

Title: Multiple instance learning

URL: [https://en.wikipedia.org/wiki/Multiple\\_instance\\_learning](https://en.wikipedia.org/wiki/Multiple_instance_learning)

PageID: 48841414

Categories: Category:Machine learning

Source: Wikipedia (CC BY-SA 4.0).

-----

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning

Decision trees

Ensembles Bagging Boosting Random forest

Bagging

Boosting

Random forest

k -NN

Linear regression

Naive Bayes

Artificial neural networks

Logistic regression

Perceptron

Relevance vector machine (RVM)

Support vector machine (SVM)

BIRCH

CURE

Hierarchical

k -means

Fuzzy

Expectation–maximization (EM)

DBSCAN

OPTICS

Mean shift

Factor analysis

CCA

ICA

LDA

NMF

PCA

PGD

t-SNE

SDL

Graphical models Bayes net Conditional random field Hidden Markov

Bayes net

Conditional random field

Hidden Markov

RANSAC

k -NN

Local outlier factor  
Isolation forest  
Autoencoder  
Deep learning  
Feedforward neural network  
Recurrent neural network LSTM GRU ESN reservoir computing  
LSTM  
GRU  
ESN  
reservoir computing  
Boltzmann machine Restricted  
Restricted  
GAN  
Diffusion model  
SOM  
Convolutional neural network U-Net LeNet AlexNet DeepDream  
U-Net  
LeNet  
AlexNet  
DeepDream  
Neural field Neural radiance field Physics-informed neural networks  
Neural radiance field  
Physics-informed neural networks  
Transformer Vision  
Vision  
Mamba  
Spiking neural network  
Memtransistor  
Electrochemical RAM (ECRAM)  
Q-learning  
Policy gradient  
SARSA  
Temporal difference (TD)  
Multi-agent Self-play  
Self-play  
Active learning  
Crowdsourcing  
Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

In machine learning , multiple-instance learning (MIL) is a type of supervised learning . Instead of receiving a set of instances which are individually labeled , the learner receives a set of labeled bags , each containing many instances. In the simple case of multiple-instance binary classification , a bag may be labeled negative if all the instances in it are negative. On the other hand, a bag is labeled positive if there is at least one instance in it which is positive. From a collection of labeled bags, the learner tries to either (i) induce a concept that will label individual instances correctly or (ii) learn how to label bags without inducing the concept.

Babenko (2008) [ 1 ] gives a simple example for MIL. Imagine several people, and each of them has a key chain that contains few keys. Some of these people are able to enter a certain room, and some aren't. The task is then to predict whether a certain key or a certain key chain can get you into

that room. To solve this problem we need to find the exact key that is common for all the "positive" key chains. If we can correctly identify this key, we can also correctly classify an entire key chain - positive if it contains the required key, or negative if it doesn't.

## Machine learning

Depending on the type and variation in training data, machine learning can be roughly categorized into three frameworks: supervised learning, unsupervised learning, and reinforcement learning. Multiple instance learning (MIL) falls under the supervised learning framework, where every training instance has a label, either discrete or real valued. MIL deals with problems with incomplete knowledge of labels in training sets. More precisely, in multiple-instance learning, the training set consists of labeled "bags", each of which is a collection of unlabeled instances. A bag is positively labeled if at least one instance in it is positive, and is negatively labeled if all instances in it are negative. The goal of the MIL is to predict the labels of new, unseen bags.

## History

Keeler et al., [ 2 ] in his work in the early 1990s was the first one to explore the area of MIL. The actual term multi-instance learning was introduced in the middle of the 1990s, by Dietterich et al. while they were investigating the problem of drug activity prediction. [ 3 ] They tried to create a learning system that could predict whether new molecule was qualified to make some drug, or not, through analyzing a collection of known molecules. Molecules can have many alternative low-energy states, but only one, or some of them, are qualified to make a drug. The problem arose because scientists could only determine if molecule is qualified, or not, but they couldn't say exactly which of its low-energy shapes are responsible for that.

One of the proposed ways to solve this problem was to use supervised learning, and regard all the low-energy shapes of the qualified molecule as positive training instances, while all of the low-energy shapes of unqualified molecules as negative instances. Dietterich et al. showed that such method would have a high false positive noise, from all low-energy shapes that are mislabeled as positive, and thus wasn't really useful. [ 3 ] Their approach was to regard each molecule as a labeled bag, and all the alternative low-energy shapes of that molecule as instances in the bag, without individual labels. Thus formulating multiple-instance learning.

Solution to the multiple instance learning problem that Dietterich et al. proposed is the axis-parallel rectangle (APR) algorithm. [ 3 ] It attempts to search for appropriate axis-parallel rectangles constructed by the conjunction of the features. They tested the algorithm on Musk dataset, [ 4 ] [ 5 ] [ dubious – discuss ] which is a concrete test data of drug activity prediction and the most popularly used benchmark in multiple-instance learning. APR algorithm achieved the best result, but APR was designed with Musk data in mind.

Problem of multi-instance learning is not unique to drug finding. In 1998, Maron and Ratan found another application of multiple instance learning to scene classification in machine vision, and devised Diverse Density framework. [ 6 ] Given an image, an instance is taken to be one or more fixed-size subimages, and the bag of instances is taken to be the entire image. An image is labeled positive if it contains the target scene - a waterfall, for example - and negative otherwise. Multiple instance learning can be used to learn the properties of the subimages which characterize the target scene. From there on, these frameworks have been applied to a wide spectrum of applications, ranging from image concept learning and text categorization, to stock market prediction.

## Examples

Take image classification for example Amores (2013) . Given an image, we want to know its target class based on its visual content. For instance, the target class might be "beach", where the image contains both "sand" and "water". In MIL terms, the image is described as a bag  $X = \{X_1, \dots, X_N\}$   $\{ \displaystyle X = \{X_{\{1\}}, \dots, X_{\{N\}} \}$ , where each  $X_i$   $\{ \displaystyle X_{\{i\}} \}$  is the feature vector (called instance ) extracted from the corresponding  $i$   $\{ \displaystyle i \}$ -th region in the image and  $N$   $\{ \displaystyle N \}$  is the total regions (instances) partitioning the image. The bag is labeled positive ("beach") if it contains both "sand" region instances and "water" region instances.

Examples of where MIL is applied are:

Molecule activity

Predicting binding sites of Calmodulin binding proteins [ 7 ]

Predicting function for alternatively spliced isoforms Li, Menon & et al. (2014) , Eksi et al. (2013)

Image classification Maron & Ratan (1998)

Text or document categorization Kotzias et al. (2015)

Predicting functional binding sites of MicroRNA targets Bandyopadhyay, Ghosh & et al. (2015)

Medical image classification Zhu et al. (2017) [ verification needed ] P.J.Sudharshan et al. (2019)  
harvtxt error: no target: CITEREF.P.J.Sudharshan\_et\_al.2019 ( help )

Numerous researchers have worked on adapting classical classification techniques, such as support vector machines or boosting , to work within the context of multiple-instance learning.

### Definitions

If the space of instances is  $X$  , then the set of bags is the set of functions  $N^X = \{ B : X \rightarrow N \}$  , which is isomorphic to the set of multi-subsets of  $X$  . For each bag  $B \in N^X$  and each instance  $x \in X$  ,  $B(x)$  is viewed as the number of times  $x$  occurs in  $B$  . [ 8 ] Let  $Y$  be the space of labels, then a "multiple instance concept" is a map  $c : N^X \rightarrow Y$  . The goal of MIL is to learn such a concept. The remainder of the article will focus on binary classification , where  $Y = \{ 0, 1 \}$  .

### Assumptions

Most of the work on multiple instance learning, including Dietterich et al. (1997) and Maron & Lozano-Pérez (1997) early papers, [ 3 ] [ 9 ] make the assumption regarding the relationship between the instances within a bag and the class label of the bag. Because of its importance, that assumption is often called standard MI assumption.

### Standard assumption

The standard assumption takes each instance  $x \in X$  to have an associated label  $y \in \{ 0, 1 \}$  which is hidden to the learner. The pair  $(x, y)$  is called an "instance-level concept". A bag is now viewed as a multiset of instance-level concepts, and is labeled positive if at least one of its instances has a positive label, and negative if all of its instances have negative labels. Formally, let  $B = \{ (x_1, y_1), \dots, (x_n, y_n) \}$  be a bag. The label of  $B$  is then  $c(B) = 1 - \prod_{i=1}^n (1 - y_i)$  . Standard MI assumption is asymmetric, which means that if the positive and negative labels are reversed, the assumption has a different meaning. Because of that, when we use this assumption, we need to be clear which label should be the positive one.

The standard assumption might be viewed as too strict, and therefore in the recent years, researchers tried to relax that position, which gave rise to other more loose assumptions. [ 10 ] The reason for this is the belief that standard MIL assumption is appropriate for the Musk dataset, but since MIL can be applied to numerous other problems, some different assumptions could probably be more appropriate. Guided by that idea, Weidmann [ 11 ] formulated a hierarchy of generalized instance-based assumptions for MIL. It consists of the standard MI assumption and three types of generalized MI assumptions, each more general than the last, in the sense that the former can be obtained as a specific choice of parameters of the latter,  $\text{standard} \subset \text{presence-based} \subset \text{threshold-based} \subset \text{count-based}$ , with the count-based assumption being the most general and the standard assumption being the least general. (Note however, that any bag meeting the count-based assumption meets the threshold-based assumption which in turn meets the presence-based assumption which, again in

turn, meet the standard assumption. In that sense it is also correct to state that the standard assumption is the weakest, hence most general, and the count-based assumption is the strongest, hence least general.) One would expect an algorithm which performs well under one of these assumptions to perform at least as well under the less general assumptions.

Presence-, threshold-, and count-based assumptions

The presence-based assumption is a generalization of the standard assumption, wherein a bag must contain all instances that belong to a set of required instance-level concepts in order to be labeled positive. Formally, let  $C_R \subseteq X \times Y$  be the set of required instance-level concepts, and let  $\#(B, c_i)$  denote the number of times the instance-level concept  $c_i$  occurs in the bag  $B$ . Then  $c(B) = 1 \Leftrightarrow \#(B, c_i) \geq 1$  for all  $c_i \in C_R$ . Note that, by taking  $C_R$  to contain only one instance-level concept, the presence-based assumption reduces to the standard assumption.

A further generalization comes with the threshold-based assumption, where each required instance-level concept must occur not only once in a bag, but some minimum (threshold) number of times in order for the bag to be labeled positive. With the notation above, to each required instance-level concept  $c_i \in C_R$  is associated a threshold  $l_i \in \mathbb{N}$ . For a bag  $B$ ,  $c(B) = 1 \Leftrightarrow \#(B, c_i) \geq l_i$  for all  $c_i \in C_R$ .

The count-based assumption is a final generalization which enforces both lower and upper bounds for the number of times a required concept can occur in a positively labeled bag. Each required instance-level concept  $c_i \in C_R$  has a lower threshold  $l_i \in \mathbb{N}$  and upper threshold  $u_i \in \mathbb{N}$  with  $l_i \leq u_i$ . A bag  $B$  is labeled according to  $c(B) = 1 \Leftrightarrow l_i \leq \#(B, c_i) \leq u_i$  for all  $c_i \in C_R$ .

GMIL assumption

Scott, Zhang, and Brown (2005) [ 12 ] describe another generalization of the standard model, which they call "generalized multiple instance learning" (GMIL). The GMIL assumption specifies a set of required instances  $Q \subseteq X$ . A bag  $X$  is labeled positive if it contains instances which are sufficiently close to at least  $r$  of the required instances  $Q$ . [ 12 ] Under only this condition, the GMIL assumption is equivalent to the presence-based assumption. [ 8 ] However, Scott et al. describe a further generalization in which there is a set of attraction points  $Q \subseteq X$  and a set of repulsion points  $\overline{Q} \subseteq X$ . A bag is labeled positive if and only if it contains instances which are sufficiently close to at least  $r$  of the attraction points and are sufficiently close to at most  $s$  of the repulsion points. [ 12 ] This condition is strictly more general than the presence-based, though it does not fall within the above hierarchy.

Collective assumption

In contrast to the previous assumptions where the bags were viewed as fixed, the collective assumption views a bag  $B$  as a distribution  $p(x|B)$  over instances  $X$ , and similarly view labels as a distribution  $p(y|x)$  over instances. The goal of an algorithm operating under the collective assumption is then to model the distribution  $p(y|B) = \int_X p(y|x) p(x|B) dx$ .

Since  $p(x|B)$  is typically considered fixed but unknown, algorithms instead focus on computing the empirical version:  $\hat{p}(y|B) = \frac{1}{n_B} \sum_{i=1}^{n_B} p(y|x_i)$ , where  $n_B$  is the number of instances in bag  $B$ . Since  $p(y|x)$  is also

typically taken to be fixed but unknown, most collective-assumption based methods focus on learning this distribution, as in the single-instance version. [ 8 ] [ 10 ]

While the collective assumption weights every instance with equal importance, Foulds extended the collective assumption to incorporate instance weights. The weighted collective assumption is then that  $\widehat{p}(y|B) = \frac{1}{\sum_{i=1}^n w(x_i)} \sum_{i=1}^n w(x_i) p(y|x_i)$ , where  $w: X \rightarrow \mathbb{R}^+$  is a weight function over instances and  $w_B = \sum_{x \in B} w(x)$ . [ 8 ]

## Algorithms

There are two major flavors of algorithms for Multiple Instance Learning: instance-based and metadata-based, or embedding-based algorithms. The term "instance-based" denotes that the algorithm attempts to find a set of representative instances based on an MI assumption and classify future bags from these representatives. By contrast, metadata-based algorithms make no assumptions about the relationship between instances and bag labels, and instead try to extract instance-independent information (or metadata) about the bags in order to learn the concept. [ 10 ] For a survey of some of the modern MI algorithms see Foulds and Frank. [ 8 ]

### Instance-based algorithms

The earliest proposed MI algorithms were a set of "iterated-discrimination" algorithms developed by Dietterich et al., and Diverse Density developed by Maron and Lozano-Pérez. [ 3 ] [ 9 ] Both of these algorithms operated under the standard assumption.

#### Iterated-discrimination

Broadly, all of the iterated-discrimination algorithms consist of two phases. The first phase is to grow an axis parallel rectangle (APR) which contains at least one instance from each positive bag and no instances from any negative bags. This is done iteratively: starting from a random instance  $x_1 \in B_1$  in a positive bag, the APR is expanded to the smallest APR covering any instance  $x_2$  in a new positive bag  $B_2$ . This process is repeated until the APR covers at least one instance from each positive bag. Then, each instance  $x_i$  contained in the APR is given a "relevance", corresponding to how many negative points it excludes from the APR if removed. The algorithm then selects candidate representative instances in order of decreasing relevance, until no instance contained in a negative bag is also contained in the APR. The algorithm repeats these growth and representative selection steps until convergence, where APR size at each iteration is taken to be only along candidate representatives.

After the first phase, the APR is thought to tightly contain only the representative attributes. The second phase expands this tight APR as follows: a Gaussian distribution is centered at each attribute and a looser APR is drawn such that positive instances will fall outside the tight APR with fixed probability. [ 4 ] Though iterated discrimination techniques work well with the standard assumption, they do not generalize well to other MI assumptions. [ 8 ]

### Diverse Density

In its simplest form, Diverse Density (DD) assumes a single representative instance  $t^*$  as the concept. This representative instance must be "dense" in that it is much closer to instances from positive bags than from negative bags, as well as "diverse" in that it is close to at least one instance from each positive bag.

Let  $B^+ = \{B_i^+ \mid 1 \leq i \leq m\}$  be the set of positively labeled bags and let  $B^- = \{B_i^- \mid 1 \leq i \leq n\}$  be the set of negatively labeled bags, then the best candidate for the representative instance is given by  $t^* = \arg \max_t DD(t)$ , where the diverse density  $DD(t) = P(t|B^+, B^-) = \arg \max_t \prod_{i=1}^m P(t|B_i^+) \prod_{i=1}^n P(t|B_i^-)$  under the assumption that bags are independently distributed given the concept  $t^*$ . Letting  $B_i^+$



$B_{ij}$  denote the  $j$ th instance of bag  $i$ , the noisy-or model gives:

$P(t | B_{ij})$  is taken to be the scaled distance  $P(t | B_{ij}) \propto \exp \left( - \sum_k s_k (x_k - (B_{ij})_k)^2 \right)$  where  $s = (s_k)$  is the scaling vector. This way, if every positive bag has an instance close to  $t$ , then  $P(t | B_{i+})$  will be high for each  $i$ , but if any negative bag  $B_{i-}$  has an instance close to  $t$ ,  $P(t | B_{i-})$  will be low. Hence,  $DD(t)$  is high only if every positive bag has an instance close to  $t$  and no negative bags have an instance close to  $t$ . The candidate concept  $\hat{t}$  can be obtained through gradient methods. Classification of new bags can then be done by evaluating proximity to  $\hat{t}$ . [ 9 ] Though Diverse Density was originally proposed by Maron et al. in 1998, more recent MIL algorithms use the DD framework, such as EM-DD in 2001 [ 13 ] and DD-SVM in 2004, [ 14 ] and MILES in 2006 [ 8 ]

A number of single-instance algorithms have also been adapted to a multiple-instance context under the standard assumption, including

Support vector machines [ 15 ]

Artificial neural networks [ 16 ]

Decision trees [ 17 ]

Boosting [ 18 ]

Post 2000, there was a movement away from the standard assumption and the development of algorithms designed to tackle the more general assumptions listed above. [ 10 ]

Weidmann [ 11 ] proposes a Two-Level Classification (TLC) algorithm to learn concepts under the count-based assumption. The first step tries to learn instance-level concepts by building a decision tree from each instance in each bag of the training set. Each bag is then mapped to a feature vector based on the counts in the decision tree. In the second step, a single-instance algorithm is run on the feature vectors to learn the concept

Scott et al. [ 12 ] proposed an algorithm, GMIL-1, to learn concepts under the GMIL assumption in 2005. GMIL-1 enumerates all axis-parallel rectangles  $\{R_i\}_{i \in I}$  in the original space of instances, and defines a new feature space of Boolean vectors. A bag  $B$  is mapped to a vector  $b = (b_i)_{i \in I}$  in this new feature space, where  $b_i = 1$  if  $APR R_i$  covers  $B$ , and  $b_i = 0$  otherwise. A single-instance algorithm can then be applied to learn the concept in this new feature space.

Because of the high dimensionality of the new feature space and the cost of explicitly enumerating all APRs of the original instance space, GMIL-1 is inefficient both in terms of computation and memory. GMIL-2 was developed as a refinement of GMIL-1 in an effort to improve efficiency. GMIL-2 pre-processes the instances to find a set of candidate representative instances. GMIL-2 then maps each bag to a Boolean vector, as in GMIL-1, but only considers APRs corresponding to unique subsets of the candidate representative instances. This significantly reduces the memory and computational requirements. [ 8 ]

Xu (2003) [ 10 ] proposed several algorithms based on logistic regression and boosting methods to learn concepts under the collective assumption.

Metadata-based (or embedding-based) algorithms

By mapping each bag to a feature vector of metadata, metadata-based algorithms allow the flexibility of using an arbitrary single-instance algorithm to perform the actual classification task. Future bags are simply mapped (embedded) into the feature space of metadata and labeled by the chosen classifier. Therefore, much of the focus for metadata-based algorithms is on what features or what type of embedding leads to effective classification. Note that some of the previously mentioned algorithms, such as TLC and GMIL could be considered metadata-based.

One approach is to let the metadata for each bag be some set of statistics over the instances in the bag. The SimpleMI algorithm takes this approach, where the metadata of a bag is taken to be a simple summary statistic, such as the average or minimum and maximum of each instance variable taken over all instances in the bag. There are other algorithms which use more complex statistics, but SimpleMI was shown to be surprisingly competitive for a number of datasets, despite its apparent lack of complexity. [ 8 ]

Another common approach is to consider the geometry of the bags themselves as metadata. This is the approach taken by the MIGraph and miGraph algorithms, which represent each bag as a graph whose nodes are the instances in the bag. There is an edge between two nodes if the distance (up to some metric on the instance space) between the corresponding instances is less than some threshold. Classification is done via an SVM with a graph kernel (MIGraph and miGraph only differ in their choice of kernel). [ 8 ] Similar approaches are taken by MILES [ 19 ] and MInD. [ 20 ] MILES represents a bag by its similarities to instances in the training set, while MInD represents a bag by its distances to other bags.

A modification of k-nearest neighbors (kNN) can also be considered a metadata-based algorithm with geometric metadata, though the mapping between bags and metadata features is not explicit. However, it is necessary to specify the metric used to compute the distance between bags. Wang and Zucker (2000) [ 21 ] suggest the (maximum and minimum, respectively) Hausdorff metrics for bags  $A$  and  $B$  :

They define two variations of kNN, Bayesian-kNN and citation-kNN, as adaptations of the traditional nearest-neighbor problem to the multiple-instance setting.

#### Generalizations

So far this article has considered multiple instance learning exclusively in the context of binary classifiers. However, the generalizations of single-instance binary classifiers can carry over to the multiple-instance case.

One such generalization is the multiple-instance multiple-label problem (MIML), where each bag can now be associated with any subset of the space of labels. Formally, if  $X$  is the space of features and  $Y$  is the space of labels, an MIML concept is a map  $c : N \times X \rightarrow 2^Y$ . Zhou and Zhang (2006) [ 22 ] propose a solution to the MIML problem via a reduction to either a multiple-instance or multiple-concept problem.

Another obvious generalization is to multiple-instance regression. Here, each bag is associated with a single real number as in standard regression. Much like the standard assumption, MI regression assumes there is one instance in each bag, called the "prime instance", which determines the label for the bag (up to noise). The ideal goal of MI regression would be to find a hyperplane which minimizes the square loss of the prime instances in each bag, but the prime instances are hidden. In fact, Ray and Page (2001) [ 23 ] show that finding a best fit hyperplane which fits one instance from each bag is intractable if there are fewer than three instances per bag, and instead develop an algorithm for approximation. Many of the algorithms developed for MI classification may also provide good approximations to the MI regression problem. [ 8 ]

See also

Supervised learning

Multi-label classification

References

Further reading

Recent reviews of the MIL literature include:

Amores (2013) , which provides an extensive review and comparative study of the different paradigms,

Foulds & Frank (2010) , which provides a thorough review of the different assumptions used by different paradigms in the literature.

Dietterich, Thomas G; Lathrop, Richard H; Lozano-Pérez, Tomás (1997). "Solving the multiple instance problem with axis-parallel rectangles". *Artificial Intelligence* . 89 ( 1– 2): 31– 71. doi : 10.1016/S0004-3702(96)00034-3 .

Herrera, Francisco; Ventura, Sebastián; Bello, Rafael; Cornelis, Chris; Zafra, Amelia; Sánchez-Tarragó, Dánel; Vluymans, Sarah (2016). *Multiple Instance Learning* . doi : 10.1007/978-3-319-47759-6 . ISBN 978-3-319-47758-9 . S2CID 24047205 .

Amores, Jaume (2013). "Multiple instance classification: Review, taxonomy and comparative study" . *Artificial Intelligence* . 201 : 81– 105. doi : 10.1016/j.artint.2013.06.003 .

Foulds, James; Frank, Eibe (2010). "A review of multi-instance learning assumptions". *The Knowledge Engineering Review* . 25 : 1– 25. CiteSeerX 10.1.1.148.2333 . doi : 10.1017/S026988890999035X . S2CID 8601873 .

Keeler, James D.; Rumelhart, David E.; Leow, Wee-Kheng (1990). "Integrated segmentation and recognition of hand-printed numerals". *Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems (NIPS 3)* . Morgan Kaufmann Publishers. pp. 557– 563. ISBN 978-1-55860-184-0 .

Li, Hong-Dong; Menon, Rajasree; Omenn, Gilbert S; Guan, Yuanfang (2014). "The emerging era of genomic data integration for analyzing splice isoform function" . *Trends in Genetics* . 30 (8): 340– 7. doi : 10.1016/j.tig.2014.05.005 . PMC 4112133 . PMID 24951248 .

Eksi, Ridvan; Li, Hong-Dong; Menon, Rajasree; Wen, Yuchen; Omenn, Gilbert S; Kretzler, Matthias; Guan, Yuanfang (2013). "Systematically Differentiating Functions for Alternatively Spliced Isoforms through Integrating RNA-seq Data" . *PLOS Computational Biology* . 9 (11): e1003314. Bibcode : 2013PLSCB...9E3314E . doi : 10.1371/journal.pcbi.1003314 . PMC 3820534 . PMID 24244129 . {{ cite journal }} : CS1 maint: article number as page number ( link )

Maron, O.; Ratan, A.L. (1998). "Multiple-instance learning for natural scene classification". *Proceedings of the Fifteenth International Conference on Machine Learning* . Morgan Kaufmann Publishers. pp. 341– 349. ISBN 978-1-55860-556-5 .

Kotzias, Dimitrios; Denil, Misha; De Freitas, Nando; Smyth, Padhraic (2015). "From Group to Individual Labels Using Deep Features". *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15* . pp. 597– 606. doi : 10.1145/2783258.2783380 . ISBN 9781450336642 . S2CID 7729996 .

Ray, Soumya; Page, David (2001). *Multiple instance regression (PDF)* . ICML.

Bandyopadhyay, Sanghamitra; Ghosh, Dip; Mitra, Ramkrishna; Zhao, Zhongming (2015). "MBSTAR: Multiple instance learning for predicting specific functional binding sites in microRNA targets" . *Scientific Reports* . 5 : 8004. Bibcode : 2015NatSR...5.8004B . doi : 10.1038/srep08004 . PMC 4648438 . PMID 25614300 .

Zhu, Wentao; Lou, Qi; Vang, Yeeleng Scott; Xie, Xiaohui (2017). "Deep Multi-instance Networks with Sparse Label Assignment for Whole Mammogram Classification". *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017* . *Lecture Notes in Computer Science*. Vol. 10435. pp. 603– 11. arXiv : 1612.05968 . doi : 10.1007/978-3-319-66179-7\_69 . ISBN 978-3-319-66178-0 . S2CID 9623929 .