electronics

MDPI

Review

A Comprehensive Review on Multiple Instance Learning

Samman Fatima [1], Sikandar Ali [1] and Hee-Cheol Kim [2,*]

[1] Department of Digital Anti-Aging Healthcare, Inje University, Gimhae 50834, Republic of Korea; samman.1511@gmail.com (S.F.); sikandarshigri77@gmail.com (S.A.)
[2] Institute of Digital Anti-Aging Healthcare, College of AI Convergence, u-AHRC, Inje University, Gimhae 50834, Republic of Korea
* Correspondence: heeki@inje.ac.kr

Abstract: Multiple-instance learning has become popular over recent years due to its use in some special scenarios. It is basically a type of weakly supervised learning where the learning dataset contains bags of instances instead of a single feature vector. Each bag is associated with a single label. This type of learning is flexible and a natural fit for multiple real-world problems. MIL has been employed to deal with a number of challenges, including object detection and identification tasks, content-based image retrieval, and computer-aided diagnosis. Medical image analysis and drug activity prediction have been the main uses of MIL in biomedical research. Many Algorithms based on MIL have been put forth over the years. In this paper, we will discuss MIL, the background of MIL and its application in multiple domains, some MIL-based methods, challenges, and lastly, the conclusions and prospects.

Keywords: artificial intelligence; deep learning; multiple instance learning; weakly supervised learning

Citation: Fatima, S.; Ali, S.; Kim, H.-C. A Comprehensive Review on Multiple Instance Learning. *Electronics* **2023**, *12*, 4323. https://doi.org/10.3390/electronics12204323

Academic Editors: Giovanni Pau and Xiangjie Kong

Received: 2 October 2023
Revised: 16 October 2023
Accepted: 16 October 2023
Published: 18 October 2023

1. Introduction

In machine learning, basically, a computer program is given some tasks to complete; if the computer program's measured performance on these tasks improves as it obtains more and more experience completing these tasks, it is claimed that the machine has learned from its experience. As a result, the machine makes decisions and predictions according to data. Traditional machine learning has three major segments, namely supervised learning, unsupervised learning, and reinforcement learning. Supervised machine learning is a subset of machine learning in which the algorithm learns using labeled training data. In this method, the model is given input data and labels for the expected outputs. For the model to accurately predict future events or categorize previously unidentified data, it must understand the relationship between inputs and outputs. In other words, when training instances have known labels, and there is consequently the least amount of ambiguity, supervised learning datasets are based on labeled inputs and their corresponding outputs, which seeks to develop a notion for accurately identifying unknown occurrences. On the other hand, Unsupervised machine learning is a sort of machine learning in which the algorithm is tasked with discovering patterns, structures, or groupings within the data on its own after being provided unlabeled data. There are no predetermined output labels to direct the learning process, in contrast to supervised learning. Instead, using methods like clustering or dimensionality reduction, the program looks for inherent relationships or commonalities between the data points. In short, when the training instances do not have labels, and there is consequently the greatest amount of uncertainty, unsupervised learning tries to understand the structure of the underlying patterns of instances. Algorithms for reinforcement learning (RL) use the learning approach by interacting with the environment (sequences of actions, observations, and rewards). Robotics and resource allocation are two areas where RL-based techniques have demonstrated outstanding performance. These

have made them one of the most promising prospects for achieving the aim of artificial intelligence (AI), creating autonomous entities that can learn in complex and unknowable contexts [1].

The amount of data required to handle significant problems has grown tremendously in recent years. A considerable amount of labeling work is necessary for large amounts of data. Since weak supervision is typically easier to get, approaches with weak supervision, like MIL, might lessen this load. For instance, object detectors could be trained to utilize web-sourced images and their associated labels as weak supervision instead of manually labeled data sets. Instead of spending money and time on expensive manual annotations, which can be only provided by experts, as the case may be in medical images for which only patient diagnoses are accessible, can be used to train computer-aided diagnosis algorithms; MIL enables the use of partially annotated data to complete the tasks with fewer resources.

The robotics industry, virtual assistants (for example, Google, etc.), video games, pattern recognition, natural language processing (NLP), data mining, traffic forecasting, online public transportation systems (one example is predicting surge prices by the Uber app during peak hours), product recommendation, share market prediction, healthcare diagnosis, online fraud detection, and search engine result prediction and refinement (for example, Google search results) are just a few of the fields where machine learning is used [2].

MIL is a type of weakly supervised learning. Training data are arranged in groupings called bags for multiple-instance learning (MIL), which uses these data. Only complete sets are subject to supervision; the individual labels of the instances contained in the bags are not made available. The research community has given this problem formulation much attention, especially in the last few years when the amount of data required to address major problems has proliferated. A significant amount of labeling work is required due to the large amounts of data. As stated earlier, in MIL, inputs are arranged in bags, and each bag has multiple instances/inputs. A single label is associated with a bag full of instances rather than every single instance. Unlike supervised learning, where all instances have predefined labels, multi-instance learning uses training instances with unknown and ambiguous labels. This arrangement of MIL is gaining popularity because of its flexibility as it leverages weakly/ambiguously annotated data [3]. Figure 1 shows the concept of single instance learning and multi-instance learning [4].

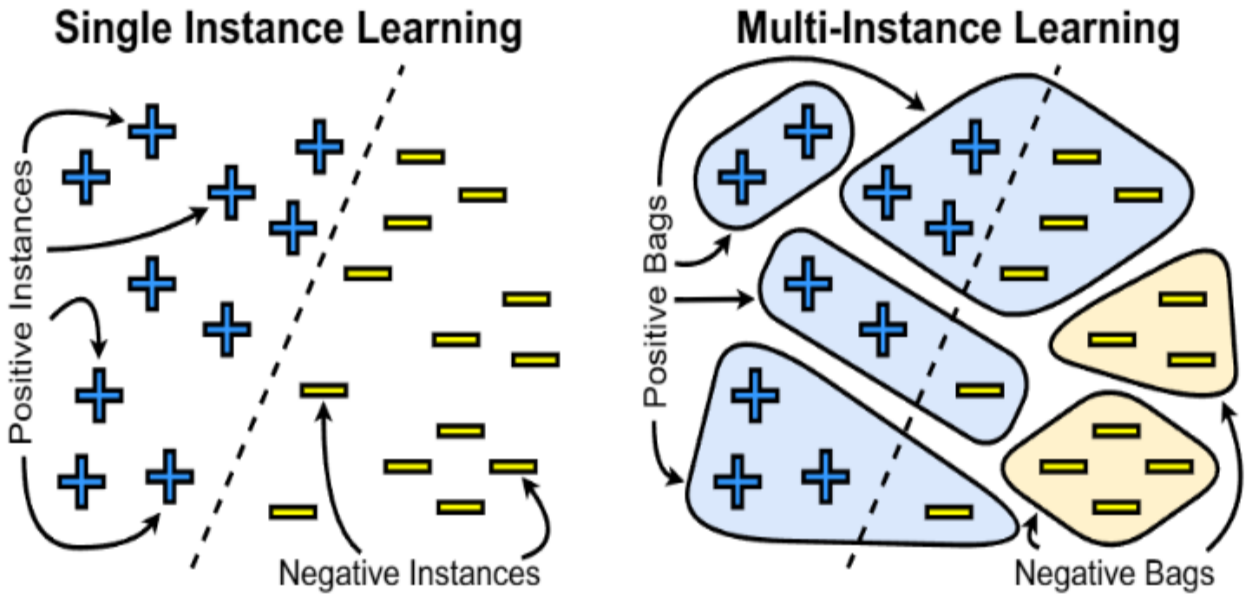

Figure 1. Difference between Single and Multi-instance learning. Reprinted with permission from Ref. [4]. Copyright 2023 Springer Nature.

Multi-instance learning has garnered much interest from the machine learning community since multi-instance problems are widespread yet distinct from those handled by prior learning frameworks. A bag in MIL is marked as positive as positive if there is at least a single positive instance and labeled as negative if all of the instances in the bag are negative [5], as illustrated in Figure 2.

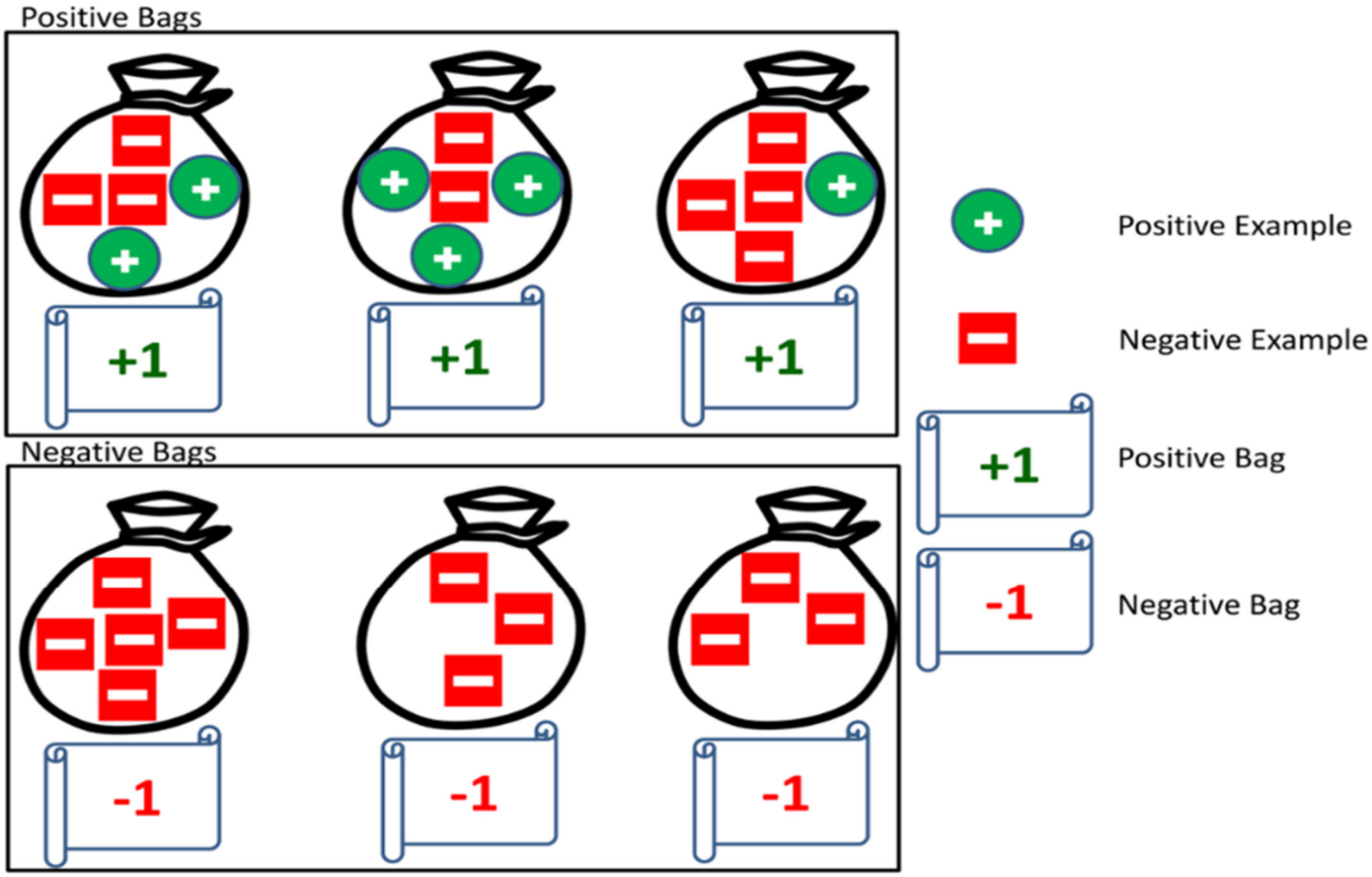

Figure 2. Positive and negative bags [5].

The main goal of applying MIL is to classify and correctly predict unseen bags of data based on the training data (labeled data) [6]. According to this paradigm, there is some ambiguity in the data on how the labels were assigned. As mentioned above, labels are given to sets or bags of inputs instead of label pairs, which is how the learning process is fed. The MIL assumes that no less than one positive input is present in every positive bag, which limits the labels to binary. Since the true input labels are unknown during training, they can be seen as latent variables. Contrary to reinforcement learning, where there is a delay in the labeling of the training instances, multi-instance learning does not have any such delay. Popular learning techniques like decision trees and neural networks, which ignore the properties of multi-instance issues, have been found to be ineffective in this situation [7]. Consider the simple example for a better understanding of the MIL concept, shown in Figure 3 below. There are some keychains with multiple keys. People can enter a special room with the use of a special key. Some people have access to that particular room, while others do not. We must first find that specific key among all the "positive" keychains in order to unravel the puzzle of who would be able to enter the room. The only way we can adequately categorize all the keychains is if we can locate the specific key. Therefore, keychains containing that specific key are positive, whilst others are negative [8,9]; in this example shown in Figure 3, that key is the green key.

Figure 3. Keychain example for MIL.

In addition, a variety of problems can be naturally phrased as MIL problems. For instance, the aim of the drug activity prediction issue is to foretell whether a given molecule will cause a specific effect. A molecule can adopt a variety of forms that either result in the desired action or not. It is impossible to observe how different conformations affect one another. As a result, molecules must be viewed as a collection of shapes, which is why the MIL formulation is used. Over the past 20 years, MIL has been employed more and more in many different application sectors because of its appealing qualities, including image and video classification, document and sound classification, sound classification, content-based image retrieval, and face recognition. Apart from this, MIL is being used in various other disciplines, such as medical imaging. Multiple instance learning is utilized to recognize and classify cancers on whole slide images (WSI). In natural language processing, it is also possible to classify documents based on their content [10].

In this paper, we will review various domains where MIL is applicable, along with some MIL algorithms, some challenges while implementing MIL, and the conclusions. The key contribution of this review is to create awareness by exposing the various application areas in which MIL can be utilized for easy problem-solving in case of partially labeled data availability scenarios. Specifically, we carried out an inventory of all the existing MIL methods in their various application domains, on which we demonstrated the potential of MIL to revolutionize problem-solving in a partially labeled data context.

2. Background Knowledge

The challenge of correctly predicting the degree of activity of medicinal drug molecules served as the initial inspiration for the MIL study. Following that, numerous MIL techniques were put out, including learning axis-parallel concepts Dietterich et al. [10], extended Citation (kNN) k-nearest neighbors (Wang and Zucker) [11], and others. They have been used for various tasks, from stock market forecasting to text classification and image concept learning.

Dietterich et al. [10] looked at the issue of drug activity prediction in the early 1990s. The objective was to give learning systems the ability to determine via analysis of a database of existing molecules whether a novel molecule was suitable for producing a certain medicine. The majority of medications or drugs are composed of tiny molecules that act by attaching to bigger protein molecules like enzymes and cell-surface receptors. One of the molecules' low-energy configurations that are eligible to make drugs can bind to the intended area tightly, whereas none of the low-energy configurations of molecules that are not capable of making drugs can. The main challenge in predicting drug activity is that every molecule could have a variety of alternative low-energy formations, as shown in Figure 4. However, biochemists only knew whether a molecule was qualified to make a drug or not which of its alternative low-energy shapes would result in qualification [5].

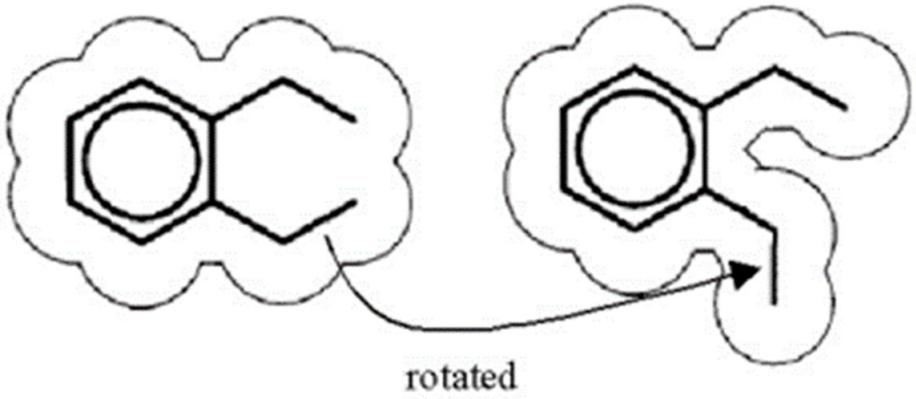

Figure 4. Changing the shape of molecule w.r.t rotating internal bond [5].

By treating all of the low-energy configurations of the "good" molecules as positive instances and all of the low-energy configurations of the "bad" molecules as negative ones, supervised learning algorithms provide a clear answer. A "good" molecule can have dozens of low-energy configurations, but only one of them may actually be a "good" shape, as demonstrated by Dietterich et al. [10], leading to a large false positive noise that makes it difficult for such an approach to be effective.

This multiple-instance learning problem was resolved using the APR algorithm. An axis-parallel hyper-rectangle (APR) is looked up in the feature space. This APR should, intuitively, include a minimum of one instance from every positive bag whilst dismissing all cases from the negative bags. It attempts to locate suitable rectangles with parallel axes by combining the qualities. To locate such a hyper-rectangle, Dietterich et al. [10] proposed three algorithms: An "outside-in" algorithm builds the smallest APR that encapsulates all instances in positive bags after that shrinks the APR to exclude false positives; The smallest APR that confines all instances from positive bags is determined using a "standard" algorithm; an "inside-out" algorithm begins with a seed point and afterward grows a rectangle from it intending to find the smallest APR that covers a minimum of one instance per positive bag but no instances from negative bags. The approach was tested using the Musk dataset, which serves as a real-world test set for drug activity prediction and the truly well-known benchmark in multiple-instance learning. The APR algorithm "inside out" delivered the greatest outcomes despite being created with Musk data in mind. Likewise, Qi Wang et al. [12] investigated saliency detection using multiple instance learning. Furthermore, MIL can also be used for the classification of histopathology breast cancer images [13].

Multi-instance learning is something that is not just limited to drug discovery. MIL was initially used for image classification and categorization by Maron and Lozano-Perez [14] and later created the Diverse Density framework in 1998. One or more fixed-size sub-images are referred to as an instance of an image, and the entire picture is referred to as the bag of instances. An image is classified as positive if it embodies the intended scene, such as a waterfall; otherwise, it is classified as negative. One can employ Multiple instance learning to discover the traits of the sub-images that make up the target scene. Since then, various tasks, such as text classification and stock market forecasting, have been carried out using these frameworks.

Danyi Xiong et al. [15] applied MIL for the detection of cancer using T-cell Receptor Sequencing, also known as TCR sequencing. TCR sequences help distinguish cancer cells from normal tissues and reflect a person's T-cell immunity system, which explains whether the cancer cells are increasing in the body or not. In the human body(bag), many T-cells contain different TCRs(instances). It is possible to use TCR structural properties to determine whether or not a patient has a tumor. TCR reconstruction software like TRUST. A computational tool called TCR Receptor Utilities for Solid Tissue uses unselected RNA sequencing data profiles from solid tissues, including malignancies, to assess TCR

sequences and MiTCR (An open-source program called MiTCR analyzes hundreds of millions of raw high-throughput sequencing reads, including sequences encoding human or mouse T-cell antigen receptor (TCR) chains quickly, thoroughly, and robustly) is used to identify the TCR sequences found in each of the sample from its raw sequencing reads.

Each and every sample is viewed as a bag of TCR sequences (instances), which, in the MIL architecture, practically function as text strings. Using the Tessa model, which includes a deep learning auto-encoder that converts intricate information (such as strings or sequences of amino acids) into numerical values, each TCR sequence is embedded into a numeric vector. When tumor-specific TCRs are present in a bag (sample), the bag becomes positive (a tumor sample), whereas negative TCRs are produced by host immune responses that are not caused by cancer, nonetheless, by some other physiological processes like autoimmune disorders infection [15]. The data processing flow of this MIL application for TCR-based cancer detection is mentioned in Figure 5 below.

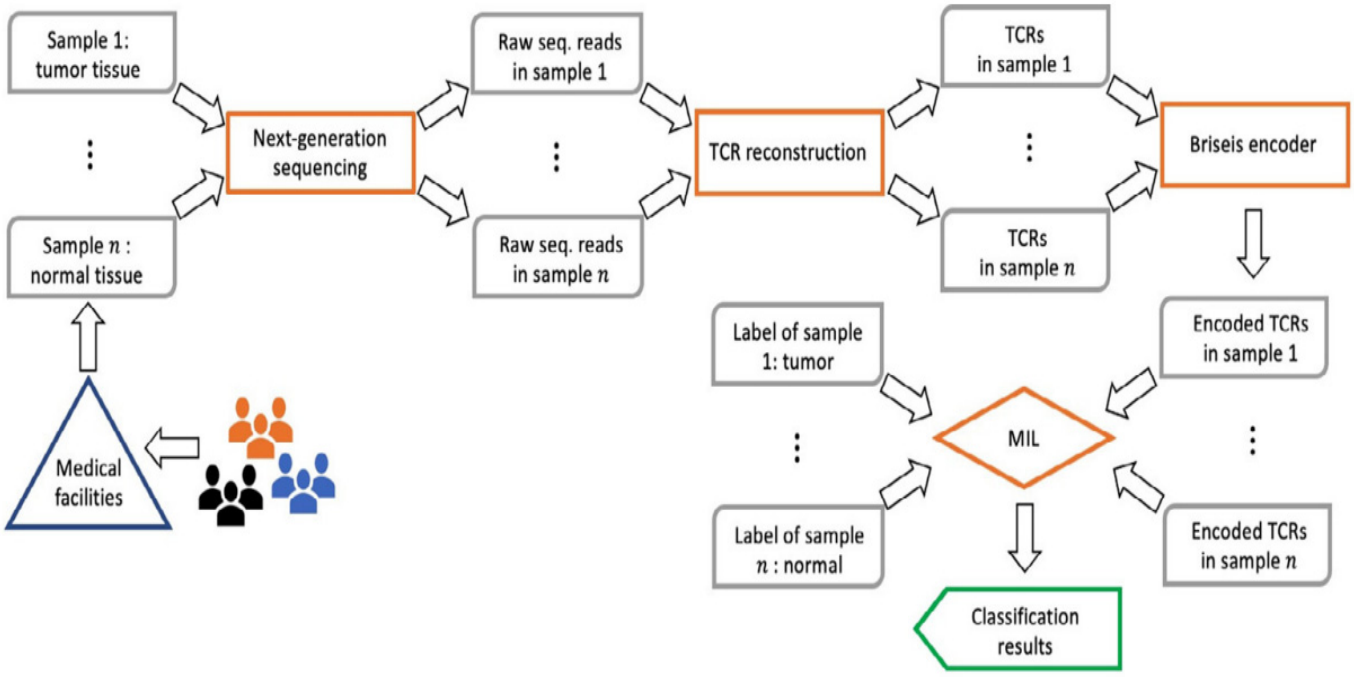

Figure 5. Data processing flow of MIL application for TCR-based cancer detection (The auto-encoder learns a d-dimensional numeric feature vector to represent each encoded TCR sequence). Reprinted with permission from Ref. [15]. Copyright 2023 Elsevier.

Yan Xu et al. [16] explored the potential of multiple instance learning for the segmentation and classification of histopathology cancer images. Moreover, MIL can also be employed for diabetic retinopathy for retinal vessel segmentation [17]. Using a simple grid sampling technique is challenging, as Marc and Veronica [18] showed, especially when the receptive field is small compared to the size of the image's key characteristics. To overcome the problems with grid sampling, they employed a sequential Monte Carlo sampling procedure for high-resolution images, sampling from the most pertinent regions throughout the training phase. Using two simulated and two histological datasets, they showed their competence for breast cancer and sun exposure categorization.

Maoying Qiao et al. [19] developed a supervised learning technique that utilized a variety of MIL-diverse dictionaries to connect representations at the instance level to labels on bags. The suggested technique makes use of labels at the level of bag data to train class-oriented dictionaries. The suggested technique incorporates a diversity Regularizer to prevent ambiguity between the class-specific dictionaries. This is considered the first example in which the diversity prior has been used to address MIL issues.

Stefanos and Lapata [20] proposed a neural network that gains the ability to predict the sentiment of different text slices or segments, such as sentences or elemental discourse units (EDUs), without segment-level supervision after being trained on document sentiment labels. They also present a new dataset called SpoT (short for Segment-level POlariTy annotations) for assessing MIL-based sentiment models, as well as an attention-based polarity score system for distinguishing between positive and negative text samples. A judgment elicitation study demonstrates that opinion extraction at the EDU level delivers more useful summaries than sentence-based alternatives [20]. At the same time, experimental findings show improved performance compared to numerous baselines. A sample of a 2-star review based on EDU (elemental discourse unit), positive and negative snippets is given in Figure 6.

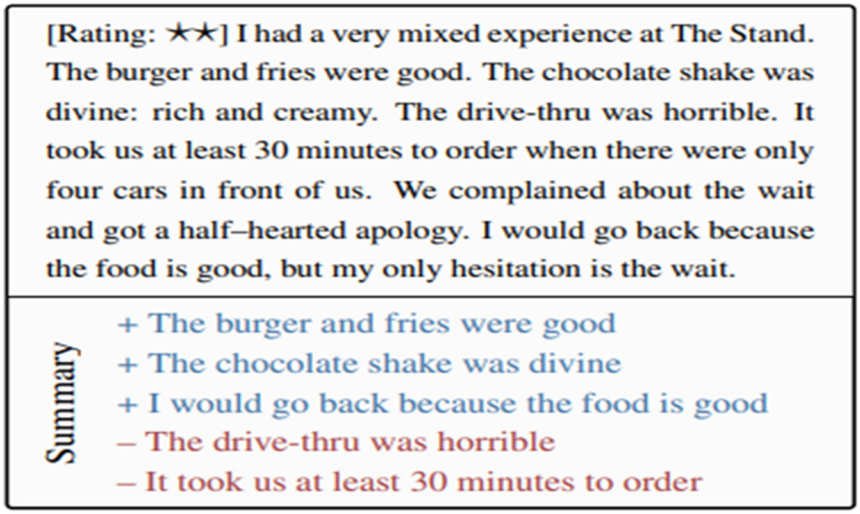

Figure 6. A summary of 2-star review based on EDU (elemental discourse unit), positive and negative snippets [20].

Michael et al. [21] explained the use of MIL in the field of pathology. The tremendous amount of information included in digital whole-slide images is a major driving force behind the creation of automated image analysis technologies. Concerning a variety of tasks in the field of digital pathology, deep neural networks, in particular, exhibit great potential. A drawback of most deep learning algorithms is that they need (human) annotations in addition to the massive volumes of visual data to perform efficient training.

Without fully annotated data, multiple-instance learning demonstrates its potency as a method for learning deep neural networks. Because labels for a complete whole slide image are frequently taken routinely, but labels for patches, areas, or pixels that's why MIL methods are very practical in this field. Whole slide images is divided into small patches so that MIL could be applied [22]. Figure 7 shows how a whole slide image is divided into patches using MIL.

Already, a significant number of publications the majority of which were released in the last three years, have been produced due to this potential. The availability of potent graphics processing units shows a rise in this field, in addition to the availability of data and a high incentive from the medical standpoint. The fundamentals of deep multiple-instance learning systems are extensively and successfully employed [21]. A complete depiction of how MIL from an advanced level is applied to WSIs. Patches that are recovered from the input images are presented in Figure 8.

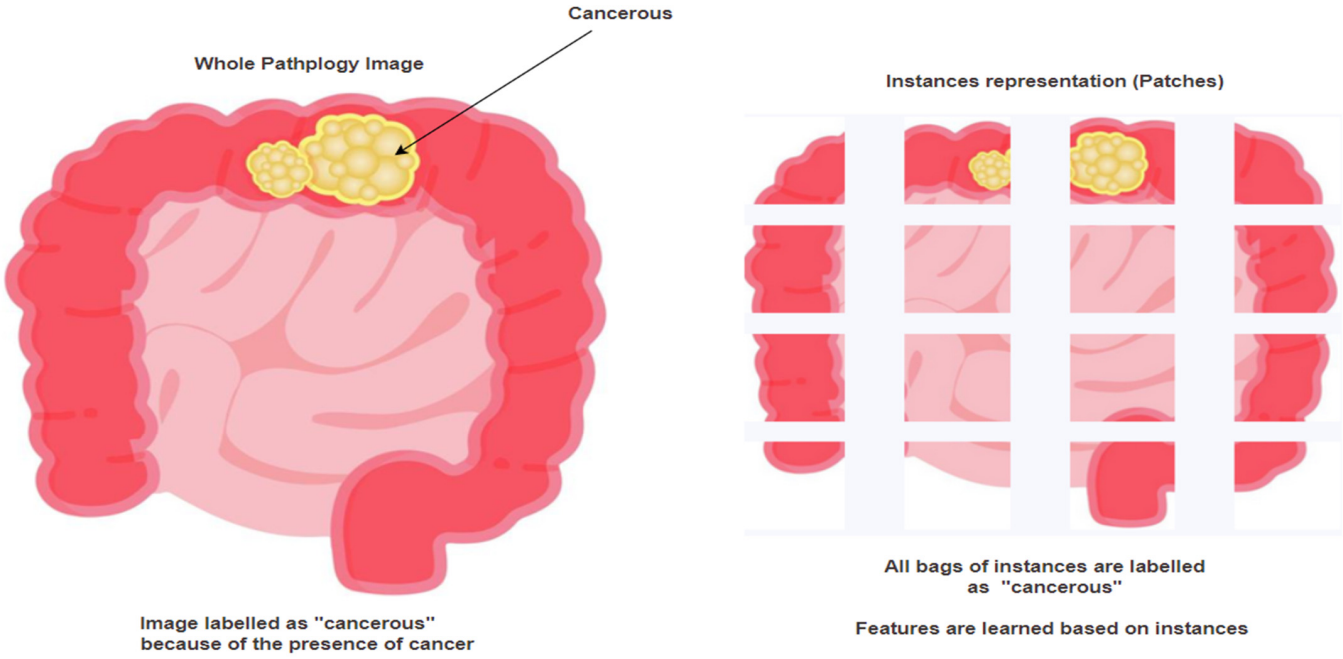

Figure 7. Difference between whole slide and patch image.

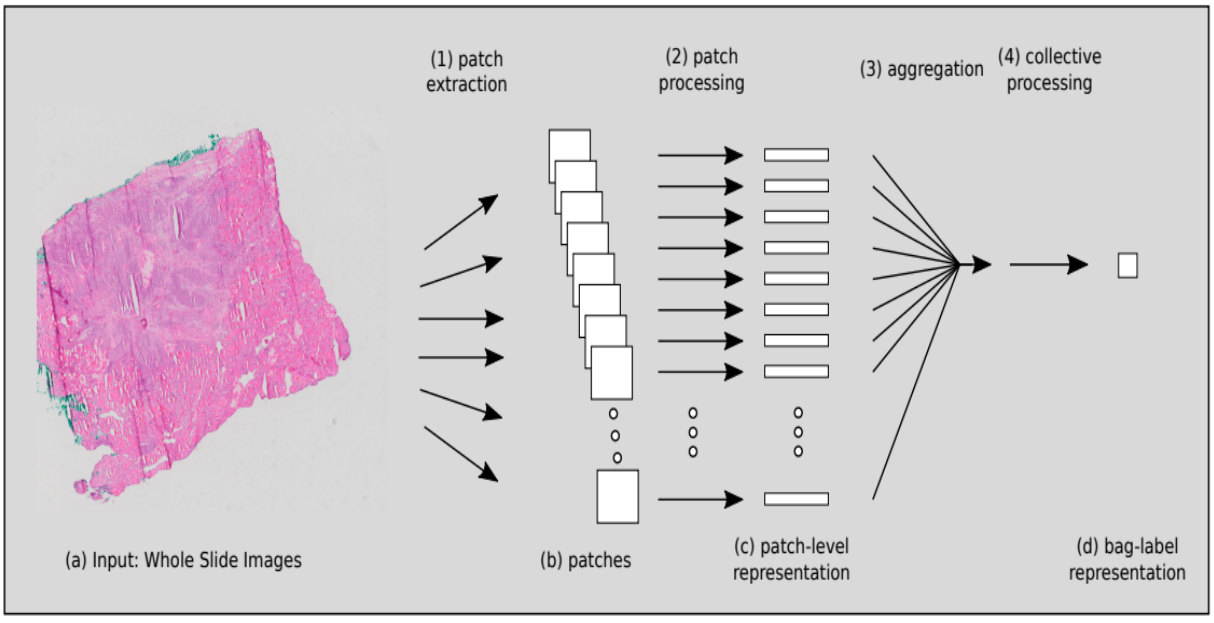

Figure 8. In [21], MIL from an advanced level is applied to WSIs. Patches (**b**) are recovered from the input images (**a**), followed by extraction of patch-level features (**c**), aggregation (converting several patch-level features into a single bag-level feature), and (**d**) collective processing (resulting in bag-level representations).

Xiangfa Song et al. [23] proposed a unique approach built on sparse coding along with a classifier ensemble for the purpose of addressing the image categorization/classification problems inside the multi-instance forming (MIL) framework. In particular, a dictionary is acquired from all of the training bags' instances. A sparse linear combination of each basis vector in the dictionary is used to represent each instance of a bag, and the bag is also represented by a single feature vector that is created by sparse representations of every instance within the bag as shown in Figure 9.

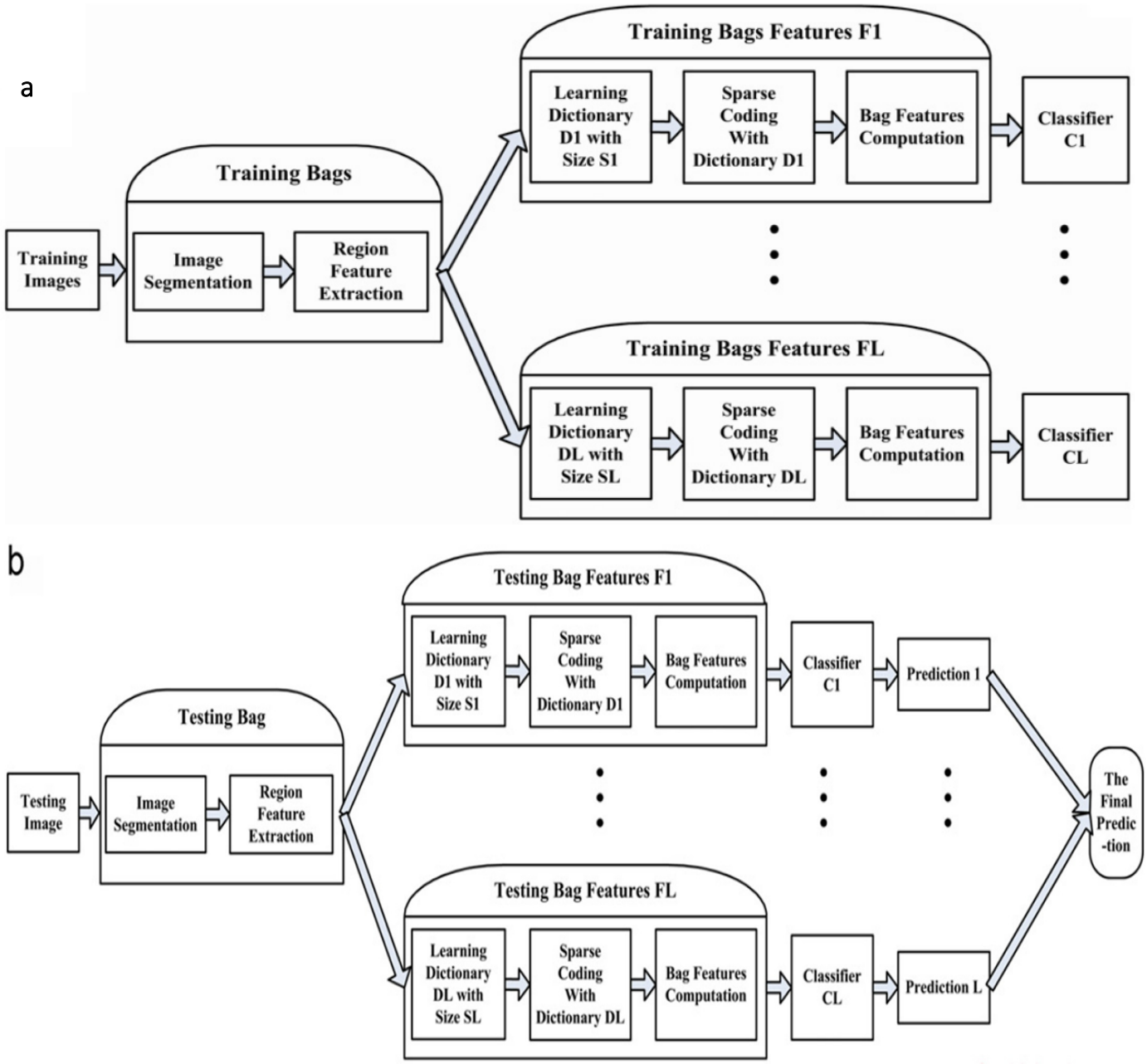

Figure 9. In [23], The Sparse coding and classifier ensemble—MIL method's framework is presented as follows: phase one (training (**a**)) and phase two (testing (**b**)). Reprinted with permission from Ref. [23]. Copyright 2023 Elsevier.

Annabella et al. [24] applied a Multiple Instance Learning (MIL) method (Lagrangian relaxation) that is appropriate for use in image processing/classification applications. The approach relies particularly on a mixed integer nonlinear formulation of the optimization problem that needs to be resolved for MIL. The algorithm categorizes the images that exhibit a particular pattern on a series of color images (Red, Green, Blue, RGB). They evaluated this technique on an artificial dataset of 100 images, where the discriminant was the presence of yellow. Even though these are early results, they seem promising and reassuring for the design of future more sophisticated segmentation systems.

Weiss and Hirsh [25] suggested that a certain sort of time series analysis problem may be solved using the multi-instance learning framework by transforming event prediction into a multi-instance problem [26].

Ruffo et al. [27] employed Relic, a multi-instance decision tree, to address network management, intrusion detection, and password verification problems related to computer security [28].

We also performed a bibliometric analysis of Multiple instance learning using some of the highly relevant keywords like deep learning, self-supervised learning, weakly supervised learning image, MIL algorithm, etc. The knowledge graph is presented in Figure 10.

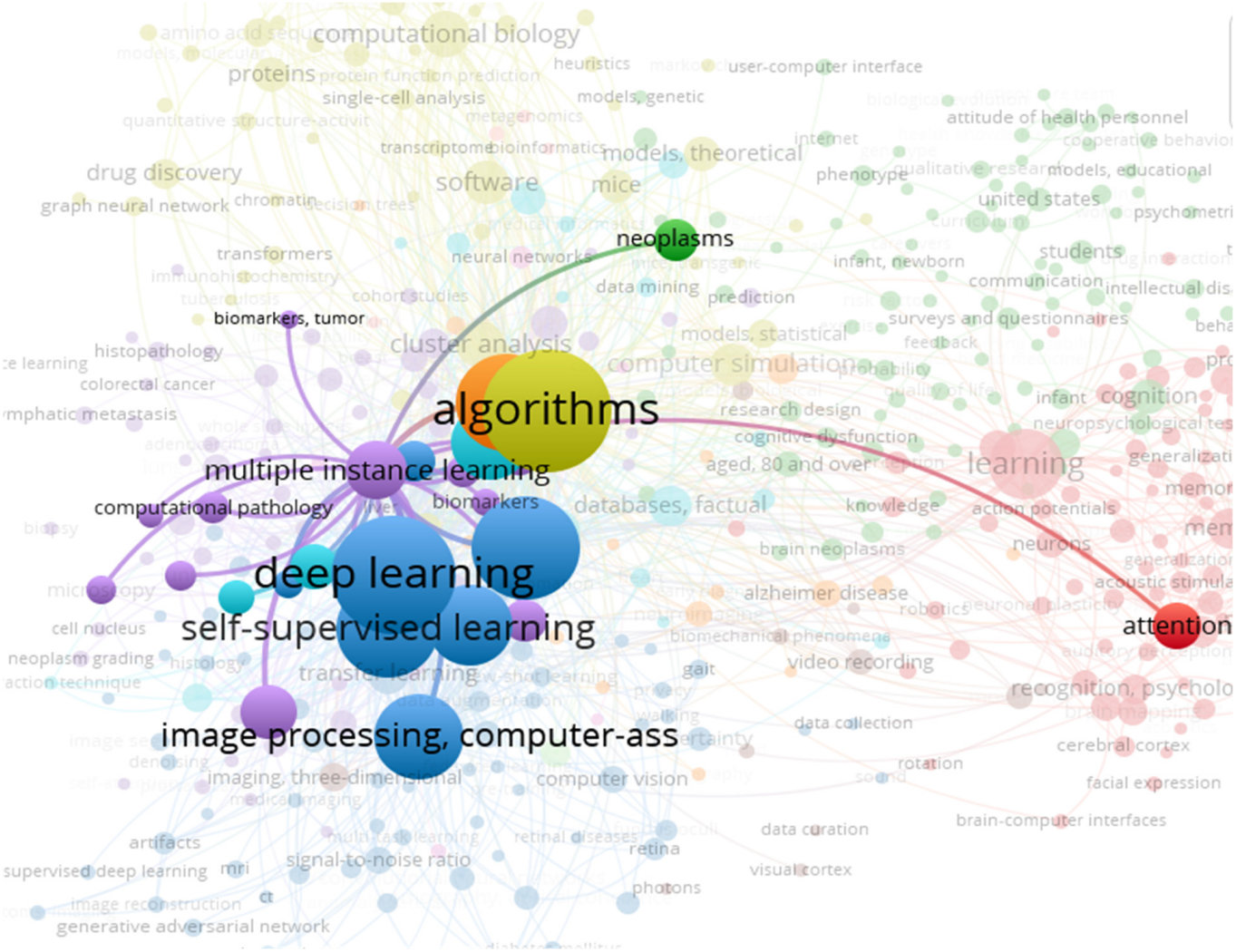

Figure 10. Knowledge graph of Multiple instance learning.

3. Multiple-Instance Learning (MIL) Methods

There are multiple types of MIL methods. These categories include Instance-space methods (IS), bag-space (BS), and embedded-space (EB) methods. Depending on how a method uses the knowledge and information that is extracted and exploited from the MI data, they are classified. These MIL methods are shown in Figure 11.

In Instance space methods, instance-level learning takes place, where f(x) is trained to distinguish between positive and negative bag instances. Instance-level scores calculated by f(x) are then merged to produce a classifier for bag-level F(X), in line with a plausible MI assumption. In order to avoid ignoring more general properties of the entire bag, IS approaches only take into account the attributes of specific instances.

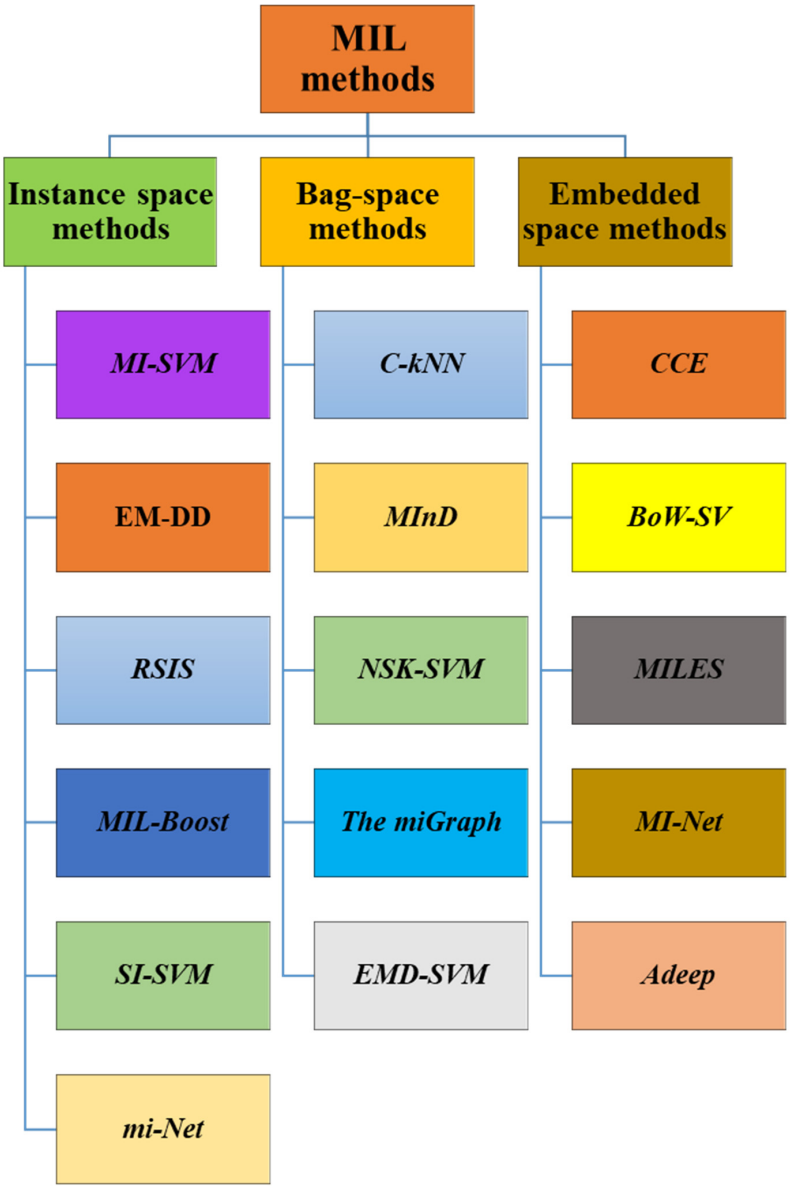

Figure 11. Some of the popular MIL methods.

The BS and ES techniques, in contrast, regard every other bag as a complete unit and train F(X) using the global bag-level data. While ES methods use a mapping function to embed multiple instances of a bag into a single "meta" instance defined on a new feature space, With the help of distance-based classifiers like k-Nearest Neighbors (kNN) and Support Vector Machine (SVM), BS techniques try to determine how similar or far apart each pair of bags are from one another and predict the labels of the bags directly [29].

These are not MIL techniques in and of themselves, but this kind of approach has been utilized as a reference point in several works [30–32] to give an idea of the relevance of employing MIL methods instead of typical supervised algorithms. In these techniques, the bag label is allocated to each instance, and bag information is ignored. Each case receives a label from the classifier during the test, and a bag is considered positive if it has no less than one positive instance. In the case of SI-SVM-TH (Single Instance Support Vector Machine with Threshold), the overall positive instances found are compared to an optimized threshold using the training data.

3.1. Instance Space Methods

IS methods disregard bag architecture and create classifiers at the level of instances after propagating bag labels to the associated instances. Then, in order to create bag labels, instance predictions are aggregated based on an appropriate MI assumption, such as the standard assumption, the collective assumption (e.g., the sum or average of individual instance predictions in a bag), and the maximum or the minimum of the instance prediction [9]. Some IS methods are mentioned below.

3.1.1. MI-SVM (Multiple Instance Support Vector Machine) and mi-SVM (Mixture of Multiple Instance Support Vector Machines)

To work in the MI setting, both MI-SVM (Multiple Instance Support Vector Machine) and mi-SVM (mixture of Multiple Instance Support Vector Machines) [33] techniques are extensions of SVM, sometimes known as a maximum-margin classifier. SVM identifies a hyperplane for binary classification that produces the greatest margin (or separation) between the two classes. All instances in negative bags have negative labels using mi-SVM, but instances in positive bags have unknown labels. A soft-margin criterion defined at the instance level is then maximized collectively over the hyperplanes and unobserved instance labels in positive bags, resulting in all instances in each negative bag being located on one side of the hyperplane and a minimum of one instance in each positive bag being positioned on the other. An SVM classifier is created with each iteration, and instance labels are changed. Once the imputed labels have stopped changing, the SVM is retrained to further refine the decision boundary using the freshly assigned labels. The margin of a positive bag is defined by the margin of the "most positive" instance, whereas the margin of a negative bag is defined by the "least negative" instance. Instead of maximizing the instance-level margin, MI-SVM represents each bag by one representative instance of the bag and maximizes the bag-level margin. When the representative instance does not vary in each bag, an SVM classifier is generated. The authors argued that mi-SVM is superior if one wants to perform an accurate instance classification; otherwise, MI-SVM is more suitable.

3.1.2. EM-DD (Expectation–Maximization Diverse Density)

Expectation–Maximization Diverse Density (DD) algorithm [34] is an extension of the Diverse Density (DD) [14] algorithm that looks for a point in the feature space with the highest DD that is as close to a number of diverse positive bags as is feasible while being as far away from the negative bags as is feasible given the neighborhood's proportion of instances of the bag. The maximum of the DD function is found using the Expectation-Maximization approach by EM DD. The classification is dependent on how far away this maximum point is.

3.1.3. RSIS (Random Subspace Instance Selection)

This method detects the witnesses in positive bags statistically by employing a technique based on random sub-spacing and clustering introduced in [35–37]. Training subgroups are sampled by applying the instances' probabilistic labels to train a set of SVMs.

3.1.4. MIL-Boost

The technique provided in [37] was generalized to create the MIL-Boost algorithm [8]. With the exception of the loss function, which is based on bag classification error, the technique is substantially the same as gradient boosting [38]. The occurrences are categorized separately, and bag labels are created by combining their labels.

3.1.5. SI-SVM (Single Instance Support Vector Machine) and SI-kNN (Single Instance k-Nearest Neighbors)

When regular (single-instance) supervised classifiers are trained on MI data using SI-SVM [30] and SI-kNN [3], the bag-membership knowledge about instances is completely

disregarded. The bag label is inherited by every instance in their implementation, and the SVM and kNN classifiers are tailored for the streamlined (single instance) problem.

3.1.6. mi-Net (Multiple Instance Neural Networks)

Wang et al. [39] coined the name "mi-Net" to refer to multiple instance neural networks (MINNs), which forecast the likelihood that a specific instance will be positive before combining instance-level probabilities to produce bag-level probabilities using a MIL pooling layer. Let us assume that MINN is composed of L layers. Each instance is first directed toward one of the numerous FC levels that serve as activation levels. After instance-level probabilities are predicted from the last FC layer or the (L-1)th layer of the MINN, the bag-level probability is collected from the last layer for each bag using a MIL pooling function (such as maximum pooling, mean pooling, and log-sum-exp pooling).

3.2. Bag-Space Methods

Compared to IS methods, which ignore the bag architecture while learning, BS methods learn the distance or similarity among each set of bags. To put it simply, BS techniques use a traditional supervised learning technique, like kNN and SVM, to learn the bag-to-bag connection before employing a suitable distance or kernel function for integrating the bags using their own member instances [15]. Some common bag-space methods are mentioned below.

3.2.1. C-kNN (Citation-k-Nearest Neighbors)

CkNN (Citation-kNN) [14] is a variation of SI-kNN (Single Instance k-Nearest Neighbors) tailored to MI data that determines the distance between two bags using the smallest Hausdorff distance in order to make sure that the estimated distance is resilient to high instance values. C-kNN is based on a two-level voting system that was motivated by the idea of references and citations in research publications. The authors proposed the terms "reference" and "citer," where references are a given bag's closest neighbors and citers are bags that view the given bag as their closest neighbor. A bag is classified as positive by employing references and citers collectively if the ratio of positive bags is higher than that of negative bags between its citers and references. Consider a bag that contains C = C+ + C- citers and R = R+ + R- references, where a subscript denotes the bag label. R+ + C+ > R- + C- identifies the target bag as positive in this case. To lessen the likelihood of producing false positives, which occur far more frequently in applications of machine learning than false negatives, the bag is put in the negative class if there is a tie. This algorithm can be modified to carry out instance classification [40].

3.2.2. MInD (Multiple Instance Learning with Bag Dissimilarities)

According to MInD [41], a vector with fields distinct from those of other bags is used to represent each bag in the training data set. These feature vectors are categorized in accordance with a standard supervised classifier, an SVM, in this instance. The publication suggests a number of dissimilarity metrics; however, the mean min provided the best overall performance.

3.2.3. NSK-SVM (Normalized Set Kernel-SVM)

An expanded version of kernel methods called NSK-SVM [42] proposes a normalized set kernel (NSK), which is used for machine learning data. The selected instance-level kernel serves as the source for the set kernel, which is particularly defined on bags. Common options include matching kernel, polynomial kernel, and radial basis function kernel. In order to lessen the influence of differing bag sizes, normalization, which is accomplished by the averaged pairwise distances amongst every instance contained in two bags, is essential. The NSK is then used to construct an SVM that can predict bag labels.

3.2.4. The miGraph

MiGraph [43] is a proposed method for bag classification by the authors that can take advantage of the relationships between instances by considering them as components of the bag that are interconnected. The observation that was made by Zhou et al. [43] is what inspired this methodology instances are hardly ever distributed (i.i.d.) independently and identically in a bag. Each bag is represented by a graph in the miGraph method, whose nodes are the instances. If the Gaussian distance across two instances is less than a predetermined threshold (such as the average distance in the bag), then there is an edge between the instances. Because instances may be reliant on one another, the weights they contribute to the bag classification are altered by the cliques visible in the graph. An SVM, along with a graph kernel (built with instance weights), classifies on the basis of between-bag similarity after all bags have been represented by their respective graphs. Utilizing an identity edge matrix (i.e., between any two instances there is no edge) can be useful in handling independent and identical instances.

3.2.5. EMD-SVM (Earth Mover's Distance-SVM)

To determine how similar any two bags are (let us say i and i'), the suggested method uses Earth Mover's Distance (EMD) [44,45]. EMD is a weighted average of the ground distances between all pairs of instances (j, j'), where instance j (j') is from bag i (i'), and vice versa. In Zhang et al. [44], the Euclidean distance is used as the ground distance measure, and the weights are obtained by resolving a linear programming issue. The obtained distances are converted to a Gaussian kernel function and then employed in an SVM for bag classification.

3.3. Embedded Space Methods

Similar to BS approaches, ES methods summarize a bag that only uses a single feature vector to extract information at the level of the bag from machine learning data and then convert a machine learning problem to a standard supervised learning problem. ES techniques, however, emphasize instance embedding [15]. Some of the embedded space methods are given below.

3.3.1. CCE (Constructive-Clustering-Based Ensemble)

In order to represent each bag, a Constructive-Clustering-based Ensemble (CCE) [28] first divides the training sets instances into C clusters using the k-means clustering algorithm. If a bag contains no less than one instance from a cluster of instances named c, the value for the associated cth feature component would be 1; if not, it is 0. An SVM can be designed to classify bags using new bag-level features. It is suggested to train several classifiers on the basis of various clustering findings and assumptions and then aggregate their predictions by a vote of the majority because there are no limits on the choice of C. In this way, CCE makes use of ensemble learning as well. Whenever there is a new bag that is presented for classification, this CCE methodology re-represents it by looking up the clustering results and then supplies the ensemble classifier with the produced feature vectors to predict the label of the bag. Be aware that any other clustering, classification, and ensemble methods in CCE may be used in place of k-means, SVM, and majority voting, respectively.

3.3.2. BoW-SVM (Bag-of-Words-SVM)

The initial stage in applying a BoW approach is compiling a sample term dictionary. By applying k-means clustering to all of the training cases, this is accomplished using BoW-SVM [29]. The most similar term found in the dictionary is then used to represent instances. The words' frequency histograms serve as a representation of bags. An SVM classifies histograms using a kernel designed for histogram comparison.

3.3.3. MILES (Multiple-Instance Learning via Embedded Instance Selection)

MILES [46], which stands for Multiple-Instance Learning by Embedded instance Selection, implies that only a portion of instances are in charge of the bag labels. Each bag is mapped into a new feature space during the embedding step using a vector representing the score of similarities among the bags being used at the time and the collection of examples from all the bags. This results in highly dimensional features, even those that are repetitive or ineffective, with the resultant feature space's dimensionality being equivalent to the overall number of instances, which may be huge. Both choosing significant features and building classifiers can be done simultaneously using SVM with the LASSO penalty [47]. Additionally, by figuring out how much each instance contributes to the classification of a bag depending on a predetermined threshold, MILES may be used for instance classification.

3.3.4. MI-Net (Multiple Instance Neural Network)

It is the first MINN (Multiple Instance Neural Network) approach in the ES techniques category. It learns how to represent bags from the features of the instances and then accordingly classifies the bags. In contrast to mi-Net, which concentrates on computing instance-level probabilities. Consider a MINN with L layers; MI-NET's pooling process, which is based on MIL, compiles all the instances into a single bag and represents it as a single feature vector, which happens in the (L-1)th layer. With a sigmoid activation function, the FC layer (also known as the Lth or the last layer) outputs bag-level probabilities from the input bag representation. In addition to the basic version mentioned above, two MI-Net variations have been proposed [39], one of which includes deep supervision [48] and the other of which takes residual connections [49] into account. Both of these can occasionally increase performance.

3.3.5. Adeep (Attention-based Deep)

Attention-based Deep MIL (ADeep) [50] is a MINN approach in addition to mi-Net and MI-Net. It alters the ES technique to improve the understanding by utilizing a cutting-edge multiple-instance learning-based pooling technique that depends on a unique attention mechanism [51], where each instance is taken as an independent unit. A weighted average of all the instances is calculated and is offered as an alternative to conventional pooling operators like max and mean, which are already specified and untrainable. Instead, a neural network consisting of two layers generates the weights and sums to 1, making them unaffected by how big or small the bag is. Naturally, instances that are more probable to be positive weigh more in the bag than the others, producing outcomes that are easier to interpret. By offering instance weights as a substitute for instance probabilities, ADeep, in this sense, connects the ES technique to the IS technique.

4. Applications of MIL

The application of a multi-instance framework in some real-life situations becomes crucial because, in such cases, the objects are characterized as bags, and each and every bag contains several feature vectors. There is no way to resolve this kind of issue by applying standard supervised learning, so in these real-world settings, the application of a multi-instance architecture becomes essential. Since it was first developed, multiple instances of learning have continually been used to solve a variety of real-world issues, including predicting drug activity, image retrieval, document classification, and document classification; researchers have proposed new application scenarios in a variety of fields, using multi-instance framework. In Figure 12. the application areas of multiple instance learning have been presented.

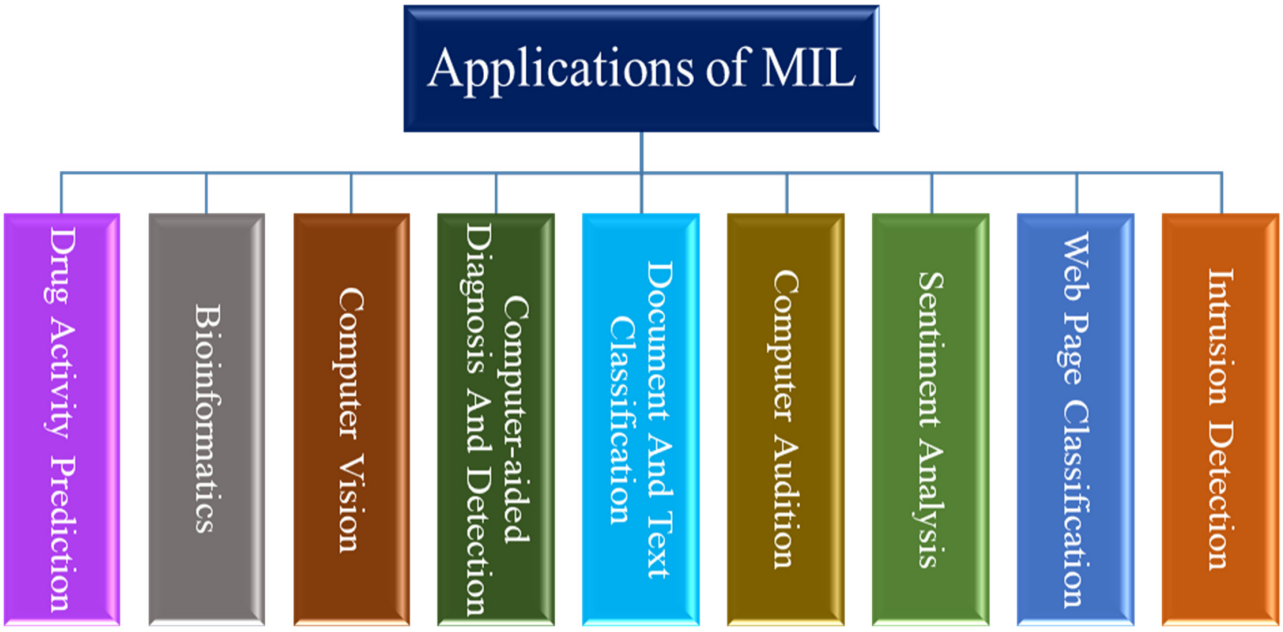

Figure 12. Some of the application areas of MIL.

4.1. Drug Activity Prediction

MIL has been used to predict the biological activity of molecules or compounds in drug discovery. Instances are molecular fragments or substructures, whereas bags are molecules. MIL can handle circumstances where the activity of individual fragments is uncertain or noisy by considering the activity of the bag (molecule) as a whole. Many different types of challenges can naturally be formulated as MIL issues. As mentioned above, the goal of the drug activity prediction problem is to determine whether a molecule will cause a specific effect or to predict the biological activity of molecules or compounds. A molecule can adopt a variety of forms that either result in the intended effect or not. It is difficult to observe the effect of every single conformation. Consequently, molecules must be viewed as a collection of conformations, which is why the MIL formulation is used. Bags represent molecules, while instances represent molecular fragments or substructures. By considering the activity of the bag (molecule) as a whole, MIL can handle situations where the activity of individual fragments is unknown or noisy. Since then, MIL has been used in numerous biological and pharmacological design applications [10].

4.2. Bioinformatics

MIL is beneficial in bioinformatics [52], where it is necessary to classify biological sequences like amino acids or DNA pairs. The objective is to identify if a protein performs a particular function, and protein sequences are considered collections of subsequences or motifs. To build a bag of overlapping sub-sequences, one common technique is to slide a window across the sequence. Only one of the subsequences is believed to be responsible for the behavior of the entire DNA or protein sequence [30].

4.3. Computer Vision

MIL is mostly applied in computer vision tasks for two major purposes: learning from data with weak annotations and characterizing complex visual concepts using sets of various subconcepts. MIL is becoming more popular in the field of medical imaging. The most widespread MIL application is undoubtedly CBIR. CBIR aims to classify images according to the objects they contain. It does not matter where an object is exactly located. Images are typically divided into smaller segments and then defined using feature vectors. The entire image represents a bag, whereas each part represents a single instance. There are

numerous approaches to split images. The image, for example, can be divided up using semantic regions, key points [53–55], or a regular grid [14]. In the latter scenario, advanced segmentation techniques are used to separate the images.

4.3.1. Object Tracking

MIL can be used in object tracking applications, where bags represent video frames or image sequences, and instances represent object proposals or bounding boxes. By considering the collective evidence from multiple instances, MIL enables robust and accurate tracking even in occlusions or partial object visibility. The localization of objects in images (or movies) requires MIL to identify instances by learning from bags. MIL is typically utilized to train visual object identification image data sets with weak labels. Or, to put it another way, labels are applied to whole images according to the items they contain. These objects can be seen in numerous places in an image; they do not all have to be in the foreground. MIL has been employed by the computer vision community to create object detectors using the massive amount of poorly labeled images that are available online. Description sentences [56–58], search engine results [59], tags connected with comparable images, and terms discovered on web pages associated with the images can all point to weak supervision of data [60].

4.3.2. Video Classification

MIL can be utilized for tasks like activity recognition or identifying anomalies in video surveillance [61–63]. Bags represent video clips or sequences, while instances are specific time frames inside such videos. MIL enables the recognition of complicated actions or the detection of unusual occurrences by taking into account all of the activity in a video clip. Sequences of the video are broken down into smaller sequences (instances), which are then categorized separately to observe the overall quantity of information displayed in the video. To identify scenes that are unsuitable for children, [64] also employs this problem formulation. MIL techniques for object tracking in films were also suggested [65–67]. For instance, in [65], a classifier is trained online to spot and follow an object of interest in a series of frames. In order to train the MIL classifier, the tracker suggests candidate windows that make up a bag. Similar to multiple instance clustering approaches [68,69], the technique creates bags using a saliency detector, which eliminates background items from positive bags to increase cluster purity. In order to build a mid-level representation of actions from a group of actions (sub-actions) found in movies, a strategy based on multiple-instance clustering is used [70].

4.4. Computer-Aided Diagnosis and Detection

MIL is increasingly being used for medical purposes. Strong labels, such as the location of anomalies in a medical scan, are typically more challenging than weak labels, such as a broad diagnosis of the person's cancer detection in WSI images. The MIL structure is helpful in this situation since patients' medical scans show both problematic and healthy parts, whereas healthy people only have healthy regions. Based on a number of diagnostic tests or examinations, MIL can be used to determine whether a disease is present or not. Patients are represented as bags, while specific test results or attributes are represented by instances. MIL can manage situations where the disease may manifest differently in various people by considering the entire test results for each patient. Applications include identifying cancer in histopathology images [16], diabetes in retinal images [71], dementia in brain MRI [72], tuberculosis in X-ray images [73], and the categorization of a chronic lung illness in CT [74]. These applications have two primary objectives, similar to other general computer vision tasks: diagnosis (i.e., predicting labels for subjects) and detection or segmentation (i.e., predicting labels for a portion of a scan). These components could be image patches, regions of interest, pixels, or voxels (3D pixels).

4.5. Document and Text Classification

One of MIL's first (1954) uses was document classification [75]. The BoW (Bag-of-Words) model shows texts as frequency histograms that quantify each word's frequency within the text. Texts and web pages are multi-part entities in this scenario, demanding the MIL classification framework. Texts can be modeled as bags since they frequently include various topics. MIL can be used to formulate text classification issues at many levels. Instances are words like in the BoW model at the lowest level. Alternatively, instances can be phrases [30,76], paragraphs [77], or passages [33,78]. MIL enables the classification of documents based on the presence or absence of relevant information without requiring explicit labeling of each segment within the document.

4.6. Computer Audition

MIL applies to computer auditions as well. A bag is created by breaking an audio file or audio clip into multiple instances, each of which possesses properties peculiar to a particular frequency range [8,79]. Some sound classification tasks can be modeled as MIL. The goal is to identify the genre of musical samples automatically. Labels are given in training for complete albums or artists but not for each snippet. The bags are collections of single-artist or album snippets. Different musical genres might be included on the same album or by the same performer; thus, the bags may have both good and bad examples. In [80], MIL is used to recognize bird songs in recordings made using single outdoor microphones. Various bird species and other noises can be heard in sound sequences. The goal is to recognize each bird's Song independently while practicing exclusively on sound files with weak labeling.

4.7. Sentiment Analysis

When performing sentiment analysis or opinion mining tasks, MIL can be used. In these tasks, instances stand in for words or text segments and bags for documents or text segments. MIL provides sentiment categorization or polarity prediction by taking into account the overall tone of a document or segment, which eliminates the need for explicit sentiment annotations at the word level. Multi-instance learning is an effective way to solve text mining challenges by converting the sentences as instances and by the use of statements that are grammatically sound and have standalone representation to address the issue of text's lacking semantic richness. In the past several years, experts have proposed employing multi-instance learning for the representation of different text-related tasks. The text is divided into different sentence units; it can then be viewed as a bag of sentences.

4.8. Web Page Classification

The MIL framework can also be used to naturally simulate web pages [81]. Like texts, websites frequently cover a wide range of subjects. For instance, a news channel's website has a number of articles on various topics. MIL, depending on a user's browsing history, has been utilized for web index-page recommendations [81,82]. Links, page names, and occasionally short descriptions of web pages are all found on an index page. A web index page is a bag in this instance, and the connected web pages are the instances. MIL can also be applied to web page classification tasks, where bags represent web pages, and instances represent snippets or regions within those pages. It allows the classification of web pages based on their overall content without the need for precise labeling of each region within the page [83].

A user is thought to be interested in at least one of the pages connected to a web index page if they have marked it as a favorite. The list of most common terms found on a web page serves as its representation. Advertisers tend to stay away from certain pages with sensitive information like adultery or war in virtual online advertisements. In [77], a MIL classifier evaluates web page content to determine which pages are appropriate for advertisements.

4.9. Intrusion Detection

MIL can be applied to detect network intrusions by considering network traffic data as bags. The aim is to classify a bag as malicious if it contains instances representing suspicious network activities. It is useful to handle packets as continuous streams in order to identify different sorts of attacks and decrease repeated alarms on the same attack. In [84], a new method is put forward for handling groups of related packets, whereas existing anomaly detection based on machine learning treats a packet as a fundamental unit. The proposed technique is in accordance with a BAG unit and a BAG creation algorithm that organizes packets. Some other common applications of MIL are mentioned in Table 1.

Table 1. Some common applications of MIL.

Application of MIL	Author Name	Description
Sound Classification	M. I. Mandel, D. P. W. Ellis [85].	It is possible to cast some sound classification tasks as MIL. The goal of [85] is to automatically identify the genre of musical snippets. Labels are given in training for complete albums or artists but not for each snippet. The bags are collections of single-artist or album snippets. Because numerous musical genres might be included on the same album or by the same singer, the bags might include both positive and negative instances.
Recognition of different bird songs	F. Briggs, X. Z. Fern, R [80].	MIL is used in [80] to detect bird songs in recordings made using microphones. Various bird species and other noises can be heard in sound sequences. The goal is to recognize each bird song independently while restricting training to sound files with inadequate labeling.
Determining Personality traits using audio signals	M.-A. Carbonneau, E. Granger, Y. Attabi, G. Gagnon [86].	According to [86], a BoW (Bag-of-words) framework is used for determining personality qualities from audio signals represented as spectrograms. In that situation, the spectrogram's discrete regions are instances, and the complete voice signals are bags.
Human activity sensors	X. Guan, R. Raich, W.-K. Wong [87] and M. Stikic, D. Larlus, S. Ebert, B. Schiele [88].	In [87,88], Wearable body sensors are used with MIL to identify human activity. The users' declaration of the actions that were carried out during a specific time period results in weak supervision. Activities typically do not last the entire period, and each period could have a distinct set of activities. In this configuration, instances are sub-periods, while full periods are bags.
Prediction of hard drive failure	J. F. Murray, G. F. Hughes, K. Kreutz-Delgado [89].	Time series are a collection of measurements on hard drives taken at regular intervals, and they are used alongside MIL to predict hard drive failure [89]. The objective is to predict when a product will fail. Time series suggest the underlying structure of bags that should not be disregarded.
Detection of buried landmines	Manandhar, K. D. Morton, L. M. Collins, P. A. Torrione [90]. A. Karem, H. Frigui [91].	MIL classifiers in [90,91] use ground-penetrating radar signals to find buried landmines. At different depths in the soil, measurements are taken when a detection takes place at a specific GPS position. The feature vectors for various depths are contained in a bag at each detection location.
Predicting the performance of stocks	O. Maron, T. Lozano-P'erez [14].	MIL is employed to choose stocks in [14]. The 100 best-performing stocks are gathered into positive bags each month, while the five worst-performing stocks are placed in negative bags. Based on these bags, an instance classifier chooses the best stocks.
Prediction for film nominations	A. Mcgovern, D. Jensen [92]	A strategy for predicting which films will be nominated for an award is described in [92]. A graph is created that represents a movie's relationships to stars, studios, genre, release date, etc. In order to assess whether test cases were successful, the MIL algorithm determines which sub-graph explains the nomination.

5. Challenges while Deploying MIL

For multiple instances of learning deployment to be successful, there are a number of issues that must be resolved. Here are a few of the main difficulties:

- Ambiguity in Instance Labels: One of the main problems with MIL is the difficulty in deciding which labels to provide for the different instances that make up a bag. The precise labeling of individual instances is still unknown because the bag is labeled according to whether or not there are positive instances. The learning task may become more difficult as a result of this ambiguity, and robust algorithms are needed to handle it correctly [93].
- Bag-level labeling: MIL requires bag-level labels, whereas conventional machine learning techniques work at the instance level. Due to this distinction, specific algorithms that can use bag-level data to anticipate the future must be created. For accurate multiple-instance learning deployment, it is essential to design efficient bag-level labeling mechanisms [82].
- Feature representation: The selection and depiction of features from the bags of instances presents another difficulty. MIL algorithms typically work with the instances in each bag's aggregated features. Achieving good performance depends on selecting the right features that capture the bag-level information while maintaining relevant instance-level properties [33].
- Complexity of Computation: The necessity to evaluate bags with several instances makes MIL methods computationally demanding. As the quantity of instances per bag rises, complexity also rises. It is difficult to deploy large-scale MIL issues without first developing effective algorithms and optimization strategies [29].
- Lack of Labeled Bags: In many practical applications, getting labeled bags might be expensive or difficult. The lack of readily available labeled bags makes it difficult to train and test MIL algorithms. The performance of MIL models must be enhanced using strategies like active learning, semi-supervised learning, or utilizing additional data to address a lack of labeled bags [3].

Challenges within the Realm of Machine Learning, Large Language Models, and XAI

Multiple instance learning faces various challenges in the realm of machine learning, large language models, and explainable artificial intelligence. Interpretability and explainability are the main focus of XAI and large language models. However, due to the inherent ambiguity in labeling and reliance on bag-level representation, sometimes MIL may not provide good interpretability results. Scalability and efficiency, data bias, and fairness of the model are also some of the key challenges. Likewise, integrating active learning and weak supervision strategies with XAI and large language models is also one of the research challenges.

6. Conclusions

MIL is a key area for machine learning research. This research work reviews the potential of MIL and outlines the key concepts of a few MIL-based methodologies. There are multiple areas where MIL could be used with promising performance, as demonstrated in this study. Multiple instance learning is a weakly supervised learning where the learning dataset contains bags of instances instead of a single feature vector; bags could be either positive or negative. MIL can be implemented using different methods. Each MIL method can be divided into three categories: instance-space (IS), bag-space (BS), and embedded-space (EB) methods. MIL's applications in various fields, including medical imaging, computer vision, image segmentation, computer audition, bioinformatics, and text categorization, are also explained. Finally, a few of the challenges encountered during the deployment of MIL are mentioned. MIL is a more adaptable form of weakly supervised learning. Furthermore, while multiple instances describe the examples better than a single instance in some scenarios, it also increases the size of the data set, so more attention must be paid to multi-instance learning optimization in order to apply the multi-instance framework to

large-scale data sets. The aim of this study is to highlight the potential areas where MIL could be helpful in addressing the challenges in respective domains, for example, cancer diagnosis in whole slide images, drug activity prediction, intrusion detection, and others.

Author Contributions: S.F. and S.A.: writing—original draft, analysis, and editing. S.F. and S.A.: conceptualization, methodology, writing—the original draft, review, and editing. H.-C.K.: formal analysis. S.A. and S.F.: validation. H.-C.K.: formal analysis, supervision. H.-C.K.: writing—review. H.-C.K.: funding acquisition, project administration, supervision. H.-C.K.: resources, writing—review. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute validation (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (Grant No: HI21C0977).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Abbreviations	**Description**
MIL	Multiple Instance Learning
RL	Reinforcement Learning
AI	Artificial Intelligence
WSI	Whole Slide Images
KNN	K Nearest Neighbors
APR	Axis-Parallel Hyper-Rectangle
TCR-Sequence	T-cell Receptor-Sequence
EDUs	Elemental Discourse Units
EB	Embedded-Space
IS	Instance-Space
BS	Bag-Space
MI-SVM	Multiple Instance Support Vector Machine
EM-DD	Expectation–Maximization Diverse Density
RSIS	Random Subspace Instance Selection
BoW	Bag of words
mi-SVM	Mixture of Multiple Instance Support Vector Machines
SI-SVM	Single Instance Support Vector Machine
SI-kNN	Single instance k-Nearest 353 Neighbors
mi-Net	Multiple instance Neural Networks
FC layer	Fully Connected Layer
CKNN	Citation-kNN
MInD	Multiple Instance Learning with Bag Dissimilarities
NSK-SVM	Normalized Set Kernel-SVM
EMD-SVM	Earth Mover's Distance-SVM
CCE	Constructive clustering based Ensemble
MILES	Multiple-Instance Learning via Embedded Instance Selection
DNA	Deoxyribonucleic Acid
CT	Computerized Tomography
GPS	Global Positioning System
MRI	Magnetic Resonance Imaging
XAI	Explainable Artificial Intelligence
Mathematical symbols	
Symbols	**Description**
C	Citer
R	Reference
L layers	Number of layers
(j, j′)	Pair of Instances

References

1. Li, Y. Deep reinforcement learning: An overview. *arXiv* **2017**, arXiv:1701.07274.
2. Ray, S. A quick review of machine learning algorithms. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 35–39.
3. Carbonneau, M.-A.; Cheplygina, V.; Granger, E.; Gagnon, G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognit.* **2018**, *77*, 329–353. [CrossRef]
4. Brand, L.; Seo, H.; Baker, L.Z.; Ellefsen, C.; Sargent, J.; Wang, H. A linear primal–dual multi-instance SVM for big data classifications. *Knowl. Inf. Syst.* **2023**, *34*, 1–32. [CrossRef]
5. Afsar Minhas, F.u.A.; Ross, E.D.; Ben-Hur, A. Amino acid composition predicts prion activity. *PLoS Comput. Biol.* **2017**, *13*, e1005465. [CrossRef]
6. Yang, J. *Review of Multi-Instance Learning and Its Applications*; Technical Report; School of Computer Science, Carnegie Mellon University: Pittsburgh, PA, USA, 2005.
7. Zhou, Z.-H. *Multi-Instance Learning: A Survey*; Technical Report; Department of Computer Science & Technology, Nanjing University: Nanjing, China, 2004; p. 1.
8. Babenko, B. *Multiple Instance Learning: Algorithms and Applications*; University of California: San Diego, CA, USA, 2008; Volume 19.
9. Maia, P. *An Introduction to Multiple Instance Learning*; NILG.AI: Porto, Portugal, 2021.
10. Dietterich, T.G.; Lathrop, R.H.; Lozano-Pérez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.* **1997**, *89*, 31–71. [CrossRef]
11. Wang, J.; Zucker, J.-D. *Solving Multiple-Instance Problem: A Lazy Learning Approach*; University of Southampton: Southampton, UK, 2000.
12. Wang, Q.; Yuan, Y.; Yan, P.; Li, X. Saliency detection by multiple-instance learning. *IEEE Trans. Cybern.* **2013**, *43*, 660–672. [CrossRef]
13. Sudharshan, P.; Petitjean, C.; Spanhol, F.; Oliveira, L.E.; Heutte, L.; Honeine, P. Multiple instance learning for histopathological breast cancer image classification. *Expert Syst. Appl.* **2019**, *117*, 103–111. [CrossRef]
14. Maron, O.; Lozano-Pérez, T. A framework for multiple-instance learning. *Adv. Neural Inf. Process. Syst.* **1997**, *10*. Available online: https://proceedings.neurips.cc/paper_files/paper/1997/file/82965d4ed8150294d4330ace00821d77-Paper.pdf (accessed on 1 October 2023).
15. Xiong, D.; Zhang, Z.; Wang, T.; Wang, X. A comparative study of multiple instance learning methods for cancer detection using T-cell receptor sequences. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3255–3268. [CrossRef]
16. Xu, Y.; Zhu, J.-Y.; Eric, I.; Chang, C.; Lai, M.; Tu, Z. Weakly supervised histopathology cancer image segmentation and classification. *Med. Image Anal.* **2014**, *18*, 591–604. [CrossRef] [PubMed]
17. Fraz, M.M.; Barman, S.A. Computer vision algorithms applied to retinal vessel segmentation and quantification of vessel caliber. *Image Anal. Model. Ophthalmol.* **2014**, *49*, 49–84.
18. Combalia, M.; Vilaplana, V. Monte-Carlo sampling applied to multiple instance learning for histological image classification. In Proceedings of the International Workshop on Deep Learning in Medical Image Analysis, Granada, Spain, 20 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 274–281.
19. Qiao, M.; Liu, L.; Yu, J.; Xu, C.; Tao, D. Diversified dictionaries for multi-instance learning. *Pattern Recognit.* **2017**, *64*, 407–416. [CrossRef]
20. Angelidis, S.; Lapata, M. Multiple instance learning networks for fine-grained sentiment analysis. *Trans. Assoc. Comput. Linguist.* **2018**, *6*, 17–31. [CrossRef]
21. Gadermayr, M.; Tschuchnig, M. Multiple instance learning for digital pathology: A review on the state-of-the-art, limitations & future potential. *arXiv* **2022**, arXiv:2206.04425.
22. Boschman, J. *Multiple-Instance Learning—One Minute Introduction*; Medium: San Francisco, CA, USA, 2021.
23. Song, X.; Jiao, L.; Yang, S.; Zhang, X.; Shang, F. Sparse coding and classifier ensemble based multi-instance learning for image categorization. *Signal Process.* **2013**, *93*, 1–11. [CrossRef]
24. Astorino, A.; Fuduli, A.; Gaudioso, M.; Vocaturo, E. A multiple instance learning algorithm for color images classification. In Proceedings of the 22nd International Database Engineering & Applications Symposium, Villa San Giovanni, Italy, 18–20 June 2018; pp. 262–266.
25. Weiss, G.M.; Hirsh, H. Learning to predict rare events in event sequences. In Proceedings of the KDD—4th International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31 August 1998; pp. 359–363.
26. Moniz, N.; Branco, P.; Torgo, L. Resampling strategies for imbalanced time series. In Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, Canada, 17–19 October 2016; pp. 282–291.
27. Gunetti, D.; Ruffo, G. Intrusion detection through behavioral data. In Proceedings of the International Symposium on Intelligent Data Analysis, Amsterdam, The Netherlands, 9–11 August 1999.
28. Zhou, Z.-H.; Zhang, M.-L. Solving multi-instance problems with classifier ensemble based on constructive clustering. *Knowl. Inf. Syst.* **2007**, *11*, 155–170. [CrossRef]
29. Amores, J. Multiple instance classification: Review, taxonomy and comparative study. *Artif. Intell.* **2013**, *201*, 81–105. [CrossRef]
30. Ray, S.; Craven, M. Supervised versus multiple instance learning: An empirical comparison. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2022; pp. 697–704.

31. Alpaydın, E.; Cheplygina, V.; Loog, M.; Tax, D.M. Single vs. multiple-instance classification. *Pattern Recognit.* **2015**, *48*, 2831–2838. [CrossRef]
32. Bunescu, R.C.; Mooney, R.J. Multiple instance learning for sparse positive bags. In Proceedings of the 24th International Conference on Machine Learning 2007, Corvalis, OR, USA, 20–24 June 2007; pp. 105–112.
33. Andrews, S.; Tsochantaridis, I.; Hofmann, T. Support vector machines for multiple-instance learning. *Adv. Neural Inf. Process. Syst.* **2002**, *15*.
34. Zhang, Q.; Goldman, S. EM-DD: An improved multiple-instance learning technique. *Adv. Neural Inf. Process. Syst.* **2001**, *14*. Available online: https://proceedings.neurips.cc/paper_files/paper/2001/file/e4dd5528f7596dcdf871aa55cfccc53c-Paper.pdf (accessed on 1 October 2023).
35. Carbonneau, M.-A.; Granger, E.; Raymond, A.J.; Gagnon, G. Robust multiple-instance learning ensembles using random subspace instance selection. *Pattern Recognit.* **2016**, *58*, 83–99. [CrossRef]
36. Carbonneau, M.-A.; Granger, E.; Gagnon, G. Witness identification in multiple instance learning using random subspaces. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 3639–3644.
37. Viola, P.; Platt, J.C.; Zhang, C. Multiple instance boosting for object recognition. In Proceedings of the Neural Information Processing Systems, Vancouver, BC, Canada, 4 December 2006.
38. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
39. Wang, X.; Yan, Y.; Tang, P.; Bai, X.; Liu, W. Revisiting multiple instance neural networks. *Pattern Recognit.* **2018**, *74*, 15–24. [CrossRef]
40. Zhou, Z.-H.; Xue, X.-B.; Jiang, Y. Locating regions of interest in CBIR with multi-instance learning techniques. In Proceedings of the AI 2005: Advances in Artificial Intelligence: 18th Australian Joint Conference on Artificial Intelligence, Sydney, Australia, 5–9 December 2005; Proceedings 18. Springer: Berlin/Heidelberg, Germany, 2005; pp. 92–101.
41. Cheplygina, V.; Tax, D.M.; Loog, M. Multiple instance learning with bag dissimilarities. *Pattern Recognit.* **2015**, *48*, 264–275. [CrossRef]
42. Gärtner, T.; Flach, P.A.; Kowalczyk, A.; Smola, A.J. Multi-instance kernels. *ICML* **2002**, *2*, 7.
43. Zhou, Z.-H.; Sun, Y.-Y.; Li, Y.-F. Multi-instance learning by treating instances as non-iid samples. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 1249–1256.
44. Zhang, J.; Marszałek, M.; Lazebnik, S.; Schmid, C. Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. J. Comput. Vis.* **2007**, *73*, 213–238. [CrossRef]
45. Rubner, Y.; Tomasi, C.; Guibas, L.J. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* **2000**, *40*, 99–121. [CrossRef]
46. Chen, Y.; Bi, J.; Wang, J.Z. MILES: Multiple-instance learning via embedded instance selection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1931–1947. [CrossRef]
47. Zhu, J.; Rosset, S.; Tibshirani, R.; Hastie, T. 1-norm support vector machines. *Adv. Neural Inf. Process. Syst.* **2003**, *16*. Available online: https://proceedings.neurips.cc/paper_files/paper/2003/file/49d4b2faeb4b7b9e745775793141e2b2-Paper.pdf (accessed on 1 October 2023).
48. Lee, C.-Y.; Xie, S.; Gallagher, P.; Zhang, Z.; Tu, Z. Deeply-supervised nets. In Proceedings of the Artificial Intelligence and Statistics 2015, San Diego, CA, USA, 9–12 May 2015; pp. 562–570.
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
50. Ilse, M.; Tomczak, J.; Welling, M. Attention-based deep multiple instance learning. In Proceedings of the International Conference on Machine Learning 2018, Stockholm, Sweden, 10–15 July 2018; pp. 2127–2136.
51. Raffel, C.; Ellis, D.P. Feed-forward networks with attention can solve some long-term memory problems. *arXiv* **2015**, arXiv:1512.08756.
52. Zhang, Y.; Chen, Y.; Ji, X. Motif discovery as a multiple-instance problem. In Proceedings of the 2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06), Arlington, VA, USA, 13–15 November 2006; pp. 805–809.
53. Yang, C.; Dong, M.; Hua, J. Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) IEEE, New York, NY, USA, 17–22 June 2006; pp. 2057–2063.
54. Chen, Y.; Wang, J.Z. Image categorization by learning and reasoning with regions. *J. Mach. Learn. Res.* **2004**, *5*, 913–939.
55. Csurka, G.; Dance, C.; Fan, L.; Willamowski, J.; Bray, C. Visual categorization with bags of keypoints. In Proceedings of the Workshop on Statistical Learning in Computer Vision, ECCV 2004, Prague, Czech Republic, 10–14 May 2004; pp. 1–2.
56. Xu, H.; Venugopalan, S.; Ramanishka, V.; Rohrbach, M.; Saenko, K. A multi-scale multiple instance video description network. *arXiv* **2015**, arXiv:1505.05914.
57. Karpathy, A.; Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3128–3137.
58. Fang, H.; Gupta, S.; Iandola, F.; Srivastava, R.K.; Deng, L.; Dollár, P.; Gao, J.; He, X.; Mitchell, M.; Platt, J.C. From captions to visual concepts and back. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1473–1482.

59. Zhu, J.-Y.; Wu, J.; Xu, Y.; Chang, E.; Tu, Z. Unsupervised object class discovery via saliency-guided multiple class learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 862–875. [CrossRef] [PubMed]
60. Wu, J.; Yu, Y.; Huang, C.; Yu, K. Deep multiple instance learning for image classification and auto-annotation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015, Boston, MA, USA, 7–12 June 2015; pp. 3460–3469.
61. Yang, X.; Li, C.; Zeng, Q.; Pan, X.; Yang, J.; Xu, H. Vehicle re-identification via spatio-temporal multi-instance learning. In Proceedings of the International Conference on Neural Computing for Advanced Applications, Hefei, China, 7–9 July 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 482–493.
62. Varga, D.; Szirányi, T. Person re-identification based on deep multi-instance learning. In Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 28 August–2 September 2017; pp. 1559–1563.
63. Liu, X.; Bi, S.; Ma, X.; Wang, J. Multi-Instance Convolutional Neural Network for multi-shot person re-identification. *Neurocomputing* **2019**, *337*, 303–314. [CrossRef]
64. Wang, J.; Li, B.; Hu, W.; Wu, O. Horror video scene recognition via multiple-instance learning. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 1325–1328.
65. Babenko, B.; Yang, M.-H.; Belongie, S. Robust object tracking with online multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 1619–1632. [CrossRef]
66. Zhang, K.; Song, H. Real-time visual tracking via online weighted multiple instance learning. *Pattern Recognit.* **2013**, *46*, 397–411. [CrossRef]
67. Lu, H.; Zhou, Q.; Wang, D.; Xiang, R. A co-training framework for visual tracking with multiple instance learning. In Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 21–23 March 2011; pp. 539–544.
68. Zhang, M.-L.; Zhou, Z.-H. Multi-instance clustering with applications to multi-instance prediction. *Appl. Intell.* **2009**, *31*, 47–68. [CrossRef]
69. Zhang, D.; Wang, F.; Si, L.; Li, T. Maximum margin multiple instance clustering with applications to image and text clustering. *IEEE Trans. Neural Netw.* **2011**, *22*, 739–751. [CrossRef]
70. Zhu, J.; Wang, B.; Yang, X.; Zhang, W.; Tu, Z. Action recognition with actons. In Proceedings of the IEEE International Conference on Computer Vision 2013, Sydney, Australia, 1–8 December 2013; pp. 3559–3566.
71. Quellec, G.; Lamard, M.; Abràmoff, M.D.; Decencière, E.; Lay, B.; Erginay, A.; Cochener, B.; Cazuguel, G. A multiple-instance learning framework for diabetic retinopathy screening. *Med. Image Anal.* **2012**, *16*, 1228–1240. [CrossRef]
72. Tong, T.; Wolz, R.; Gao, Q.; Guerrero, R.; Hajnal, J.V.; Rueckert, D.; Alzheimer's Disease Neuroimaging Initiative. Multiple instance learning for classification of dementia in brain MRI. *Med. Image Anal.* **2014**, *18*, 808–818. [CrossRef]
73. Melendez, J.; Van Ginneken, B.; Maduskar, P.; Philipsen, R.H.; Reither, K.; Breuninger, M.; Adetifa, I.M.; Maane, R.; Ayles, H.; Sánchez, C.I. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays. *IEEE Trans. Med. Imaging* **2014**, *34*, 179–192. [CrossRef]
74. Cheplygina, V.; Sørensen, L.; Tax, D.M.; Pedersen, J.H.; Loog, M.; De Bruijne, M. Classification of COPD with multiple instance learning. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Washington, DC, USA, 24–28 August 2014; pp. 1508–1513.
75. Harris, Z.S. Distributional Structure. *Word* **1954**, *10*, 146–162. [CrossRef]
76. Pappas, N.; Popescu-Belis, A. Explaining the stars: Weighted multiple-instance learning for aspect-based sentiment analysis. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 455–466.
77. Zhang, Y.; Surendran, A.C.; Platt, J.C.; Narasimhan, M. Learning from multi-topic web documents for contextual advertisement. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NV, USA, 24–27 August 2008; pp. 1051–1059.
78. Zhang, D.; He, J.; Lawrence, R. Mi2ls: Multi-instance learning from multiple informationsources. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 149–157.
79. Tian, Y.; Shi, J.; Li, B.; Duan, Z.; Xu, C. Audio-visual event localization in unconstrained videos. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 247–263.
80. Briggs, F.; Fern, X.Z.; Raich, R. Rank-loss support instance machines for MIML instance annotation. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2012; pp. 534–542.
81. Zafra, A.; Ventura, S.; Herrera-Viedma, E.; Romero, C. Multiple instance learning with genetic programming for web mining. In Proceedings of the International Work-Conference on Artificial Neural Networks, San Sebastián, Spain, 20–22 June 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 919–927.
82. Zhou, Z.-H.; Zhou, Z.-H. Semi-supervised learning. In *Machine Learning*; Springer: Singapore, 2021; pp. 315–341.
83. Birk, A. Robot learning and self-sufficiency: What the energy-level can tell us about a robot's performance. In Proceedings of the European Workshop on Learning Robots, Brighton, UK, 1–2 August 1997; Springer: Berlin/Heidelberg, Germany, 1997; pp. 109–125.

84. Weon, I.-Y.; Song, D.-H.; Ko, S.-B.; Lee, C.-H. A multiple instance learning problem approach model to anomaly network intrusion detection. *J. Inf. Process. Syst.* **2005**, *1*, 14–21. [CrossRef]
85. Mandel, M.I.; Ellis, D.P. Multiple-instance learning for music information retrieval. In Proceedings of the ISMIR 2008: Proceedings of the 9th International Conference of Music Information Retrieval, Philadelphia, PA, USA, 14–18 September 2008.
86. Carbonneau, M.-A.; Granger, E.; Attabi, Y.; Gagnon, G. Feature learning from spectrograms for assessment of personality traits. *IEEE Trans. Affect. Comput.* **2017**, *11*, 25–31. [CrossRef]
87. Guan, X.; Raich, R.; Wong, W.-K. Efficient multi-instance learning for activity recognition from time series data using an auto-regressive hidden markov model. In Proceedings of the International Conference on Machine Learning 2016, New York, NY, USA, 19–24 June 2016; pp. 2330–2339.
88. Stikic, M.; Larlus, D.; Ebert, S.; Schiele, B. Weakly supervised recognition of daily life activities with wearable sensors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2521–2537. [CrossRef]
89. Murray, J.F.; Hughes, G.F.; Kreutz-Delgado, K.; Schuurmans, D. Machine Learning Methods for Predicting Failures in Hard Drives: A Multiple-Instance Application. *J. Mach. Learn. Res.* **2005**, *6*, 783–816.
90. Manandhar, A.; Morton, K.D., Jr.; Collins, L.M.; Torrione, P.A. Multiple instance learning for landmine detection using ground penetrating radar. In Proceedings of the Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XVII, Baltimore, MD, USA, 23–27 April 2012; SPIE: Bellingham, WA, USA, 2012; pp. 668–678.
91. Karem, A.; Frigui, H. A multiple instance learning approach for landmine detection using ground penetrating radar. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 878–881.
92. McGovern, A.; Jensen, D. Identifying predictive structures in relational data using multiple instance learning. In Proceedings of the 20th International Conference on Machine Learning (ICML-03) 2003, Washington, DC, USA, 21–24 August 2003; pp. 528–535.
93. Maron, O.; Ratan, A.L. Multiple-instance learning for natural scene classification. In Proceedings of the ICML—International Conference on Machine Learning 1998, Madison, WI, USA, 24–27 July 1998; pp. 341–349.