problems, the concept of deterministic annealing (DA) (Gehler & Chapelle, 2007) is combined with support vector machines for MIL applications in Gehler and Chapelle (2007). The algorithm proposed in this study is named the deterministic annealing for identifying all labels (AL-SVM). The study proposes a new objective function to locate better local minima in the optimization process. This is done to handle uncertainty in the labeling of instances within bags. The DA approach addresses this challenge by formulating the problem as an optimization problem. To find the optimal class labels for the instances within the bags, it uses a form of simulated annealing (Henderson, Jacobson, & Johnson, 2003).

Recent implementations of MIL-based SVM include primal–dual multi-instance support vector machine (SVM) designed for efficient handling of large-scale Multi-Instance Learning (MIL) datasets (Brand et al., 2023). This approach utilizes a multi-block variation of the alternating direction method of multipliers, bypassing iterative quadratic programming problems typical in traditional SVM-based MIL algorithms. It resolves scalability issues by avoiding computationally intensive optimizations, includes enhancements to accommodate a large number of features and bags, and extends to learn nonlinear decision boundaries through kernel extensions. In another recent study (Huang, Liu, Jin, & Mu, 2022), researchers proposed a Bag Dissimilarity Regularized (BDR) framework for Multi-Instance Learning (MIL). This framework combines both implicit and explicit bag representations. It includes a regularization term that captures intrinsic geometric information from bag dissimilarities and an efficient method for explicit bag embedding. The researchers developed two BDR methods based on support vector machines and broad learning systems to minimize computational overhead.

The concepts of hard margin loss (MIHMSV), Ramp loss (MIRLSVM), and Hinge loss (MIHLSVM) are investigated in Brooks (2011). The paper proposes three integer programs and two constraint programs and compares their performance regarding the time required to reach an optimal solution. The paper also presents a heuristic approach capable of handling large problems. The proposed SVM-based MIL approaches are able to obtain comparable performance to other methods. However, it is important to note that the study primarily aims to reduce computation time rather than improve performance. In addition, the performances of the proposed techniques are not fairly compared. These results are obtained using leave-one-out cross-validation rather than repeating tenfold cross-validation several times, as is done in other MIL studies (Leistner, Saffari, & Bischof, 2010; Vezhnevets & Buhmann, 2010; Zhou, Sun, & Li, 2009). Table 1 discusses a few other SVM-based MIL approaches. These algorithms mostly operate on instance-level classification. A new approach called Simultaneous Imputation Multi-Instance Support Vector Machine (SI-MISVM) has been proposed in Brand, Baker, and Wang (2021). This approach combines techniques from multi-instance learning and matrix completion to handle missing data and predict outcomes over time. It uses Primal SVM to identify biomarkers related to COVID-19.

A support vector machine-based approach for the detection of breast cancer is proposed in Seo, Brand, Barco, and Wang (2022). The authors proposed an innovative algorithm called Primal-Dual Multi-Instance SVM (pdMISVM) to effectively address the optimization difficulties caused by coupled primal variables. The proposed method involves using the Alternating Direction Method of Multipliers (ADMM) to split the primal variables followed by the integration of constraints to facilitate optimization. The algorithm updates primal and dual variables iteratively until it reaches a predefined tolerance. The study also discusses class-hyperplane updates and avoiding least-squares calculations through gradient descent. The concept of a Fast Multi-Modal Support Vector Machine (FMMSVM) is introduced in Seo and Wang (2023). The proposed approach enables the simultaneous imputation of missing modalities and the prediction of clinical outcomes. The optimization

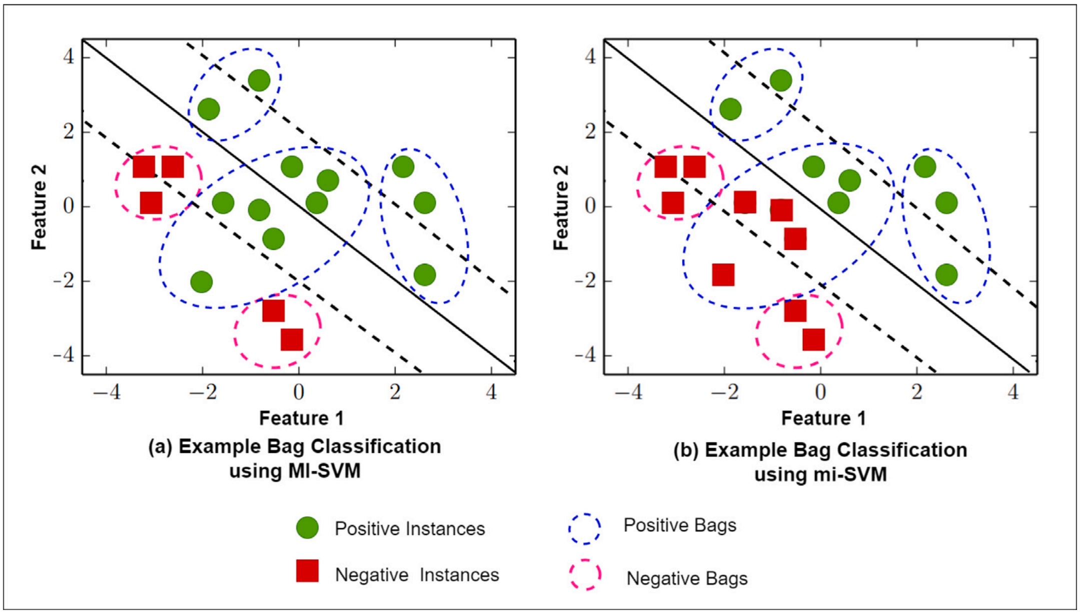

Fig. 3. The example of a hyperplane found by MI-SVM is illustrated in (a), and the obtained hyperplane using mi-SVM is shown in (b). MI-SVM performs bag-level classification, where all the instances of the bag are labeled positive if at least one instance in the bag is classified as positive. On the other hand, mi-SVM assumes that each instance can be positive or negative and imputes all the instances labels in the training process, and then treats the MIL task as a supervised SVM task.

algorithm utilized is based on a Support Vector Machine (SVM) and scales linearly with the number of input data features. This guarantees efficiency when dealing with large datasets. A kernelized version of the semiproximal SVM algorithm for multiple instance learning (MIL) is proposed in Avolio and Fuduli (2023). The proposed method combines Support Vector Machines (SVM) with the efficiency of proximal SVM (PSVM) techniques, mi-KSPSVM provides an enhanced solution for this task. The algorithm employs the Radial Basis Function (RBF) kernel to govern the trade-off between margin maximization and error minimization. Traditional multiple instance learning (MIL) algorithms face issues with scalability when it comes to handling large datasets. To address this issue, a primal–dual multi-instance support vector machine (pdMISVM) has been proposed in a research paper (Brand, Seo, Baker, Ellefsen, Sargent, & Wang, 2024). This algorithm uses a multi-block variation of the alternating direction method of multipliers (ADMM) to efficiently handle large-scale data without the need for iterative solutions of quadratic programming problems, which is common in SVM-based MIL algorithms. Furthermore, the research paper extends the method to handle nonlinear decision boundaries through kernelization

Current MIL algorithms may experience reduced classification performance due to limited use of MIL metadata and assumptions about data distribution (Andrews, Tsochantaridis et al., 2002; Zhou & Xu, 2007), a tackle this limitation, a novel Double Similarities Kernalized Weighted Multi-instance Learning (DSKMIL) is proposed in Zhang, Wu et al. (2024). This takes into account both bag and instance information, incorporating intra-bag instance relationships, instance-to-bag importance, and bag-to-bag similarity. Unlike other models, DSKMIL does not rely on assumptions about the relationship between bag and instance labels. The designed kernel function supports vector machines (SVMs) in MIL tasks and delivers high-quality classification performance.

Table 1
Description of SVM-based MIL techniques.

Algorithms	Description
Generalized support vector machine (Andrews, Hofmann and Tsochantaridis, 2002)	The mixed integer program is approximated using a heuristic approach with a re-labeling approach
Semi-proximal support vector machine (Avolio & Fuduli, 2020)	Generates proximal hyperplanes in the middle of two parallel planes to cluster positive bag instances. The second hyperplane serves as a supporting hyperplane for the −ve bags
Robust support vector machine for MIL (R-SVM) (Poursaeidi & Kundakcioglu, 2014)	a three-phase heuristic algorithm is proposed and different loss functions are empirically evaluated in the context of MI classification, including hinge, ramp, and hard margin loss.
Kernel SVM for MIL (Zhang, Platt, & Viola, 2005)	Employs learning of convex combination of multiple kernels during the optimization process

4.2. Density-based MIL techniques

The Diverse Density (DD) procedure was developed as the first probabilistic model for MIL (Maron & Lozano-Pérez, 1997). This is a probabilistic technique, where the diversity density of a point in feature space measures "how far the negative instances are from this point and how many different positive bags are near to that point". The concept of selecting diverse instances within bags to ensure comprehensive representation is discussed in Amores (2013). Including instances that contribute most to the bag's classification enhances the robustness and generalization of MIL models. Andrews, Tsochantaridis et al. (2002) proposed ensemble methods that leverage diverse subsets of training data or algorithms to improve model performance.

This approach maximizes the DD measure across all feature space to retrieve the true concept. It assumes that the probability of a bag-level class can be affected by a single instance in a bag. Thus, the bag-level construction is represented by two different models noisy-or and most-likely-cause model. The first model considers that the positive bag labels can be attributed to each individual instance, while the second model assumes only a single representative instance for each bag. With gradient ascent, the DD algorithm looks for the point that maximizes the diverse density function from multiple starting points (e.g., starting from each positive bag). A clustering-based method is proposed in Maron and Lozano-Pérez (1997) to understand the underlying density and diversity of the instances in the bag. Carbonneau et al.

(2016) proposed bag-level density estimation techniques that compute the density of positive instances in multiple subspaces using K-means algorithms. The concept of subspace clustering is further investigated in other studies (Waqas, Tahir, Khan, 2023; Waqas, Tahir, & Qureshi, 2021; Waqas, Tahir, Qureshi, 2023), where the distribution of instances is determined using GMM and FCM clustering. By understanding the underlying density, models can focus on regions of higher importance or confidence within bags, which contributes to improved decision-making. These methods use clustering techniques to identify clusters or subgroups within bags, which allows for a more detailed analysis of instances that share similar characteristics or importance. This leads to better bag-level representations. The supervised version of Kernel density estimation (SKDE) is introduced as an extension of the unsupervised KDE framework in Du, Wu, He, and Yang (2013). SKDE incorporates class labels to identify modes effectively, while a supervised mean shift locates SKDE modes. This method offers better multi-modal concept learning and noise robustness compared to DDE. Additionally, the paper proposes the use of bag mapping into a concept space and multi-class SVM classifiers.

The approach of DD is extended with expectation maximization (EM-DD) in Zhang and Goldman (2001). This method combines the most likely cause model of DD with EM. The maximization function causes a non-differentiable optimization problem in the DD algorithm with the most likely cause model. To avoid such a non-differentiable problem, EMDD aims to solve the maximum function in the expectation phase (E-step). The E-step in EMDD selects a single representative instance from the bags for the initial hypothesis space. For the Maximization step, EMDD employs the most likely-cause model of the DD algorithm to maximize the likelihood. The fundamental concept behind the DD method is to locate a candidate point with a high diversity density in the feature space.

Deep learning methods (Ilse et al., 2018; Shi et al., 2020), with attention mechanisms, handle structured MIL data well. However, they do not model instance labels, limiting the ability to quantify instance-level uncertainty. Probabilistic MIL methods, like Gaussian Process GPs, can address this issue. Therefore, a new probabilistic MIL model called PG-VGPMIL is proposed in Castro-Macías, Morales-Álvarez, Wu, Molina, and Katsaggelos (2024). This model is based on the Pólya-Gamma random variables and is equivalent to the existing VGPMIL model. This equivalence helps to develop a general inference framework for the logistic observation model called VGPMIL and can accommodate various Generalized Scale Mixture (GSM) densities. The paper also introduces GVGPMIL, which is a particular realization of VGPMIL using the Gamma distribution.

4.3. Decision tree-based MIL techniques

In Blockeel, Page, and Srinivasan (2005), a decision tree (DT) approach is proposed for MIL settings. The algorithm uses the best-first node expansion technique during the tree construction process. For multi-instance learning, the algorithm gives more weight to examples from smaller bags and focuses on pure positive leaves. Furthermore, Several MIL-specific heuristics are incorporated in splitting decision tree nodes. Several other extensions of classical tree learning algorithms are presented, e.g., multiple-decision trees (ID3-MI) and multiple-decision rules (Ripper- Mi) are proposed in Chevaleyre and Zucker (2001). The concept of random forest for MIL is proposed in Leistner et al. (2010). The proposed approach initially considers class labels of instances as random variables and optimizes class labels during the training process by using the proposed iterative homotopy technique. The concept of random forest for binary MIL classification problems is extended in Leistner et al. (2010). The proposed algorithm focuses on model interpretability using proposed instance selection and randomized tree construction (ISRT). This algorithm aims to train a selector to choose instances of positive bags that contribute the most to the positive bag. For each positive bag, the selector assigns a high value to at least one instance in the positive bag and a low value to all instances in the negative bag by applying an SVM-based optimization approach similar to MI-SVM (Andrews, Tsochantaridis et al., 2002). Finally, based on the selected instances, ISRT builds a random forest by searching for the best splitting parameter that is performed using standard purity measures. The process allows discriminatory information to be extracted from both instances and bags.

4.4. Distance-based MIL approaches

Conventional lazy learning involves selecting the closest neighbors by calculating the Euclidean distance between them, such as the KNN approach. The most frequent class label of the nearest neighbors is then predicted as the instance unknown class label. In this type of learning, the distance measure is used to make a decision. Since the labels of the individual instances are unknown, a standard distance measure such as the Euclidean distance does not apply to MIL problems. Therefore, it is necessary to enhance the distance function.

To compute the distance between the two bags based on the standard KNN approach, Wang and Zucker (2000) proposed a modified Hausdorff Distance-based MIL approach called Citation-KNN. This approach computes Hausdorff's distance to find references and citer bags. Citation KNN assumes that the closest neighbors of one bag may not be enough to solve MI problems. Therefore, the results are derived from combining the results of citer bags.

The nearest distribution method for MIL is proposed in Xu (2001). The proposed algorithm estimates the Gaussian distribution for each bag feature based on two major assumptions. First, all data dimensions are equally relevant to the classification. Second, each bag contains enough instances. Once the distribution is derived, the original data is discarded, and the distance between the training and testing bag distributions is computed using Kullback–Leibler (KL) distance (Kullback & Leibler, 1951), also known as relative entropy. Afterward, the classification process is carried out as in a standard KNN algorithm.

In Auer and Ortner (2004), the authors present an optimal ball method for MIL is proposed. This method is inspired by APR (Dietterich et al., 1997) and works as a weak learner for the boosting. The algorithm locates a circular area (optimal ball) in the feature space by finding the Euclidian distance between each positive bag instance and all other positive bags. The circle separates the positive and negative concepts such that all negative bags are located outside the ball. Based on the accuracy evaluation, the circle's radius is maximized by placing an instance from the positive bag at the center. A new bag is classified as negative if every instance is outside the optimal circle.

The concept of similar and dissimilar bags instead of relying on fully labeled bags in MIL is proposed in Feng et al. (2021). the authors propose a convex formulation for training a bag-level classifier through empirical risk minimization, along with deriving a theoretical generalization error bound. Additionally, they introduce a robust baseline method for training an instance-level classifier by minimizing the instance-level empirical risk. The concept of utilizing triplet comparison bags is introduced in Shu et al. (2024), the proposed method uses triplet comparison data to train a bag-level classifier using empirical risk minimization. This reduces the cost and improves accuracy compared to traditional methods that require fully labeled bags. The article provides theoretical analysis and extensive experiments demonstrating the effectiveness of the proposed method.

5. Instance selection techniques in MIL

The processing and classification of intricate bag representations lead to a complex hypothesis space (Wei et al., 2016). Thus, Multi-instance learning (MIL) problems are often converted into supervised learning (single instance single label learning) to tackle hypothesis space complexity. A straightforward approach propagates bag-level labels to instances inside the bag (Andrews, Tsochantaridis et al.,

Table 2
The details of instance selection and bag-to-vector transformation algorithms.

Algorithms	Category	MIL assumption	Classification granularity		Description
			Bag level	Instance level	
MILES	Instance selection and bag-to-vector transformation	Non-standard	✓		It uses bag embedding and instance similarity measurements to map bags to the new feature space and removes unnecessary instances sequentially
RSIS	Instance selection	Standard		✓	Estimates an instance's positive score via subspace clustering. Selects a single sample instance from each +ve bag.
MI-Wrapper	Bag-to-vector transformation	Non-standard		✓	Simplify the MIL task by assigning bag labels to the instances and Computing weights to each instance in the bag in proportion to the bag size.
Simple-MI	Bag-to-vector transformation	Non-standard	✓		Summarizes the bag of instances with the mean vector of instances.
MILMPC	Instance selection and bag-to-vector transformation	Non-standard	✓		Select instances by considering similar instances as a multi-point concept based on the nearest neighbor of instances in positive bags.
MI-Graph	Bag-to-graph kernel transformation	Non-standard	✓		Consider the instances as non-i.i.d and adopt a graph kernel learning technique.
miVLAD	Bag-to-vector transformation	Non-standard	✓		Transforms bag into a high-dimensional vector representation based on the K-means dictionary learning.
miFV	Bag-to-vector transformation	Non-standard	✓		Produces a high dimensional vector for the bag with higher-order(HO) stats of the soft assignment method.
DGMIR-FV	Bag-to-vector transformation	Non-standard	✓		Uses an estimation network for instance selection and Fisher vector encoding for bag label prediction.
FCBE-miFV	Bag-to-vector transformation	Non-standard	✓		Uses fuzzy subspace clustering & ensemble-based Fisher vector (FV) encoding to solve MIL problem
RMI-FV	Bag-to-vector transformation	Non-standard	✓		Adopts is stacking ensemble architecture with instance relevance estimation for MIL.

2002) to learn a bag-level propositional classifier. However, assigning bag labels to instances can result in incorrect class labels for the negative instances in the positive bag. An alternative is to generate a bag representation using its statistic properties of instances. In Dong (2006), three bag representation techniques are proposed, including instance-wise min–max, geometric mean, and arithmetic mean pooling operations. These bag representation techniques can achieve comparable performance but discard essential instance-level information in the bag.

The utilization of instance selection can be advantageous in mitigating the complexity associated with the learning process (Carbonneau et al., 2016; Waqas, Tahir, Khan, 2023; Waqas, Tahir, Qureshi, 2023). The aforementioned procedure involves the identification and inclusion of significant examples from a collection, which are then integrated into the classification process.

For example, multi-instance embedded instance selection (MILES) (Chen et al., 2006) maps the bags to the new feature space using bag embeddings and instance similarity measures. MILES initially considers all the training instances relevant and iteratively removes redundant instances using L_1-SVM regularization and performs classification and instance selection simultaneously. As a result, the multiple-instance learning problem is transformed into a standard supervised learning problem that does not assume instance labels are associated with bag labels.

Random subspace instance selection approach (RSIS) (Carbonneau et al., 2016) performs instance selection based on a subspace clustering approach to estimate the positive score of an instance. This approach selects a single representative instance from each positive bag and trains a pool of classifiers to perform instance-level classification. RSIS trains several classifiers by adopting this guided random instance selection process, and the learners' output is combined through average voting criterion.

The instance selection approach in Waqas et al. (2021) and Waqas, Tahir, Qureshi (2023) proposed instance distribution learning techniques based on traditional Gaussian Mixture Model (GMM) and deep GMM. These techniques are robust enough to identify the essential instances from each bag. The concept of instance relevance estimation is proposed in Waqas, Tahir, Khan (2023), where one or more relevant instances from each bag are selected based on the Fuzzy C Mean (FCM) clustering process. These techniques are based on the non-standard assumption of MIL. The instance selection techniques (Waqas, Tahir, Khan, 2023; Waqas et al., 2021; Waqas, Tahir, Qureshi, 2023) are based on the subspace clustering techniques inspired by RSIS (Carbonneau et al., 2016) and are more suitable for critical decision-making applications (Ding et al., 2020). Therefore, the training process of RSIS is illustrated in Fig. 4.

A deep learning-based approach called deep multi-instance selection (DMIS) (Li, Zhan, Yang, & Shi, 2021) uses Gumbel Softmax or Gumbel Top-K (Jang, Gu, & Poole, 2016) to automatically identify regions of interest (ROI) in an end-to-end process. This approach effectively reduces interference from redundant instances during predictions. Both theoretical analysis and empirical investigations have demonstrated that DMIS outperforms classical MIL methods in terms of generalization ability, ROI positioning, and comprehensibility on both synthetic and real-world datasets.

6. Bag-to-vector transformation techniques for MIL

A simple solution to hypothesis space's complexity of MIL problems is to assign labels at the bag level to each instance and train a classifier based on the modified data. Simple-MI (Amores, 2013) is an approach

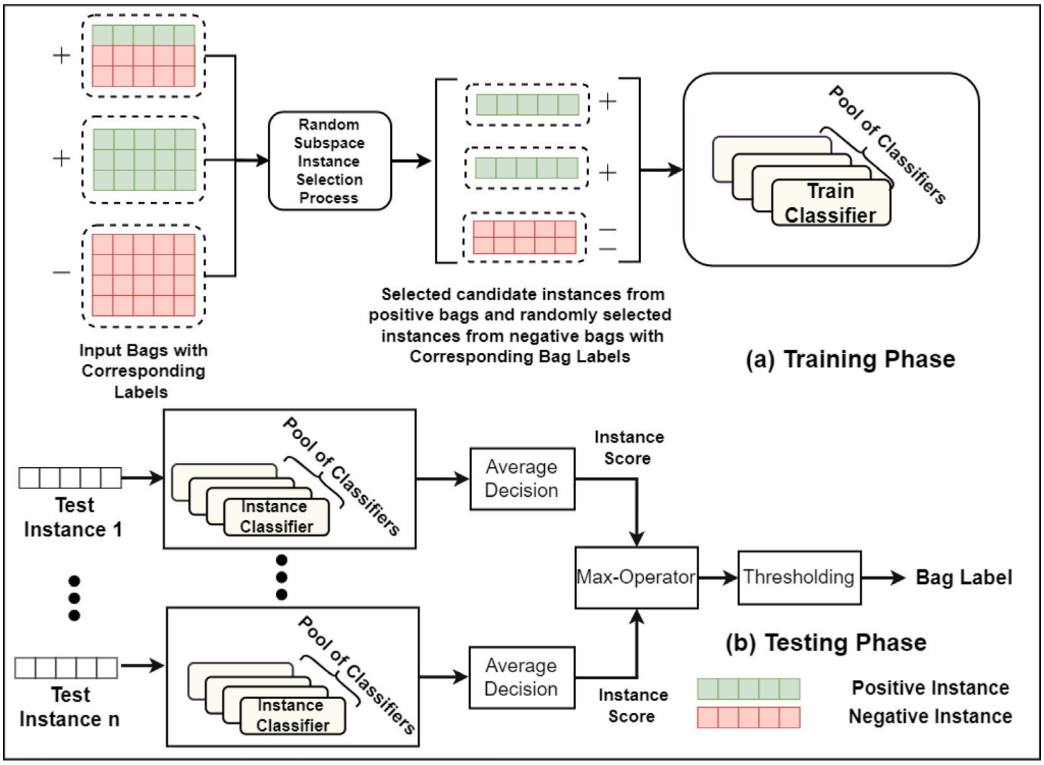

Fig. 4. The instance selection process adopted in the training process of RSIS (Carbonneau et al., 2016). (a) shows the algorithm's training phase, where a candidate instance from each positive bag is selected. While negative instances are randomly sampled from negative bags. An instance-level classifier is trained with the selected instances and bag–level labels in each instance selection iteration. In the testing phase (b), the pool of trained classifiers generates a prediction for each bag instance, and predictions are combined through the average voting process. The instances in the bags are shown in dotted squares.

to simplifying the MIL problem to a standard single instance single learning problem. Simple-MI is the baseline MIL method that summarizes the bag of instances with the mean vector of instances. A classifier is trained using the summarized bag representation vector, which inherits the bag label. In the testing phase, the test bags are also summarized similarly. A similar approach is adopted in MI-Wrapper (Frank & Xu, 2003) by discarding the standard MIL assumption and replacing it with the collective assumption. To ensure that all bags are treated equally, the algorithm assigns weights to each instance in the bag in weighted proportion to the size of the bag. This technique provides a fast solution to the MIL task with comparable performance to several algorithms (Andrews, Tsochantaridis et al., 2002; Hong et al., 2013; Leistner et al., 2010).

In Yuan et al. (2020), the authors proposed a joint approach for multipoint-based instance selection bag embedding MIL (MILMPC). The proposed approach selects instances by considering a set of similar instances as a multi-point concept. MILMPC generates a set of candidate concepts based on the nearest neighbor of instances in all positive bags. With the selected multi-point concept, the bag is transformed into a vector representation by concatenating the vectors with minimum distance operation. Zhou et al. (2009) adopt a graph kernel learning technique to transform a given bag into an undirected weighted graph (Mi-Graph). The nodes in the generated graph represent instances of the bag, and if the distance between the two nodes is smaller than a preset threshold, then a weighted edge is established between the nodes. The weight of the edge expresses the affinity of the two nodes. This approach solves the MIL problem using the kernel learning approach, which is useful where details of the bag structure play an essential role in the bag classification process (Dietterich et al., 1997).

The bag encoding techniques are also helpful in reducing the hypothesis-space complication by transforming bags of multiple instances into a high-dimensional vector representation. For example, a multiple-instance vector of locally aggregated descriptors (miVLAD) (Wei et al., 2016) encodes the given bag to a high dimensional vector representation. This method is based on codebook learning using a clustering process. After that, miVLAD (Wei et al., 2016) maps bag to a vector representation using the generated codebook and accumulating the difference between the instance and the centroid it is assigned to. The bag-to-vector generation process used in miVLAD is presented in Fig. 5, where 5(a) illustrated the instance clustering process, 5(b) shows the process of difference computation between clusters and corresponding labels, and vector generation process for each cluster center is illustrated 5(c) respectively.

An alternative encoding scheme is multiple-instance Fisher vector encoding (miFV) (Wei et al., 2016). It approximates instance distribution using the Gaussian mixture model (GMM), essentially a soft assignment of instances to K clusters using GMM clustering. miFV describes the bag of a variable number of instances to a gradient vector representation with respect to the GMM parameters. Due to the robustness of the FV encoding techniques, it is further utilized with ensemble-based bag classification approaches in Waqas, Tahir, Khan (2023), Waqas et al. (2021) and Waqas, Tahir, Qureshi (2023). The study in Zhang, Xu et al. (2024) investigates instance aggregation and classification based on instance importance. It uses a weighted average aggregation method to identify instances that contain valuable information, which leads to better interpretability and classification performance. The proposed approach offers a transparent approach to MIL that aligns with human-like analysis. The details of several instance selection and bag-to-vector transformations are discussed in Table 2.

7. Applications of instance selection and Bag-to-vector transformation techniques

The instance selection and Bag-to-vector transformation techniques play a critical role in Multiple Instance Learning (MIL). These techniques are to deal with the complication of hypothesis space, which

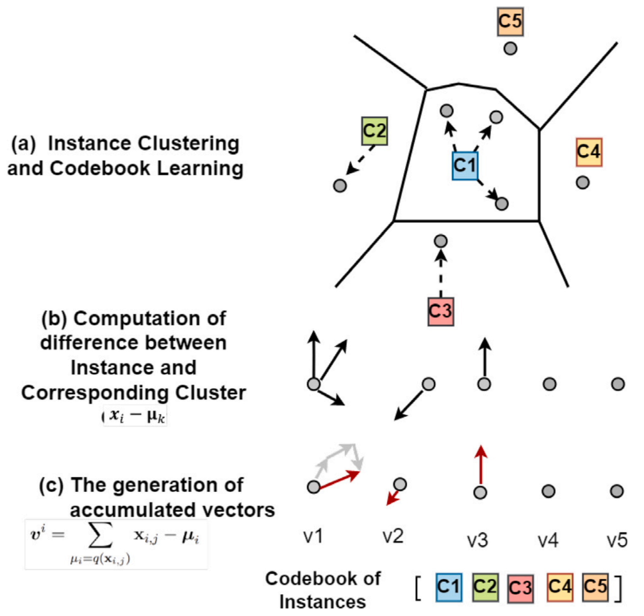

Fig. 5. The working of miVLAD (Wei et al., 2016). The solid arrows represent the vector representation of instances.

is the major problem while designing MIL algorithms (Waqas, Tahir, Khan, 2023; Wei et al., 2016). since they aid in the identification of the most informative instances in each bag (Chen et al., 2006). Several strategies exist for instance selection, including hard and soft clustering-based techniques (Carbonneau et al., 2016; Waqas, Tahir, Qureshi, 2023), which select instances with the highest confidence scores, and Density, which prioritizes instances representatives within the bag.

The instance selection methods are used for various real-world applications, particularly in medical image classification (Quellec et al., 2017). For example, large whole slide images (WSI) or 3D CT scans in medical imaging are typically divided into small patches and treated as instances. However, instance selection techniques can help to identify the important regions of interest (ROI) in the image instead of using all patches, which can aid in detecting metastatic breast cancer (Tang et al., 2023). This approach can make it easier to choose informative image patches as instances, allowing MIL models to effectively distinguish between WSIs from patients with metastatic breast cancer (positive bags) and those without (negative bags) (Yan et al., 2023). The instance selection process is applied to other MIL applications such as image classification (Fu, Robles-Kelly, & Zhou, 2010), object localization, sound classification (Choi et al., 2024; Krishna, Bhattu, Somayajulu, Kumar, & Reddy, 2022; Pereira & Maia, 2024; Ren, Yang et al., 2023), text and sentiment classification (Li et al., 2021).

Bag-to-vector transformation techniques are commonly used in several real-world tasks, such as image classification, document classification, and drug response prediction (Wang & Zucker, 2000; Waqas et al., 2021; Wei, Wu, & Zhou, 2014; Wei et al., 2016). These techniques address the complexity of the hypothesis space by transforming the entire bag into a vector representation. This representation process often involves clustering and instance scoring processes. Conventional clustering-based bag-to-vector transformation algorithms require less computational overhead, making them well-suited for applications with small to moderate-sized datasets (Waqas, Tahir, Khan, 2023; Waqas et al., 2021; Wei et al., 2014, 2016). However, state-of-the-art bag-to-vector transformation techniques based on neural networks and deep learning have also been proposed for applications involving larger datasets, such as video classification (Ren, Yang et al., 2023) and disease diagnosis through large-scale medical imaging (WSI and CT) (Shao et al., 2023). These advanced techniques are discussed in the following sections. In some cases, the combination of instance selection and bag-to-vector transformation techniques are also proposed for moderate-size datasets (Waqas, Tahir, Khan, 2023; Waqas, Tahir, Qureshi, 2023).

8. Neural network based MIL solutions

Before the rise of deep neural networks (DNN), the use of a regular neural network is proposed for MIL in Ramon and De Raedt (2000). This approach is called a multi-instance neural network (MINN). The instance-level classification is adopted in this approach, and bag labels are computed by aggregating the instance-level information. This idea provided the foundation to employ neural networks in the domain of MIL. However, at that time, the neural networks and deep learning techniques (Ahmed, Ahmad, Shuja and Affan, 2022) were not the algorithms of choice due to data and hardware limitations. Deep learning has achieved a breakthrough in artificial intelligence and machine learning in recent years and has been employed in a range of applications, including computer vision (Han et al., 2022), speech recognition (Jolad & Khanai, 2022), and many others (Adate & Tripathy, 2022). Popular neural architectures such as ResNet (He, Zhang, Ren, & Sun, 2016), GoogleNet (Szegedy et al., 2015), and AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) have become famous for solving diverse machine learning challenges. Thus, the deep learning-based approaches are also applied to several MIL applications (Carbonneau et al., 2018). Traditional neural network-based MIL techniques usually utilize the instance representation as a feature given (Ramon & De Raedt, 2000). However, recent research also used convolution layers to extract the features (Miech et al., 2020). MIL poses a unique challenge of varying the number of unlabeled instances in the bag. Neural network-based MIL techniques tackle this challenge using pooling operations. These pooling operations are discussed in detail in the next subsections. The details of neural network-based MIL techniques are given in Table 3.

8.1. MIL pooling-based neural network

The pooling operation can be used to generate a vector representation for the bag at the instance level. This vector representation can then be given as input to the classification layers of the network. Alternatively, an instance-level classification model can be trained to obtain instance-level class scores, which can then be pooled to obtain the bag-level label. A generic instance pooling process for MIL bag classification is shown in Fig. 6(a), and the process of instance score pooling is illustrated in Fig. 6(b). When dealing with bag-level classification problems, the instance-wise MIL pooling techniques are robust. However, this approach is less interpretable. The interpretability of the model is important for critical applications, such as computer-aided diagnosis (Carbonneau et al., 2018).

The concept of pooling in MIL is revised in Wang, Yan, Tang, Bai and Liu (2018), where neural networks are evaluated as solutions to MIL problems using a variety of pooling operations. The study investigates instance and bag-level prediction techniques using two proposed neural network architectures including (mi-Net and MI-Net). The MI-net is a favorable option for bag-level classification, as it utilizes the pooling of instance embedding to generate scores to derive the ultimate bag-level score. Similarly, mi-Net pools instance to generate The neural network designs that have been suggested can conduct end-to-end learning for both instance-level and bag-level classifications. Additionally, the study proposed residual connections (He et al., 2016) and deep

Yan et al. (2018) proposed a dynamic pooling operator-based dynamic routing capsule network (DP-MINN). This approach incorporates the contextual information in the bag and uses an iterative average update technique to pool the instances (Choi, Seo, Im, & Kang, 2019; Peer, Stabinger, & Rodriguez-Sanchez, 2018). Unlike simple neurons

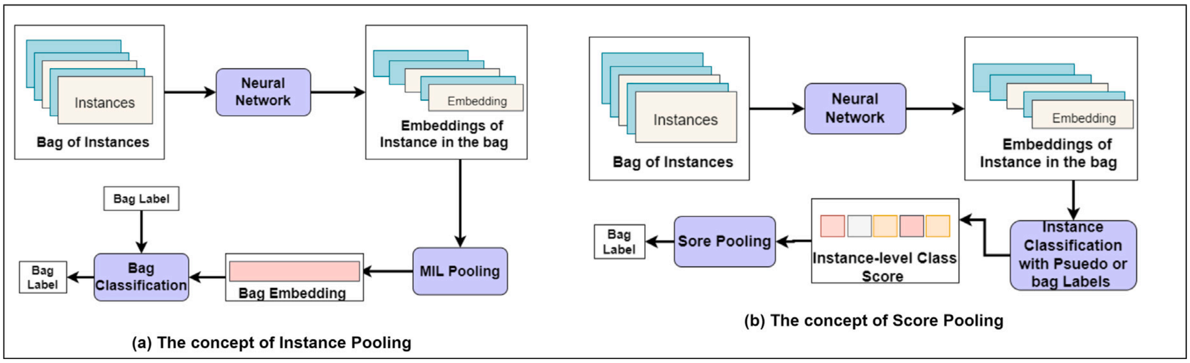

Fig. 6. Neural network-based Pooling Scheme for MIL, (a) illustrates the Instance pooling Scheme, while (b) depicts the Instance-wise class score pooling Scheme for MIL.

in neural networks, capsules perform operations on the input and encapsulate the output in the form of highly informative vectors.

The adaptive pooling is investigated by Zhou, Sun, Liu, Zha, and Zeng (2017) for the MIL video tagging application. The proposed learnable pooling function dynamically adapts based on different classes and allows end-to-end training. The proposed pooling function is based on generalized mean (Gulcehre, Cho, Pascanu, & Bengio, 2014), which adapts to adjust various classes. In addition, the combination of instance level and bag loss is also proposed. A discriminative pooling is proposed in Wang, Cherian, Porikli and Gould (2018) for action recognition based on separating nonlinear hyperplanes and addresses the problem of equally treating all frames' contributions in the pooling process. The proposed algorithms treat the parameters of the support vector hyperplane as a pooled representative descriptor for the video segment. A certainty pooling approach for MIL (CP-MIL) is proposed in Gildenblat, Ben-Shaul, Lapp, and Klaiman (2021). The proposed approach uses the Monte-Carlo dropout (Gal, Islam, & Ghahramani, 2017) and inverse standard deviation method to measure the uncertainty of instances. CP-MIL ensures the generation of weak gradients for ambiguous instances in the bag, and the bag's label is obtained by using instance-level certainty scores. Liu, Chen, Wang, and Zhang (2021) proposed a power pooling approach by attaching additional event-based trainable parameters for sound event classification. The proposed pooling function estimates instance-level probabilities and calculates a threshold for positive and negative gradients using the power exponent as a trainable parameter. The trainable parameter is used as an exponent of the frame-level probabilities. Transformation-invariant pooling (TI-Pooling) for MIL is proposed in Laptev, Savinov, Buhmann, and Pollefeys (2016). The proposed approach addresses the challenges of a larger number of model parameters, extra training time, and a high risk of variance or bias in the prediction process. These limitations are minimized by using a Siamese architecture (Bromley, Guyon, LeCun, Säckinger, & Shah, 1993) with a weight-sharing approach on the transformation set prior to FC layers. This process minimizes the redundancy in learned features and only uses the representative instances in the classification process, improving the overall predicted performance.

The challenges related to bag embedding generation using the neural network are studied in Lin and Zhang (2020), where an instance-embedding regularization technique is proposed for MIL. The proposed approach considers the instance similarity and maximizes the instance correlation in the embedding generation process. This technique considers the instances' relation into accounts, assumes that the instances are non-i.i.d., and encourages related instances to possess closer embeddings. A similar study related to bag embedding generation for histopathology images is proposed in Chikontwe, Kim, Nam, Go, and Park (2020). The proposed approach performs a joint learning for bag- and instance-level embeddings with a proposed centric loss, which minimizes the intra-class variation by mapping the embeddings of instances inside the bag to a single centroid position.

A novel ranking Loss-based MIL approach is proposed in Asif and ul Amir Afsar Minhas (2019) to differentiate between positive and negative bags by a simple pairwise bag-level ranking loss function. The proposed objective function ensures that the model assigns a higher score to the positive instances in the positive bags compared to instances in negative bags. This process helps to maximize the AUC score. However, it ignores the contribution of other instances in the bag classification process. A differentiable instance selection approach for MIL is proposed in Li et al. (2021), where the identification of the region of interest is proposed in an end-to-end trainable manner. The proposed approach utilizes the concept of gumble-softmax (Jang et al., 2016) to optimize the hard instance selection process. The proposed approach is more suitable for standard assumption, where a positive bag contains a single positive instance.

8.2. Attention-based pooling algorithms

The attention-based pooling approach can be regarded as a weighted average of the bag instance, where each instance receives a weight according to the participation in the bag label, with the following constraint:

$$\forall_i : \alpha_i \geq 0, \sum_{i=1}^{ni} \alpha_i = 1, \tag{6}$$

where α_i represents the weight of the ith instance in the bag and the vector representation for the bag is computed as:

$$\mathbf{z}_i = \sum_{i=1}^{ni} \alpha_i \mathbf{h}_i, \tag{7}$$

where the $\mathbf{h}_i$ represents the latent representation of the ith instance in the bag.

The weights of the instances play a vital role in the robust bag embedding generation process and in critical decision-making where the interpretation of the model is also needed, such as medical image classification and disease diagnostic (Carbonneau et al., 2018).

The attention-based pooling for MIL was first introduced by Ilse et al. (2018), where two attention-based pooling architectures were proposed. The first architecture, the MIL attention pooling (MIL-Attention), uses a two-layer MLP to generate the instance weights. The second architecture is gated MIL attention pooling (Gated-MIL-Attention), where additional layers and gating mechanisms are introduced for improved gradient flow in the network. The details of MIL-Attention and Gated-MIL-Attention architectures are given in Eqs. (8) and (9), respectively.

$$a_i = \frac{\exp\left\{\mathbf{w}^\top \tanh\left(\mathbf{V}\mathbf{h}_k^\top\right)\right\}}{\sum_{j=1}^{ni} \exp\left\{\mathbf{w}^\top \tanh\left(\mathbf{V}\mathbf{h}_j^\top\right)\right\}} \tag{8}$$

$$a_i = \frac{\exp\left\{\mathbf{w}^\top \left(\tanh\left(\mathbf{V}\mathbf{h}_k^\top\right) \odot \sigma\left(\mathbf{U}\mathbf{h}_k^\top\right)\right)\right\}}{\sum_{j=1}^{ni} \exp\left\{\mathbf{w}^\top \left(\tanh\left(\mathbf{V}\mathbf{h}_j^\top\right) \odot \sigma\left(\mathbf{U}\mathbf{h}_j^\top\right)\right)\right\}} \tag{9}$$

Table 3
The details of neural networks-based MIL algorithms.

Neural network pooling based MIL algorithms	Description	Classification granularity		Pooling technique
		Bag level	Instance level	
mi-net (Wang, Yan et al., 2018)	It generates instance embeddings and classifies each instance independently and pools the instance-level scores		✓	Instance score pooling
MI-Net (Wang, Yan et al., 2018)	It generates bag embedding by pooling instance embeddings and performs bag classification directly.	✓		Instance pooling
Dynamic pooling (DP-MINN) (Yan et al., 2018)	Uses a dynamic routing capsule and iterative update to pool the instances	✓		Iterative weighted-sum pooling
Adaptive pooling (Zhou et al., 2017)	It uses a generalized-mean operator, an alternative to the average pool.	✓		Instance pooling
Certainty pooling (Gildenblat et al., 2021)	Uses the Monte-Carlo dropout and inverse standard deviation method to measure the certainty of instances in the bag.	✓		Instance pooling
Power pooling (Liu et al., 2021)	Estimates instance-level probabilities and calculates a threshold for positive and negative gradients using the power exponent as a trainable parameter		✓	Instance score pooling

Table 4
The available and produced datasets is multi-instance learning.

Datasets of MIL produced, created and available	Description	Classification		Citations
		Journal	Conference	
Bio Creative Dataset (Ray & Craven, 2005a) [Dataset Link]	This problem belongs to the area of text categorization. We have different documents consisting of paragraphs as feature vectors, the Task is to annotate GO code	✓	✓	62
Birds Sound Dataset (Briggs, Fern and Raich, 2012) [Dataset Link]	The bird songs problem was initially multi-label multi-class. Each bag records one or more birds. The bag gets all bird labels from the recording. Choosing a bird "target class" turns this into an MI issue.		✓	240
Corel (20-Classes) Dataset (Chen et al., 2006) [Dataset Link]	Image categorization issue Corel has 20 classes. A positive class is assigned. Bags are images, while instances are image fragments. Segments are shown by the mean of 4×4 patch features.	✓		918
Elephant, Fox & Tiger Dataset (Andrews, Tsochantaridis et al., 2002) [Dataset Link]	The datasets are Fox, Tiger, and Elephant. Bags are images, instances are image fragments. For each category, +ve bags contain the animal while −ve contains other animals.	✓		1993
Messidor Eye-Fundus 1200 Images Dataset (Kandemir & Hamprecht, 2015) [Dataset Link]	Eye fundus photos from 654 diabetics and 546 healthy people. Original data images are rescaled to 700×700 pixels and divided into 135×135 pixel patches. Foreground-deficient patches are removed.	✓		1101
Musk Dataset (Dietterich et al., 1997) [Dataset Link]	Musk1 and Musk2 databases Both predict if a chemical smells musky Each bag represents a molecule & each instance its conformer. Conformers determine molecular characteristics.	✓		3320
Mutagenesis Dataset (Srinivasan & Muggleton, 1995) [Dataset Link]	Drug activity prediction problem: mutagenesis. The dataset has easy (1) and hard (2) variants.		✓	83
Histology Images of Colorectal Adenocarcinomas Data (Colon Data) (Sirinukunwattana, Raza, et al., 2016) [Dataset Link]	images of 10 whole-slide images from 9 patients were cropped to 500×500 pixels after non-overlapping parts were identified. The pixel resolution is 0.55 μm (20× optical magnification) Omnix VL120 scanned all slides.		✓	1211
Breast Cancer Dataset (Gelasca & Byun, 2008) [Dataset Link]	Lacks marking on 58, 896×768 H&E biopsy pictures. A malignant image has at least one cancer cell. Benign otherwise. Each image splits into 32×32 patches, and 672 pictures were created. At least 75% of white patch pixels are removed, yielding 58 bag sizes.	✓		168

The $\mathbf{V} \in \mathbb{R}^{L\times M}$, $\mathbf{U} \in \mathbb{R}^{L\times M}$ and $\mathbf{w} \in \mathbb{R}^{L\times 1}$ are the weights of MLP layers. The $\odot$, $\tanh(:)$, and $\sigma(:)$ represent the elementwise multiplication, tangent-hyperbolic, and sigmoid activation functions respectively. Shi et al. (2020) Further investigated the attention approach for MIL, compared to the previous attention approach (Ilse et al., 2018), where separate attention layers are employed to compute the instance weights,

Table 5
The available and produced datasets is multi-instance learning.

Datasets of MIL produced, created and available	Description	Classification		Citations
		Journal	Conference	
Fungi Species Dataset (Zielinski & Sroka-Oleksiak, 2019) [Dataset Link]	Multi-class categorization includes 180 photos for 9 fungus species, divided into 2 preparations with 10 images per species. Photos are 5760 × 3600 × 3 pixels	✓		15
(CAMELYON16 and 17 (Litjens et al., 2018) [Dataset Link] [Dataset Link]	CAMELYON16 and CAMELYON17 challenges have 399 and 1000 WSIs, respectively. This yielded 1399 distinct WSIs and 2.95 GB of data.	✓		303
MIL Colon Cancer Dataset (Patches) (Sirinukunwattana et al., 2016) [Dataset Link]	This dataset consists of 100 H&E images belonging to binary classes. Each image has several patches.	✓		1250
MIL-based MNIST Dataset (Ilse et al., 2018) [Dataset Link]	The dataset is synthesized form MNIST- Digit Dataset. Where bag contains one or more digits, and bag is labeled as positive if it contains the digit "9" in it.		✓	1450
MIL-based CIFAR-10 Dataset (Shi et al., 2020) [Dataset Link]	This dataset is synthesized using the well-known CIFAR-10 dataset. The bags in this dataset are composed of images of different objects.		✓	71

the proposed approach simultaneously uses classification layers as attention layers. Furthermore, additional attention-based cross-entropy loss is proposed for model training in combination with bag loss. The proposed loss function allows the model to train in multi-class MIL scenarios while learning attention-based instance weights. The workflow of attention-based MIL pooling is shown in Fig. 7. Attention-based MIL pooling is an effective and interpretable approach. It emphasizes the importance of instances in the bag through instance weighting. Model interpretation is essential For high-risk applications such as medical image analysis and classification (Chikontwe et al., 2022; Lu et al., 2021; Oner, Kye-Jet, Lee, & Sung, 2023; Qaiser et al., 2021; Tourniaire, Ilie, Hofman, Ayache, & Delingette, 2021), disease detection (Han et al., 2020; Wu, Schmidt, Hernández-Sánchez, Molina, & Katsaggelos, 2021; Yu et al., 2021; Zheng et al., 2022), text classification (Xiao, Jin, Cheng, & Hao, 2022), and many others. Additionally, The concept of recent MLP-Mixer architecture (Tolstikhin et al., 2021) with MIL pooling is also explored to generate robust bag embeddings in Waqas, Khan, Ahmed and Raza (2023), the study extends MLP-Mixer to support the concept of varying number of instances in the bag and assumptions of MIL. The authors propose a token mixing technique that allows them to use the relationship of instances in the bag to generate a robust bag encoding.

The study in Castro-Macías et al. (2024) introduces the concept of contrastive learning (SCL) for MIL, which aims to learn balanced feature spaces while handling imbalance within whole slide images (WSIs). This approach gradually moves from learning bag-level representations to optimal classifier learning, resulting in the ability to handle tasks related to cancer pathology, such as subtyping non-small cell lung cancer (NSCLC) and renal cell carcinoma (RCC). The method's performance is evaluated and tested on these tasks. In Waqas et al. (2024), the concept of attention pooling is analyzed for generating i.i.d and non-i.i.d bag representations. The study aims to obtain bag-level information for both assumptions i.i.d and non-i.i.d and select one of them for further processing through an end-to-end trainable approach.

9. Publically available MIL datasets

As a versatile framework, Multiple Instance Learning (MIL) has been applied to a variety of fields to address complex problems. It has been used in drug discovery (Dietterich et al., 1997), healthcare (Ren, Zhao et al., 2023; Shanmugam, Blalock, & Guttag, 2019), object detection (Chowdhury et al., 2023), and text categorization (Toraman, 2011). For example, Musk's dataset (Dietterich et al., 1997) has been widely used for identifying chemical compounds exhibiting specific properties, which has improved the efficiency of drug development. In healthcare, datasets such as the Histology images of colorectal adenocarcinomas Data (Colon Data) (Sirinukunwattana et al., 2016) and CAMELYON-16 and CAMELYON-17 (Litjens et al., 2018) for cancer detection and diagnosis aiding in the early detection of diseases and evaluation of MIL algorithms. MIL is also applicable to computer vision tasks, as demonstrated by datasets such as Fox, Elephant, and Tiger datasets (Andrews, Tsochantaridis et al., 2002), and Corel (20-Classes) (Chen et al., 2006). These datasets are benchmark MIL datasets and valuable resources for the development of algorithms. Datasets such as Newsgroup (Zhou et al., 2009), and Web Recommendation (Zhou, Jiang, & Li, 2005) are also used in several MIL studies.

The details related to publically available MIL datasets are presented in Tables 5 and 4. By utilizing these datasets the performance and diversity of MIL algorithms are compared. Customized datasets drive MIL algorithm advancements and interdisciplinary research. As datasets expand, MIL's applications will multiply, unlocking solutions in domains like environmental monitoring and finance.

10. Unsupervised MIL techniques

Multiple Instance Learning (MIL) has also been used in various unsupervised scenarios. Unsupervised Multiple Instance Learning (U-MIL) takes it a step further by addressing situations where bag-level labels are not available. In UMIL, the learning process occurs without any manual annotations on bags or instances. The objective is to extract meaningful patterns or representations directly from the data without relying on labeled examples. UMIL is especially useful in situations where obtaining any type of supervision is difficult, making it a more generalized and challenging setting than MIL (Herrera et al., 2016b; Zhang & Zhou, 2009).

In Henegar, Clément, and Zucker (2006), the concept of Unsupervised Multiple Instance Learning (U-MIL) was presented to extend the widely used MIL concept in unsupervised settings. The study suggests three algorithmic solutions; agglomerative clustering, partition clustering, ans the citationkNN (Wang & Zucker, 2000) approach. These solutions are designed to cluster data objects described by the multiple-instance representation effectively. The evaluation involved relating gene expression data to biological annotations and representing them in the UMIL framework. Standard clustering quality measures were employed to assess the performance of the algorithms. The paper also demonstrates its practical applicability through a thorough bioinformatics analysis.

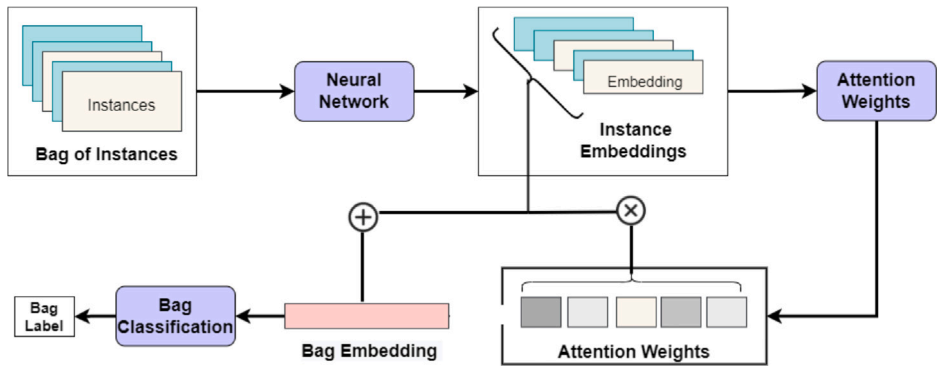

Fig. 7. The working of attention-based pooling.

Another study in Alam and Shehu (2021) addresses the challenge of decoy selection in protein structure prediction, focusing on inaccessible regions of the protein collection. The authors introduce an approach by formulating the problem as an instance of unsupervised multi-instance learning. The method involves organizing decoys into bags, identifying relevant bags, and drawing individual instances for prediction using both non-parametric and parametric algorithms. This research provides a promising avenue for advancing decoy selection techniques in protein structure prediction, showcasing the potential of unsupervised multi-instance learning to improve accuracy and efficiency in this critical area of structural biology.

In Alam and Shehu (2020), the authors focus on solving the challenge of identifying native tertiary protein structures among template-free predictions in computational biology. To tackle this issue, the paper used unsupervised multi-instance learning U-MIL, which involves a three-stage process: organizing structures into bags, identifying relevant bags, and drawing individual structures using both non-parametric and parametric algorithms. This approach offers a promising solution for increasing the accuracy of template-free protein structure predictions by selecting biologically relevant structures from diverse possibilities. A new unsupervised method for computational pathology has been proposed in Tavolara, Gurcan, and Niazi (2022). It uses a tile-wise encoder and a multiple-instance learning framework to generate slide-level representations without the need for slide-level labels. The method has been applied to lung cancer subtyping and breast cancer proliferation scoring tasks. The Unsupervised Multiple Instance Learning (U-MIL) method is used for unsupervised computer vision problems, as stated in Wang and Yuan (2018). In this study, the U-MIL approach was proposed to optimize the bag label and instance label jointly in a unified framework under the constraint of the Noisy-OR model (Zhang et al., 2005). This approach can be easily applied to object discovery in wild images by treating the object proposals extracted from images as instances, and the corresponding images as bags.

Unsupervised Multiple Instance Learning (U-MIL) has limitations primarily due to the absence of explicit labels, which makes it challenging to evaluate and validate learned models. Algorithms in unsupervised tasks must infer patterns without the guidance of labeled examples, leading to difficulties in model interpretation. Defining effective objective functions becomes problematic in the absence of clear goals based on labeled data, and evaluating performance relies on indirect metrics, which can potentially impact the practical utility of learned representations. Additionally, U-MIL can be sensitive to initialization, may struggle with sparse or highly variable data, and demands extensive computational resources. Addressing these challenges is crucial to expanding the practical applicability of U-MIL and enhancing its effectiveness in various real-world scenarios.

11. Open challenges and future direction

Multiple Instance Learning (MIL) is a machine learning approach that is particularly useful for handling data with weak or imprecise labels. Unlike traditional supervised learning, MIL embraces the concept of bags containing multiple instances, allowing it to manage ambiguity more effectively. This has led to the development of algorithms that can cope with scenarios where instances are not explicitly labeled, making them useful tools in applications such as medical diagnosis, object recognition, and text categorization. MIL's adaptability to varying levels of label ambiguity has made it a valuable tool in many real-world settings.

Although Multiple Instance Learning (MIL) provides important characteristics that help in weakly labeled scenarios, the complex bag representation of data leads to a complex hypothesis space (Waqas, Tahir, Khan, 2023; Waqas, Tahir, Qureshi, 2023; Wei et al., 2014, 2016; Yuan et al., 2020). This situation arises when the learner finds it extremely difficult to converge. For example, algorithms like MI-SVM (Andrews, Tsochantaridis et al., 2002), MI-Boosting (Hong et al., 2013), and MI-Wrapper (Frank & Xu, 2003) are unable to deal with the complicated hypothesis spaces. Furthermore, several SVM-based MIL techniques (Andrews, Tsochantaridis et al., 2002; Zhou & Xu, 2007) have non-convex objective functions, which can result in convergence to local minima or require multiple starts. Therefore, exploring convex optimization functions and methods to reduce the complexity of hypothesis space is an open research avenue.

Several MIL algorithms (Andrews, Tsochantaridis et al., 2002; Carbonneau et al., 2016; Waqas et al., 2021; Waqas, Tahir, Qureshi, 2023; Wei et al., 2014, 2016) are built on the assumption that instances are independent and drawn from the same distribution. This assumption is known as the i.i.d. assumption. However, this assumption may not hold true for MIL (Multiple Instance Learning) problems where instances are grouped into bags. In MIL, each bag contains multiple instances, and the labels are assigned at the bag level, which leads to a departure from the i.i.d. assumption. This is because instances within the same bag are not independent, as they share a collective label. This assumption may limit the performance of the MIL algorithm. Several studies (Bi et al., 2023; Rymarczyk et al., 2021; Yu et al., 2021; Zhou et al., 2009) have explored this dimension of the research using a graph-based model and using self-attention, kernel attention process. However, these studies majorly focused on specific types of data. Therefore, it requires the development of specialized MIL algorithms that can take into account and utilize these complex relationships within bags. Additionally, handling interdependencies among instances within bags requires more complex and computationally intensive algorithms. Furthermore, interpreting the decision-making process of these models remains a challenge due to the inherent complexities of non-i.i.d data.

It would be interesting to explore whether all bags follow the same distribution or not. It is not theoretically guaranteed that all bags in the dataset follow the same distribution. However, it would be a topic of exploration whether one bag in the dataset follows i.i.d while the other bag follows non-i.i.d. Furthermore, in the context of Multiple Instance Learning (MIL), domain shift may occur when the characteristics of bags and instances in the testing environment are different from those in the training dataset (Gonthier, Ladjal, & Gousseau, 2022; Pocevičiūtė, Eilertsen, Garvin, & Lundström, 2023). This shift can include changes in feature distributions, instance relationships, label ambiguity, or bag structures. It is necessary to adapt models to enable them to generalize more effectively across different domains, to address domain shift in multiple instance learning (MIL). There are various techniques like domain adaptation, transfer learning (Xiao, Liang, & Liu, 2020), and instance re-weighting (Zhang & Zhou, 2017) can be used to align distributions between source and target domains or to learn robust representations that can be transferred across domains would be an open research area for the researchers.

It has been observed that the effectiveness of several sub-space clustering-based techniques for instance selection, such as RSIS (Carbonneau et al., 2016), RMI-FV (Waqas et al., 2021), and DGMIR-FV (Waqas, Tahir, Qureshi, 2023), depends on the performance and convergence of clustering algorithms. The application of different clustering techniques can significantly affect the performance of these algorithms. Therefore, it would be an area of interest for researchers to develop clustering algorithms with objective functions that support MIL assumptions. Additionally, the instance selection techniques consider the instances in the bag are i.i.d. It would be interesting to explore methods that can consider instances related to the instance selection process.

Graph Neural Networks (GNNs) have emerged as a promising approach for many supervised problems, offering the ability to effectively capture complex relationships among instances within data. However, limited studies are available on using GNNs for MIL (Pal, Valkanas, Regol, & Coates, 2022; Tu, Huang, He, & Zhou, 2019; Xiong et al., 2022; Zhao et al., 2020). The topic of MIL-based GNN can excel in modeling bag structures by leveraging graph-based architectures, enabling the extraction of rich bag-level representations. Their capability to learn from graph structures and aggregate instance-level information can be a potential topic of interest for researchers. Additionally, converting the bag representation into a graph structure can also be explored to identify a suitable transformation technique (Zhou et al., 2009).

The utilization of Generative Artificial Intelligence and large language model (LLMs) (Hadi et al., 2023; Kingma & Welling, 2013; Park, Ko, Huh, & Kim, 2021) can also be applied to Multiple Instance Learning (MIL) frameworks, but it requires customization to produce data that conforms to the structure of bags (Koohi-Moghadam & Bae, 2023; Musalamadugu & Kannan, 2023; Pinaya et al., 2023). In Multiple Instance Learning (MIL), positive bags need to have specific characteristics. For example, when generating bags from Whole Slide Images (WSIs) for cancer diagnosis, the generative approach can create patches that show visible cancer symptoms. On the other hand, for negative bags, it is essential to ensure the absence of such patches that illustrate cancer-related characteristics. Customizing Generative AI models to meet these bag-level requirements is crucial to creating synthetic instances that align with the specific characteristics needed for positive and negative bags in MIL applications. This tailored approach helps improve the relevance and quality of generated instances, enhancing the usefulness of Generative AI within MIL frameworks, especially in specialized domains like medical imaging (Kasban, El-Bendary, & Salama, 2015). Furthermore, ensuring the quality and regulation of generative data would also be an interesting investigation avenue. In conclusion, the exploration of the above lessons learned and future directions in Multiple Instance Learning (MIL) underscores several critical insights and evergrowing avenues for advancement, making it an open research choice for researchers worldwide.

12. Conclusion

In this review, we aim to provide a detailed understanding of Multiple Instance Learning (MIL) by discussing its background, methodology, applications, challenges, and future directions. We begin by explaining what MIL is, why it is important for real-world problems, and the basic assumptions underlying MIL. Later, we presented how existing supervised learning algorithms can be customized to support MIL assumptions and requirements, which were not covered in the previous studies. We also described instance selection and bag encoding strategies, along with a detailed discussion of Neural network-based pooling techniques used in state-of-the-art studies. Finally, the paper also discusses lessons learned from the past and directions for future research endeavors in MIL. The paper attempted to comprehensively cover various aspects of MIL, providing foundational knowledge for beginners, while also delving into minute details to serve as a practical guide for MIL practitioners. This manuscript is a comprehensive resource for researchers in multiple instance learning. It covers basic concepts and advanced methodologies, promoting ongoing progress and understanding in weakly supervised learning.

CRediT authorship contribution statement

Muhammad Waqas: Conceptualization, Software, Data Curation, Writing, Writing – review & editing. **Syed Umaid Ahmed:** Writing, Writing – review & editing. **Muhammad Atif Tahir:** Supervision, Validation. **Jia Wu:** Writing, Validation, Supervision. **Rizwan Qureshi:** Writing – review & editing, Validation.

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.

Data availability

No data was used for the research described in the article.

References

Adate, A., & Tripathy, B. (2022). A survey on deep learning methodologies of recent applications. In *Deep learning in data analytics* (pp. 145–170). Springer.

Ahmed, S. U., Affan, M., Raza, M. I., & Hashmi, M. H. (2022). Inspecting mega solar plants through computer vision and drone technologies. In *2022 international conference on frontiers of information technology* (pp. 18–23). IEEE.

Ahmed, S. U., Ahmad, Q. N., Shuja, J., & Affan, M. (2022). Analysis of deep learning algorithms on edge in microscopic fabric dataset. In *2022 17th international conference on emerging technologies* (pp. 54–58). IEEE.

Ahmed, S. U., Khalid, H., Affan, M., Khan, T. A., & Ahmad, M. (2020). Smart surveillance and tracking system. In *2020 IEEE 23rd international multitopic conference* (pp. 1–5). IEEE.

Ahmed, S. U., Shuja, J., & Tahir, M. A. (2023). Leaf classification on flavia dataset: A detailed review. *Sustainable Computing: Informatics and Systems*, Article 100907.

Alam, F., & Shehu, A. (2020). From unsupervised multi-instance learning to identification of near-native protein structures. *EPiC Series in Computing*, 70.

Alam, F. F., & Shehu, A. (2021). Unsupervised multi-instance learning for protein structure determination. *Journal of Bioinformatics and Computational Biology*, *19*, Article 2140002.

Amores, J. (2013). Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence*, *201*, 81–105.

Andrews, S., Hofmann, T., & Tsochantaridis, I. (2002). Multiple instance learning with generalized support vector machines. In *AAAI/IAAI* (pp. 943–944).

Andrews, S., Tsochantaridis, I., & Hofmann, T. (2002). Support vector machines for multiple-instance learning. In *NIPS* (p. 7).

Asif, A., & ul Amir Afsar Minhas, F. (2019). An embarrassingly simple approach to neural multiple instance classification. *Pattern Recognition Letters*, *128*, 474–479.

Aslani, S., & Jacob, J. (2023). Utilisation of deep learning for covid-19 diagnosis. *Clinical Radiology*, *78*, 150–157.

Astorino, A., Fuduli, A., Gaudioso, M., & Vocaturo, E. (2019). Multiple instance learning algorithm for medical image classification. In *SEBD* (pp. 1–8).

Auer, P., & Ortner, R. (2004). A boosting approach to multiple instance learning. In *European conference on machine learning* (pp. 63–74). Springer.

Avolio, M., & Fuduli, A. (2020). A semiproximal support vector machine approach for binary multiple instance learning. *IEEE Transactions on Neural Networks and Learning Systems*, *32*, 3566–3577.

Avolio, M., & Fuduli, A. (2023). The semiproximal svm approach for multiple instance learning: a kernel-based computational study. *Optimization Letters*, 1–15.

Babenko, B., Yang, M. H., & Belongie, S. (2009). Visual tracking with online multiple instance learning. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 983–990). IEEE.

Babenko, B., Yang, M. H., & Belongie, S. (2010). Robust object tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*, 1619–1632.

Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, *318*, 2199–2210.

Bi, Q., Sun, X., Yu, S., Ma, K., Bian, C., Ning, M., et al. (2023). Mil-vit: A multiple instance vision transformer for fundus image classification. *Journal of Visual Communication and Image Representation*, *97*, Article 103956.

Blockeel, H., Page, D., & Srinivasan, A. (2005). Multi-instance tree learning. In *Proceedings of the 22nd international conference on machine learning* (pp. 57–64).

Bontempo, G., Porrello, A., Bolelli, F., Calderara, S., & Ficarra, E. (2023). Das-mil: Distilling across scales for mil classification of histological wsis. In *International conference on medical image computing and computer-assisted intervention* (pp. 248–258). Springer.

Brand, L., Baker, L. Z., & Wang, H. (2021). A multi-instance support vector machine with incomplete data for clinical outcome prediction of covid-19. In *Proceedings of the 12th ACM conference on bioinformatics, computational biology, and health informatics* (pp. 1–6).

Brand, L., Seo, H., Baker, L. Z., Ellefsen, C., Sargent, J., & Wang, H. (2023). A linear primal–dual multi-instance svm for big data classifications. *Knowledge and Information Systems*, 1–32.

Brand, L., Seo, H., Baker, L. Z., Ellefsen, C., Sargent, J., & Wang, H. (2024). A linear primal–dual multi-instance svm for big data classifications. *Knowledge and Information Systems*, *66*, 307–338.

Brendal, C. (2023). *MRI artifact detection: schematic design in perceptual training for radiological board exam preparation* (Ph.D. thesis), Johns Hopkins University.

Briggs, F., Fern, X. Z., & Raich, R. (2012). Rank-loss support instance machines for miml instance annotation. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 534–542).

Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S. J., et al. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, *131*, 4640–4650.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a siamese time delay neural network. *Advances in Neural Information Processing Systems*, 6.

Brooks, J. P. (2011). Support vector machines with the ramp loss and the hard margin loss. *Operations Research*, *59*, 467–479.

Carbonneau, M. A., Cheplygina, V., Granger, E., & Gagnon, G. (2018). Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, *77*, 329–353.

Carbonneau, M. A., Granger, E., Raymond, A. J., & Gagnon, G. (2016). Robust multiple-instance learning ensembles using random subspace instance selection. *Pattern Recognition*, *58*, 83–99.

Castro-Macías, F. M., Morales-Álvarez, P., Wu, Y., Molina, R., & Katsaggelos, A. K. (2024). Hyperbolic secant representation of the logistic function: Application to probabilistic multiple instance learning for ct intracranial hemorrhage detection. *Artificial Intelligence*, Article 104115.

Chen, Y., Bi, J., & Wang, J. Z. (2006). Miles: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 1931–1947.

Chen, W., Li, G., Zhang, X., Wang, S., Li, L., & Huang, Q. (2023). Weakly supervised text-based actor-action video segmentation by clip-level multi-instance learning. *ACM Transactions on Multimedia Computing, Communications and Applications*, *19*, 1–22.

Chen, L., Tong, T., Ho, C. P., Patel, R., Cohen, D., Dawson, A. C., et al. (2015). Identification of cerebral small vessel disease using multiple instance learning. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part i 18* (pp. 523–530). Springer.

Chen, Y., & Wang, J. Z. (2004). Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, *5*, 913–939.

Chen, J., Zeng, H., Zhang, C., Shi, Z., Dekker, A., Wee, L., et al. (2022). Lung cancer diagnosis using deep attention-based multiple instance learning and radiomics. *Medical Physics*, *49*, 3134–3143.

Chevaleyre, Y., & Zucker, J. D. (2001). Solving multiple-instance and multiple-part learning problems with decision trees and rule sets. application to the mutagenesis problem. In *Conference of the Canadian society for computational studies of intelligence* (pp. 204–214). Springer.

Chikontwe, P., Kim, M., Nam, S. J., Go, H., & Park, S. H. (2020). Multiple instance learning with center embeddings for histopathology classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 519–528). Springer.

Chikontwe, P., Nam, S. J., Go, H., Kim, M., Sung, H. J., & Park, S. H. (2022). Feature re-calibration based multiple instance learning for whole slide image classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 420–430). Springer.

Choi, W. G., Chang, J. H., Yang, J. M., & Moon, H. G. (2024). Instance-level loss based multiple-instance learning framework for acoustic scene classification. *Applied Acoustics*, *216*, Article 109757.

Choi, J., Seo, H., Im, S., & Kang, M. (2019). Attention routing between capsules. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*.

Chowdhury, P. N., Bhunia, A. K., Sain, A., Koley, S., Xiang, T., & Song, Y. Z. (2023). What can human sketches do for object detection? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 15083–15094).

Cinbis, R. G., Verbeek, J., & Schmid, C. (2016). Weakly supervised object localization with multi-fold multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 189–203.

Deng, R., Cui, C., Remedios, L. W., Bao, S., Womick, R. M., Chiron, S., et al. (2022). Cross-scale attention guided multi-instance learning for crohn's disease diagnosis with pathological images. In *International workshop on multiscale multimodal medical imaging* (pp. 24–33). Springer.

Deng, R., Cui, C., Remedios, L. W., Bao, S., Womick, R. M., Chiron, S., et al. (2024). Cross-scale multi-instance learning for pathological image diagnosis. *Medical Image Analysis*, Article 103124.

Dietterich, T. G., Lathrop, R. H., & Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, *89*, 31–71.

Ding, R. X., Palomares, I., Wang, X., Yang, G. R., Liu, B., Dong, Y., et al. (2020). Large-scale decision-making: Characterization, taxonomy, challenges and future directions from an artificial intelligence and applications perspective. *Information Fusion*, *59*, 84–102.

Dong, L. (2006). *A comparison of multi-instance learning algorithms* (Ph.D. thesis), The University of Waikato.

Doran, G., & Ray, S. (2016). Multiple-instance learning from distributions. *Journal of Machine Learning Research*, *17*, 4384–4433.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Dou, B., Zhu, Z., Merkurjev, E., Ke, L., Chen, L., Jiang, J., et al. (2023). Machine learning methods for small data challenges in molecular science. *Chemical Reviews*, *123*, 8736–8780.

Du, R., Wu, Q., He, X., & Yang, J. (2013). Mil-skde: Multiple-instance learning with supervised kernel density estimation. *Signal Processing*, *93*, 1471–1484.

Dundar, M., Krishnapuram, B., Rao, R., & Fung, G. (2006). Multiple instance learning for computer aided diagnosis. *Advances in Neural Information Processing Systems*, 19.

Fatima, S., Ali, S., & Kim, H. C. (2023). A comprehensive review on multiple instance learning. *Electronics*, *12*, 4323.

Feng, L., Shu, S., Cao, Y., Tao, L., Wei, H., Xiang, T., et al. (2021). Multiple-instance learning from similar and dissimilar bags. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 374–382).

Feng, S., Xiong, W., Li, B., Lang, C., & Huang, X. (2014). Hierarchical sparse representation based multi-instance semi-supervised learning with application to image categorization. *Signal Processing*, *94*, 595–607.

Foulds, J., & Frank, E. (2010). A review of multi-instance learning assumptions. *The Knowledge Engineering Review*, *25*, 1–25.

Frade, J., Pereira, T., Morgado, J., Silva, F., Freitas, C., Mendes, J., et al. (2022). Multiple instance learning for lung pathophysiological findings detection using ct scans. *Medical & Biological Engineering & Computing*, *60*, 1569–1584.

Frank, E., & Xu, X. (2003). *Applying propositional learning algorithms to multi-instance data*. University of Waikato.

Fu, G., Nan, X., Liu, H., Patel, R. Y., Daga, P. R., Chen, Y., et al. (2012). Implementation of multiple-instance learning in drug activity prediction. In *BMC bioinformatics* (pp. 1–12). BioMed Central.

Fu, Z., Robles-Kelly, A., & Zhou, J. (2010). Milis: Multiple instance learning with instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*, 958–977.

Fuhrman, J., Yip, R., Zhu, Y., Jirapatnakul, A. C., Li, F., Henschke, C. I., et al. (2023). Evaluation of emphysema on thoracic low-dose cts through attention-based multiple instance deep learning. *Scientific Reports*, *13*, 1187.

Gadermayr, M., & Tschuchnig, M. (2024). Multiple instance learning for digital pathology: A review of the state-of-the-art, limitations & future potential. *Computerized Medical Imaging and Graphics*, Article 102337.

Gal, Y., Islam, R., & Ghahramani, Z. (2017). Deep bayesian active learning with image data. In *International conference on machine learning* (pp. 1183–1192). PMLR.

Gao, Z., Mao, A., Wu, K., Li, Y., Zhao, L., Zhang, X., et al. (2023). Childhood leukemia classification via information bottleneck enhanced hierarchical multi-instance learning. *IEEE Transactions on Medical Imaging*.

Gao, W., Wan, F., Yue, J., Xu, S., & Ye, Q. (2022). Discrepant multiple instance learning for weakly supervised object detection. *Pattern Recognition*, *122*, Article 108233.

Gehler, P. V., & Chapelle, O. (2007). Deterministic annealing for multiple-instance learning. In *Artificial intelligence and statistics* (pp. 123–130). PMLR.

Gelasca, E. D., & Byun, e. (2008). Evaluation and benchmark for biological image segmentation. In *2008 15th IEEE international conference on image processing* (pp. 1816–1819). IEEE.

Gildenblat, J., Ben-Shaul, I., Lapp, Z., & Klaiman, E. (2021). Certainty pooling for multiple instance learning. In *International conference on pattern recognition* (pp. 141–153). Springer.

Godson, L., Alemi, N., Nsengimana, J., Cook, G. P., Clarke, E. L., Treanor, D., et al. (2024). Immune subtyping of melanoma whole slide images using multiple instance learning. *Medical Image Analysis*, *93*, Article 103097.

Gonthier, N., Ladjal, S., & Gousseau, Y. (2022). Multiple instance learning on deep features for weakly supervised object detection with extreme domain shifts. *Computer Vision and Image Understanding*, *214*, Article 103299.

Gulcehre, C., Cho, K., Pascanu, R., & Bengio, Y. (2014). Learned-norm pooling for deep feedforward and recurrent neural networks. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 530–546). Springer.

Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., et al. (2023). Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *Authorea Preprints*.

Han, Y., & Qi, X. (2005). A complementary svms-based image annotation system. In *IEEE international conference on image processing 2005* (pp. I–1185). IEEE.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., et al. (2022). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Han, Z., Wei, B., Hong, Y., Li, T., Cong, J., Zhu, X., et al. (2020). Accurate screening of covid-19 using attention-based deep 3d multiple instance learning. *IEEE Transactions on Medical Imaging*, *39*, 2584–2594.

Hanif, M., Waqas, M., Muneer, A., Alwadain, A., Tahir, M. A., & Rafi, M. (2023). Deepsdc: Deep ensemble learner for the classification of social-media flooding events. *Sustainability*, *15*, 6049.

He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., et al. (2023). Transformers in medical image analysis. *Intelligent Medicine*, *3*, 59–78.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hebbar, R., Papadopoulos, P., Reyes, R., Danvers, A. F., Polsinelli, A. J., Moseley, S. A., et al. (2021). Deep multiple instance learning for foreground speech localization in ambient audio from wearable devices. *EURASIP Journal on Audio, Speech, and Music Processing*, *2021*, 1–8.

Henderson, D., Jacobson, S. H., & Johnson, A. W. (2003). The theory and practice of simulated annealing. *Handbook of Metaheuristics*, *28*, 7–319.

Henegar, C., Clément, K., & Zucker, J. D. (2006). Unsupervised multiple-instance learning for functional profiling of genomic data. In *Machine learning: ECML 2006: 17th European conference on machine learning Berlin, Germany, September 18-22, 2006 proceedings 17* (pp. 186–197). Springer.

Herrera, F., Ventura, S., Bello, R., Cornelis, C., Zafra, A., Sánchez-Tarragó, D., et al. (2016a). *Multiple instance learning*. Springer.

Herrera, F., Ventura, S., Bello, R., Cornelis, C., Zafra, A., Sánchez-Tarragó, D., et al. (2016b). Unsupervised multiple instance learning. In *Multiple instance learning: foundations and algorithms* (pp. 141–167).

Hong, R., Wang, M., Gao, Y., Tao, D., Li, X., & Wu, X. (2013). Image annotation by multiple-instance learning with discriminative feature mapping and selection. *IEEE Transactions on Cybernetics*, *44*, 669–680.

Huang, S., Liu, Z., Jin, W., & Mu, Y. (2022). Bag dissimilarity regularized multi-instance learning. *Pattern Recognition*, *126*, Article 108583.

Huang, F., Qi, J., Lu, H., Zhang, L., & Ruan, X. (2017). Salient object detection via multiple instance learning. *IEEE Transactions on Image Processing*, *26*, 1911–1922.

Ilse, M., Tomczak, J., & Welling, M. (2018). Attention-based deep multiple instance learning. In *International conference on machine learning* (pp. 2127–2136). PMLR.

Jang, E., Gu, S., & Poole, B. (2016). Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144.

Jiang, H., Zhou, Y., Lin, Y., Chan, R. C., Liu, J., & Chen, H. (2023). Deep learning for computational cytology: A survey. *Medical Image Analysis*, *84*, Article 102691.

Joachims, T., et al. (1999). Transductive inference for text classification using support vector machines. In *ICML* (pp. 200–209).

Jolad, B., & Khanai, R. (2022). Anns for automatic speech recognition—a survey. In *Expert clouds and applications* (pp. 35–48). Springer.

Kamoona, A. M., Gostar, A. K., Bab-Hadiashar, A., & Hoseinnezhad, R. (2023). Multiple instance-based video anomaly detection using deep temporal encoding–decoding. *Expert Systems with Applications*, *214*, Article 119079. http://dx.doi.org/10.1016/j.eswa.2022.119079.

Kandemir, M., & Hamprecht, F. A. (2015). Computer-aided diagnosis from weak supervision: A benchmarking study. *Computerized Medical Imaging and Graphics*, *42*, 44–50.

Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, *65*, Article 101759.

Kasban, H., El-Bendary, M., & Salama, D. (2015). A comparative study of medical imaging techniques. *International Journal of Information Science and Intelligent System*, *4*, 37–58.

Khan, S. M., & Shah, M. (2006). A multiview approach to tracking people in crowded scenes using a planar homography constraint. In *Computer vision–ECCV 2006: 9th European conference on computer vision, Graz, Austria, May 7-13, 2006, proceedings, part IV 9* (pp. 133–146). Springer.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Koohi-Moghadam, M., & Bae, K. T. (2023). Generative ai in medical imaging: Applications, challenges, and ethics. *Journal of Medical Systems*, *47*, 94.

Krishna, N. S., Bhattu, S. N., Somayajulu, D. V., Kumar, N. N., & Reddy, K. J. S. (2022). Gssmilp for anomaly classification in surveillance videos. *Expert Systems with Applications*, *203*, Article 117451.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*, 79–86.

Laptev, D., Savinov, N., Buhmann, J. M., & Pollefeys, M. (2016). Ti-pooling: transformation-invariant pooling for feature learning in convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 289–297).

Le, K. H., Tran, T. V., Pham, H. H., Nguyen, H. T., Le, T. T., & Nguyen, H. Q. (2023). Learning from multiple expert annotators for enhancing anomaly detection in medical image analysis. *IEEE Access*, *11*, 14105–14114.

Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B., Seo, J. B., et al. (2017). Deep learning in medical imaging: general overview. *Korean Journal of Radiology*, *18*, 570–584.

Leistner, C., Saffari, A., & Bischof, H. (2010). Miforests: Multiple-instance learning with randomized trees. In *European conference on computer vision* (pp. 29–42). Springer.

Li, J., Chen, J., Tang, Y., Wang, C., Landman, B. A., & Zhou, S. K. (2023). Transforming medical imaging with transformers? a comparative review of key properties, current progresses, and future perspectives. *Medical Image Analysis*, Article 102762.

Li, C., & Li, G. (2020). Learning multiple instance deep representation for objects tracking. *Journal of Visual Communication and Image Representation*, *71*, Article 102737.

Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H., et al. (2023). Deep learning attention mechanism in medical image analysis: Basics and beyonds. *International Journal of Network Dynamics and Intelligence*, *9*, 3–116.

Li, C., Shi, C., Zhang, H., Chen, Y., & Zhang, S. (2015). Multiple instance learning for computer aided detection and diagnosis of gastric cancer with dual-energy ct imaging. *Journal of Biomedical Informatics*, *57*, 358–368.

Li, X. C., Zhan, D. C., Yang, J. Q., & Shi, Y. (2021). Deep multiple instance selection. *Science China. Information Sciences*, *64*, 1–15.

Lin, T., Xu, H., Yang, C., & Xu, Y. (2022). Interventional multi-instance learning with deconfounded instance-level prediction. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 1601–1609).

Lin, Y., & Zhang, H. (2020). Regularized instance embedding for deep multi-instance learning. *Applied Sciences*, *10*, 64.

Litjens, G. B., et al. (2018). 1399 H & e-stained sentinel lymph node sections of breast cancer patients: the Camelyon dataset. *Applied Intelligence*, *53*, Article giy065.

Liu, Y., Chen, H., Wang, Y., & Zhang, P. (2021). Power pooling: An adaptive pooling function for weakly labelled sound event detection. In *2021 international joint conference on neural networks* (pp. 1–7). IEEE.

Liu, P., Ji, L., Ye, F., & Fu, B. (2024). Advmil: Adversarial multiple instance learning for the survival analysis on whole-slide images. *Medical Image Analysis*, *91*, Article 103020.

Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., & Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, *5*, 555–570.

Manandhar, A., Morton, K. D., Jr., Collins, L. M., & Torrione, P. A. (2012). Multiple instance learning for landmine detection using ground penetrating radar. In *Detection and sensing of mines, explosive objects, and obscured targets XVII* (pp. 668–678). SPIE.

Manzari, O. N., Ahmadabadi, H., Kashiani, H., Shokouhi, S. B., & Ayatollahi, A. (2023). Medvit: a robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, *157*, Article 106791.

Maron, O., & Lozano-Pérez, T. (1997). A framework for multiple-instance learning. *Advances in Neural Information Processing Systems*, 10.

Maron, O., & Ratan, A. L. (1998). Multiple-instance learning for natural scene classification. In *ICML* (pp. 341–349). Citeseer.

Maser, M., Park, J. W., Lin, J. Y. Y., Lee, J. H., Frey, N. C., & Watkins, A. (2023). Supsiam: Non-contrastive auxiliary loss for learning from molecular conformers. arXiv preprint arXiv:2302.07754.

Melendez, J., Van Ginneken, B., Maduskar, P., Philipsen, R. H., Reither, K., Breuninger, M., et al. (2014). A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest x-rays. *IEEE Transactions on Medical Imaging*, *34*, 179–192.

Meng, Y., Bridge, J., Addison, C., Wang, M., Merritt, C., Franks, S., et al. (2023). Bilateral adaptive graph convolutional network on ct based covid-19 diagnosis with uncertainty-aware consensus-assisted multiple instance learning. *Medical Image Analysis*, *84*, Article 102722.

Meyer, D., & Wien, F. (2001). Support vector machines. *R News*, *1*, 23–26.

Miech, A., Alayrac, J. B., Smaira, L., Laptev, I., Sivic, J., & Zisserman, A. (2020). End-to-end learning of visual representations from uncurated instructional videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9879–9889).

Morales-Álvarez, P., Schmidt, A., Hernández-Lobato, J. M., & Molina, R. (2024). Introducing instance label correlation in multiple instance learning, application to cancer detection on histopathological images. *Pattern Recognition*, *146*, Article 110057.

Musalamadugu, T. S., & Kannan, H. (2023). Generative ai for medical imaging analysis and applications. *Future Medicine AI*, FMAI5.

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*, *b4*, 51–62.

Nayak, S. R., & Mishra, J. (2023). Analysis of medical images using fractal geometry. In *Research anthology on improving medical imaging techniques for analysis and intervention* (pp. 1547–1562). IGI Global.

Oner, M. U., Kye-Jet, J. M. S., Lee, H. K., & Sung, W. K. (2023). Distribution based mil pooling filters: Experiments on a lymph node metastases dataset. *Medical Image Analysis*, *87*, Article 102813.

Pal, S., Valkanas, A., Regol, F., & Coates, M. (2022). Bag graph: Multiple instance learning using bayesian graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 7922–7930).

Papandreou, G., Kokkinos, I., & Savalle, P. A. (2015). Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 390–399).

Park, S. W., Ko, J. S., Huh, J. H., & Kim, J. C. (2021). Review on generative adversarial networks: focusing on computer vision and its applications. *Electronics*, *10*, 1216.

Peer, D., Stabinger, S., & Rodriguez-Sanchez, A. (2018). Training deep capsule networks. (pp. 1–7). arXiv preprint arXiv:1812.09707.

Pereira, S. S., & Maia, J. E. B. (2024). Mc-mil: video surveillance anomaly detection with multi-instance learning and multiple overlapped cameras. *Neural Computing and Applications*, 1–17.

Pérez-Cano, J., Wu, Y., Schmidt, A., López-Pérez, M., Morales-Álvarez, P., Molina, R., et al. (2024). An end-to-end approach to combine attention feature extraction and gaussian process models for deep multiple instance learning in ct hemorrhage detection. *Expert Systems with Applications*, *240*, Article 122296.

Pinaya, W. H., Graham, M. S., Kerfoot, E., Tudosiu, P. D., Dafflon, J., Fernandez, V., et al. (2023). Generative ai for medical imaging: extending the monai framework. arXiv preprint arXiv:2307.15208.

Pocevičiūtė, M., Eilertsen, G., Garvin, S., & Lundström, C. (2023). Detecting domain shift in multiple instance learning for digital pathology using Fréchet domain distance. In *International conference on medical image computing and computer-assisted intervention* (pp. 157–167). Springer.

Poursaeidi, M. H., & Kundakcioglu, O. E. (2014). Robust support vector machines for multiple instance learning. *Annals of Operations Research*, *216*, 205–227.

Prince, J. L., & Links, J. M. (2023). Medical imaging signals and systems.

Provost, F. J., Fawcett, T., Kohavi, R., et al. (1998). The case against accuracy estimation for comparing induction algorithms. In *ICML* (pp. 445–453).

Qaiser, T., Winzeck, S., Barfoot, T., Barwick, T., Doran, S. J., Kaiser, M. F., et al. (2021). Multiple instance learning with auxiliary task weighting for multiple myeloma classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 786–796). Springer.

Qi, S., Xu, C., Li, C., Tian, B., Xia, S., Ren, J., et al. (2021). Dr-mil: deep represented multiple instance learning distinguishes covid-19 from community-acquired pneumonia in ct images. *Computer Methods and Programs in Biomedicine*, *211*, Article 106406.

Quellec, G., Cazuguel, G., Cochener, B., & Lamard, M. (2017). Multiple-instance learning for medical image and video analysis. *IEEE Reviews in Biomedical Engineering*, *10*, 213–234.

Quellec, G., Lamard, M., Abràmoff, M. D., Decencière, E., Lay, B., Erginay, A., et al. (2012). A multiple-instance learning framework for diabetic retinopathy screening. *Medical Image Analysis*, *16*, 1228–1240.

Quellec, G., Lamard, M., Cozic, M., Coatrieux, G., & Cazuguel, G. (2016). Multiple-instance learning for anomaly detection in digital mammography. *IEEE Transactions on Medical Imaging*, *35*, 1604–1614.

Rahmani, R., & Goldman, S. A. (2006). Missl: Multiple-instance semi-supervised learning. In *Proceedings of the 23rd international conference on machine learning* (pp. 705–712).

Ramon, J., & De Raedt, L. (2000). Multi instance neural networks. In *Proceedings of the ICML-2000 workshop on attribute-value and relational learning* (pp. 53–60).

Ray, S., & Craven, M. (2005a). Learning statistical models for annotating proteins with function information using biomedical text. *BMC Bioinformatics*, *6*, 1–9.

Ray, S., & Craven, M. (2005b). Supervised versus multiple instance learning: An empirical comparison. In *Proceedings of the 22nd international conference on machine learning* (pp. 697–704).

Redekop, E., Sarma, K. V., Kinnaird, A., Sisk, A., Raman, S. S., Marks, L. S., et al. (2022). Attention-guided prostate lesion localization and grade group classification with multiple instance learning. In *International conference on medical imaging with deep learning* (pp. 975–987). PMLR.

Ren, H., Yang, W., Zhang, T., & Zhang, Y. (2023). Proposal-based multiple instance learning for weakly-supervised temporal action localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2394–2404).

Ren, Q., Zhao, Y., He, B., Wu, B., Mai, S., Xu, F., et al. (2023). Iib-mil: Integrated instance-level and bag-level multiple instances learning with label disambiguation for pathological image analysis. In *International conference on medical image computing and computer-assisted intervention* (pp. 560–569). Springer.

Rymarczyk, D., Borowa, A., Tabor, J., & Zielinski, B. (2021). Kernel self-attention for weakly-supervised image classification using deep multiple instance learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 1721–1730).

Safta, W., & Frigui, H. (2018). Multiple instance learning for benign vs. malignant classification of lung nodules in ct scans. In *2018 IEEE international symposium on signal processing and information technology* (pp. 490–494). IEEE.

Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., et al. (2019). Deep learning in medical imaging and radiation therapy. *Medical Physics*, *46*, e1–e36.

Salton, G. (1971). *The SMART retrieval system—experiments in automatic document processing*. Prentice-Hall, Inc.

Saravanan, R., & Sujatha, P. (2018). A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In *2018 second international conference on intelligent computing and control systems* (pp. 945–949). IEEE.

Schirris, Y., Gavves, E., Nederlof, I., Horlings, H. M., & Teuwen, J. (2022). Deepsmile: Contrastive self-supervised pre-training benefits msi and hrd classification directly from h & e whole-slide images in colorectal and breast cancer. *Medical Image Analysis*, *79*, Article 102464.

Seo, H., Brand, L., Barco, L. S., & Wang, H. (2022). Scaling multi-instance support vector machine to breast cancer detection on the breakhis dataset. *Bioinformatics*, *38*, i92–i100.

Seo, H., & Wang, H. (2023). Fast multi-modal multi-instance support vector machine for fine-grained chest x-ray recognition. In *2023 IEEE international conference on data mining* (pp. 1295–1300). IEEE.

Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., et al. (2023). Transformers in medical imaging: A survey. *Medical Image Analysis*, Article 102802.

Shanmugam, D., Blalock, D., & Guttag, J. (2019). Multiple instance learning for ecg risk stratification. In *Machine learning for healthcare conference* (pp. 124–139). PMLR.

Shao, Z., Wang, Y., Chen, Y., Bian, H., Liu, S., Wang, H., et al. (2023). Lnpl-mil: Learning from noisy pseudo labels for promoting multiple instance learning in whole slide image. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 21495–21505).

Shi, X., Xing, F., Xie, Y., Zhang, Z., Cui, L., & Yang, L. (2020). Loss-based attention for deep multiple instance learning. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 5742–5749).

Shu, S., Wang, D. B., Yuan, S., Wei, H., Jiang, J., Feng, L., et al. (2024). Multiple-instance learning from triplet comparison bags. *ACM Transactions on Knowledge Discovery from Data*, *18*, 1–18.

Sirinukunwattana, K., Raza, et al. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, *35*, 1196–1206.

Srinivasan, A., & Muggleton, e. (1995). Comparing the use of background knowledge by inductive logic programming systems. In *Proceedings of the 5th international workshop on inductive logic programming* (pp. 199–230).

Srinivasan, A., Muggleton, S., King, R. D., & Sternberg, M. J. (1994). Mutagenesis: Ilp experiments in a non-determinate biological domain. In *Proceedings of the 4th international workshop on inductive logic programming* (pp. 127–232). Citeseer.

Stegmüller, T., Bozorgtabar, B., Spahr, A., & Thiran, J. P. (2023). Scorenet: Learning non-uniform attention and augmentation for transformer-based histopathological image classification. (pp. 6170–6179).

Su, Z., Tavolara, T. E., Carreno-Galeano, G., Lee, S. J., Gurcan, M. N., & Niazi, M. (2022). Attention2majority: Weak multiple instance learning for regenerative kidney grading on whole slide images. *Medical Image Analysis*, *79*, Article 102462.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

Tang, W., Huang, S., Zhang, X., Zhou, F., Zhang, Y., & Liu, B. (2023). Multiple instance learning framework with masked hard instance mining for whole slide image classification. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4078–4087).

Tavolara, T. E., Gurcan, M. N., & Niazi, M. K. K. (2022). Contrastive multiple instance learning: An unsupervised framework for learning slide-level representations of whole slide histopathology images without labels. *Cancers*, *14*, 5778.

Tax, D. M., & Duin, R. P. (2008). Learning curves for the analysis of multiple instance classifiers. In *Structural, syntactic, and statistical pattern recognition: joint IAPR international workshop, SSPR & SPR 2008, Orlando, USA, December 4-6, 2008. proceedings* (pp. 724–733). Springer.

Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., et al. (2021). Mlp-mixer: An all-mlp architecture for vision. *Advances in Neural Information Processing Systems*, *34*, 24261–24272.

Tong, T. W., et al. (2014). Multiple instance learning for classification of dementia in brain mri. *Medical Image Analysis*, *18*, 808–818.

Toraman, Ç. (2011). *Text categorization and ensemble pruning in Turkish news portals* (Ph.D. thesis), Bilkent Universitesi (Turkey).

Tourniaire, P., Ilie, M., Hofman, P., Ayache, N., & Delingette, H. (2021). Attention-based multiple instance learning with mixed supervision on the camelyon16 dataset. In *MICCAI workshop on computational pathology* (pp. 216–226). PMLR.

Tu, M., Huang, J., He, X., & Zhou, B. (2019). Multiple instance learning with graph neural networks. arXiv preprint arXiv:1906.04881.

Vasen, H. F., Wasser, M., Van Mil, A., Tollenaar, R. A., Konstantinovski, M., Gruis, N. A., et al. (2011). Magnetic resonance imaging surveillance detects early-stage pancreatic cancer in carriers of a p16-leiden mutation. *Gastroenterology*, *140*, 850–856.

Vezhnevets, A., & Buhmann, J. M. (2010). Towards weakly supervised semantic segmentation by means of multiple instance and multitask learning. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 3249–3256). IEEE.

Wan, F., Liu, C., Ke, W., Ji, X., Jiao, J., & Ye, Q. (2019). C-mil: Continuation multiple instance learning for weakly supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2199–2208).

Wang, J., Cherian, A., Porikli, F., & Gould, S. (2018). Video representation learning using discriminative pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1149–1158).

Wang, Y., Li, J., & Metze, F. (2019). A comparison of five multiple instance learning pooling functions for sound event detection with weak labeling. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing* (pp. 31–35). IEEE.

Wang, S., McKenna, M. T., Nguyen, T. B., Burns, J. E., Petrick, N., Sahiner, B., et al. (2012). Seeing is believing: Video classification for computed tomographic colonography using multiple-instance learning. *IEEE Transactions on Medical Imaging*, *31*, 1141–1153.

Wang, X., Yan, Y., Tang, P., Bai, X., & Liu, W. (2018). Revisiting multiple instance neural networks. *Pattern Recognition*, *74*, 15–24.

Wang, Z., & Yuan, J. (2018). Unsupervised multiple-instance learning for instance search. In *2018 IEEE international conference on multimedia and expo* (pp. 1–6). IEEE.

Wang, J., & Zucker, J. d. (2000). Solving the multiple-instance problem: A lazy learning approach. In *International conference on machine learning* (pp. 1119–1126).

Waqas, M., Khan, Z., Ahmed, S. U., & Raza, A. (2023). Mil-mixer: A robust bag encoding strategy for multiple instance learning (mil) using mlp-mixer. In *2023 18th international conference on emerging technologies* (pp. 22–26). IEEE.

Waqas, M., Khan, Z., Anjum, S., & Tahir, M. A. (2020). Lung-wise tuberculosis analysis and automatic ct report generation with hybrid feature and ensemble learning. In *CLEF (working notes)* (pp. 1–10).

Waqas, M., Tahir, M. A., Danish, M., Al-Maadeed, S., Bouridane, A., & Wu, J. (2024). Simultaneous instance pooling and bag representation selection approach for multiple-instance learning (mil) using vision transformer. *Neural Computing and Applications*, 1–22.

Waqas, M., Tahir, M. A., & Khan, S. A. (2023). Robust bag classification approach for multi-instance learning via subspace fuzzy clustering. *Expert Systems with Applications*, *214*, Article 119113. http://dx.doi.org/10.1016/j.eswa.2022.119113.

Waqas, M., Tahir, M. A., & Qureshi, R. (2021). Ensemble-based instance relevance estimation in multiple-instance learning. In *2021 9th European workshop on visual information processing* (pp. 1–6). IEEE.

Waqas, M., Tahir, M. A., & Qureshi, R. (2023). Deep Gaussian mixture model based instance relevance estimation for multiple instance learning applications. (pp. 10310–10325).

Wei, X. S., Wu, J., & Zhou, Z. H. (2014). Scalable multi-instance learning. In *2014 IEEE international conference on data mining* (pp. 1037–1042). IEEE.

Wei, X. S., Wu, J., & Zhou, Z. H. (2016). Scalable algorithms for multi-instance learning. *IEEE Transactions on Neural Networks and Learning Systems*, *28*, 975–987.

Wei, X. S., & Zhou, Z. H. (2016). An empirical study on image bag generators for multi-instance learning. *Machine Learning*, *105*, 155–198.

Wu, Y., Schmidt, A., Hernández-Sánchez, E., Molina, R., & Katsaggelos, A. K. (2021). Combining attention-based multiple instance learning and gaussian processes for ct hemorrhage detection. In *International conference on medical image computing and computer-assisted intervention* (pp. 582–591). Springer.

Xiao, Y., Jin, Y., Cheng, R., & Hao, K. (2022). Hybrid attention-based transformer block model for distant supervision relation extraction. *Neurocomputing*, *470*, 29–39.

Xiao, Y., Liang, F., & Liu, B. (2020). A transfer learning-based multi-instance learning method with weak labels. *IEEE Transactions on Cybernetics*, *52*, 287–300.

Xiong, J., Li, Z., Wang, G., Fu, Z., Zhong, F., Xu, T., et al. (2022). Multi-instance learning of graph neural networks for aqueous p k a prediction. *Bioinformatics*, *38*, 792–798.

Xu, X. (2001). *A nearest distribution approach to multiple-instance learning*: *Rapport technique*, Hamilton, New Zealand: Department of Computer Science, University of Waikato.

Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Eric, I., et al. (2014). Deep learning of feature representation with multiple instance learning for medical image analysis. In *2014 IEEE international conference on acoustics, speech and signal processing* (pp. 1626–1630). IEEE.

Yan, R., Shen, Y., Zhang, X., Xu, P., Wang, J., Li, J., et al. (2023). Histopathological bladder cancer gene mutation prediction with hierarchical deep multiple-instance learning. *Medical Image Analysis*, *87*, Article 102824.

Yan, Y., Wang, X., Guo, X., Fang, J., Liu, W., & Huang, J. (2018). Deep multi-instance learning with dynamic pooling. In *Asian conference on machine learning* (pp. 662–677). PMLR.

Yang, Y., Tu, Y., Lei, H., & Long, W. (2023). Hamil: Hierarchical aggregation-based multi-instance learning for microscopy image classification. *Pattern Recognition*, *136*, Article 109245.

Yu, S., Ma, K., Bi, Q., Bian, C., Ning, M., He, N., et al. (2021). Mil-vt: Multiple instance learning enhanced vision transformer for fundus image classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 45–54). Springer.

Yu, J. G., Wu, Z., Ming, Y., Deng, S., Li, Y., Ou, C., et al. (2023). Prototypical multiple instance learning for predicting lymph node metastasis of breast cancer from whole-slide pathological images. *Medical Image Analysis*, *85*, Article 102748.

Yuan, L., Xu, G., Zhao, L., Wen, X., & Xu, H. (2020). Multiple-instance learning via multiple-point concept based instance selection. *International Journal of Machine Learning and Cybernetics*, *11*, 2113–2126.

Zhang, D., Duan, Y., Guo, J., Wang, Y., Yang, Y., Li, Z., et al. (2022). Using multi-scale convolutional neural network based on multi-instance learning to predict the efficacy of neoadjuvant chemoradiotherapy for rectal cancer. *IEEE Journal of Translational Engineering in Health and Medicine*, *10*, 1–8.

Zhang, Q., & Goldman, S. A. (2001). Em-dd: An improved multiple-instance learning technique. In *Advances in neural information processing systems* (pp. 1073–1080).

Zhang, D., Han, J., Cheng, G., & Yang, M. H. (2021). Weakly supervised object localization and detection: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*, 5866–5885.

Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S. E., et al. (2022). Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18802–18812).

Zhang, C., Platt, J., & Viola, P. (2005). Multiple instance boosting for object detection. *Advances in Neural Information Processing Systems*, 18.

Zhang, D., Wang, F., Si, L., & Li, T. (2011). Maximum margin multiple instance clustering with applications to image and text clustering. *IEEE Transactions on Neural Networks*, *22*, 739–751.

Zhang, J., Wu, Y., Hao, F., Liu, X., Li, M., Zhou, D., et al. (2024). Double similarities weighted multi-instance learning kernel and its application. *Expert Systems with Applications*, *238*, Article 121900.

Zhang, X., Xu, Y., & Liu, X. (2024). Multiple instance learning from similarity-confidence bags. *Pattern Recognition*, *146*, Article 109984.

Zhang, M. L., & Zhou, Z. H. (2009). Multi-instance clustering with applications to multi-instance prediction. *Applied Intelligence*, *31*, 47–68.

Zhang, Y. L., & Zhou, Z. H. (2017). Multi-instance learning with key instance shift. In *IJCAI* (pp. 3441–3447).

Zhao, Y., Yang, F., Fang, Y., Liu, H., Zhou, N., Zhang, J., et al. (2020). Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4837–4846).

Zheng, B., Zhu, Y., Shi, Q., Yang, D., Shao, Y., & Xu, T. (2022). Ma-net: Mutex attention network for covid-19 diagnosis on ct images. *Applied Intelligence*, 1–16.

Zhou, Z. H. (2004). *Multi-instance learning: a survey*: *Tech. rep 1*, Department of Computer Science & Technology, Nanjing University.

Zhou, Z. H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, *5*, 44–53.

Zhou, Z. H., Jiang, K., & Li, M. (2005). Multi-instance learning based web mining. *Applied Intelligence*, *22*, 135–147.

Zhou, Z. H., Sun, Y. Y., & Li, Y. F. (2009). Multi-instance learning by treating instances as non-iid samples. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1249–1256).

Zhou, Y., Sun, X., Liu, D., Zha, Z., & Zeng, W. (2017). Adaptive pooling in multi-instance learning for web video annotation. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 318–327).

Zhou, Z. H., & Xu, J. M. (2007). On the relation between multi-instance learning and semi-supervised learning. In *Proceedings of the 24th international conference on machine learning* (pp. 1167–1174).

Zielinski, B., & Sroka-Oleksiak, e. (2019). Deep learning approach to description and classification of fungi microscopic images. CoRR.