Deep Multi-Instance Learning with Induced Self-Attention for Medical Image Classification

Zhenliang Li[1,2,3], Liming Yuan[1,2,3*], Haixia Xu[1,2,3], Rui Cheng[1,2,3], Xianbin Wen[1,2,3]
[1]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China
[2]Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin 300384, China
[3]Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin 300384, China

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313518

***Abstract*—Existing Multi-Instance learning (MIL) methods for medical image classification typically segment an image (bag) into small patches (instances) and learn a classifier to predict the label of an unknown bag. Most of such methods assume that instances within a bag are independently and identically distributed. However, instances in the same bag often interact with each other. In this paper, we propose an Induced Self-Attention based deep MIL method that uses the self-attention mechanism for learning the global structure information within a bag. To alleviate the computational complexity of the naive implementation of self-attention, we introduce an inducing point based scheme into the self-attention block. We show empirically that the proposed method is superior to other deep MIL methods in terms of performance and interpretability on three medical image data sets. We also employ a synthetic MIL data set to provide an intensive analysis of the effectiveness of our method. The experimental results reveal that the induced self-attention mechanism can learn very discriminative and different features for target and non-target instances within a bag, and thus fits more generalized MIL problems.**

***Index Terms*—multi-instance learning, deep learning, self-attention, medical image classification**

I. Introduction

A well-designed medical image classification system can assist doctors in making pathological diagnosis and locating possible lesions. Some fully supervised methods have achieved promising performance by using carefully labeled images [1–4]. However, providing fine-grained labels for images is very expensive or even impossible in medical image analysis. Alternatively, Multi-Instance Learning (MIL) [5], as one typical form of weakly supervised learning, depends on only coarsely labeled data, and thus is very attractive in solving medical image classification problems. Xu *et al.* [6] have demonstrated empirically that MIL methods using only coarse labels outperform its fully supervised counterparts.

The MIL methods regard a bag as the basic processing unit. Each medical image is represented as a bag of patches (instances), and each patch is represented by a fixed-dimensional feature vector. An image is marked as benign or malignant. The standard MIL assumption states that images labeled malignant contain at least one diseased patch, whereas images labeled benign contain only normal patches. Based on this assumption, MIL approaches [7] convert medical image problems from single instance learning to multiple instance learning, and design models with deep structure [8–11].

Most existing MIL methods typically consider instances within a bag as independently and identically distributed (i.i.d.) samples [12, 13]. These models process each instance independently before the feature aggregation, regardless of the important structure information among instances [11]. Not all data sets are suitable for this strategy, especially medical data sets. For example, histopathology images are complex structured objects composed of diseased areas, normal areas and areas with artifacts, over-staining, and failed autofocusing [14]. Treating different patches within one image as related instances is obviously more meaningful than treating them as independent instances.

In this paper, we propose a deep MIL method based on self-attention which uses structure information for medical image classification. We formulate the MIL model similar to the baselines [8, 9] and train it by optimizing the log-likelihood function. As shown in Fig. 1, the model consists of four components: (i) some convolutional layers used to generate instances; (ii) a permutation-invariant feature extractor; (iii) a permutation-invariant feature aggregator; and (iv) a final classifier used to produce the bag probability. Unlike baselines that assume instances in a bag are independent, we assume that instances are related to each other. We introduce the Self-Attention block (SA) as the feature extractor to learn high-level representations that contain structure information. Inspired by the inducing point [15], we further introduce the Induced Self-Attention block (ISA) as the feature extractor. Finally, we propose to replace the attention operator by a Multiquery Attention Pooling block (MAP). The main contributions of this paper are summarized as follows:

- We propose an Induced Self-Attention based deep MIL method for medical image classification. We show that the structure information is valuable and can help to improve classification performance and interpretability.
- The proposed method outperforms other methods on three medical image data sets. We further empirically verify that the proposed method fits more generalized MIL problems.

*Corresponding author. E-mail: yuanliming@tjut.edu.cn
This work was supported by the National Natural Science Foundation of China (61472278), the Major Project of Tianjin (18ZXZNGX00150), the Natural Science Foundation of Tianjin (18JCYBJC84800), the Scientific Research Key Project of Tianjin Municipal Education Commission (2017ZD13), and the Scientific Research Project of Tianjin Municipal Education Commission (2017KJ255).

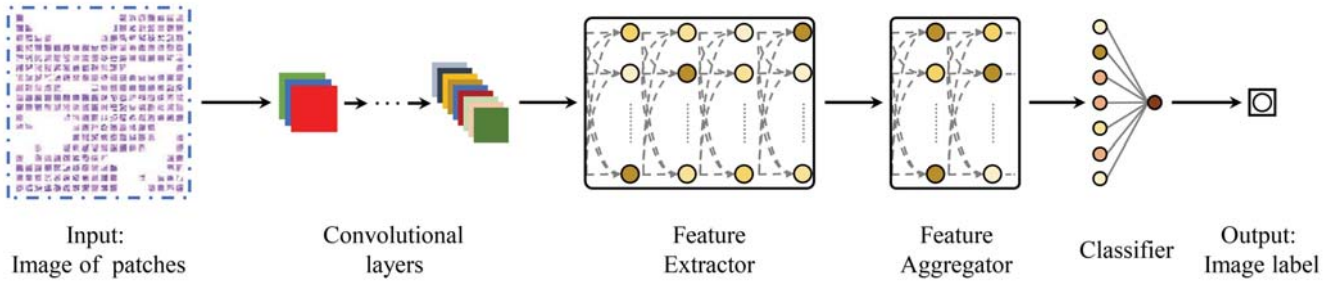


Fig. 1. The overall structure of the proposed Induced Self-Attention based deep multi-instance learning.

II. Proposed Method

A. Multi-Instance Learning

In multi-instance learning, one sample is a bag of instances, each bag has a binary label correspondingly, each instance has an unknown label. In this task, a bag is a medical image, a label is a diagnosis report (benign or malignant) of the image, and any instance is a small patch divided from the image which may involve lesions or ROIs. The standard MIL assumption is that positive bags contain at least one positive instance, negative bags contain only negative instances. However, generalized MIL assumption assume that any number of concepts exist in the data set [16]. Positive bags contain all concepts which are constrained by multiple positive examples under a rule. The remaining bags are negative bags.

Not all medical tasks can be converted to standard MIL problems, such as identifying antagonist drugs, protein superfamily identification [17]. However, these complex tasks can be handled under the generalized MIL assumption.

B. Set transformer

Set Transformer [15] is an attention-based permutation-invariant neural network module for set-input problems. Similar to other transformer architectures, a Set transformer includes an encoder and a decoder, each of which is constructed by some blocks such as MAB, ISAB, PMAB.

MAB. Multihead Attention, firstly proposed by Vaswani *et al.* [18], consists of h different heads to learn information and representation from different subspaces independently. In order to process sets of data, key and value in attention are from the same input $Y \in \mathbb{R}^{n\times d}$, and query is given by another input $X \in \mathbb{R}^{n\times d}$. The formal definition is

$$\begin{aligned} &\mathrm{MAB}(X,Y) = \mathrm{LayerNorm}(H + \mathrm{rFF}(H)),\\ &\text{where } H = \mathrm{LayerNorm}(X + \mathrm{Multihead}(X,Y,Y)). \end{aligned} \tag{1}$$

ISAB. Induced Set Attention Block (ISAB) is proposed to bypass the quadratic computational complexity caused by pairwise dot product. Given a input $X \in \mathbb{R}^{n\times d}$ and inducing points $I \in \mathbb{R}^{m\times d}$. The ISAB first maps the inducing points I to intermediate variable H using the input X, and then maps input X to the outputs using H:

$$\begin{aligned} &\mathrm{ISAB}_m(X) = \mathrm{MAB}(X,H) \in \mathbb{R}^{n\times d},\\ &\text{where } H = \mathrm{MAB}(I,X) \in \mathbb{R}^{m\times d}. \end{aligned} \tag{2}$$

Attention is computed between a set of size m and a set of size n in (2). Therefore, the time complexity is reduced from $\mathcal{O}(n^2)$ to $\mathcal{O}(nm)$.

PMAB. Different to manual permutation-invariant aggregation operators, Pooling by Multihead Attention Block (PMAB) is a pooling operator based on multihead attention. The learnable parameters are a set of k seed vectors $S \in \mathbb{R}^{k\times d}$. Given the input $X \in \mathbb{R}^{n\times d}$, the formal definition is

$$\mathrm{PMAB}_k(X) = \mathrm{MAB}(S, \mathrm{rFF}(X)) \in \mathbb{R}^{k\times d}. \tag{3}$$

C. Induced Self-Attention based MIL

In this section, we present three blocks of the proposed method: Self-Attention block (SA) to model interactions among instances; Induced Self-Attention block (ISA) to reduce time complexity of dot product; and Multiquery Attention Pooling block (MAP) to flexibly aggregate features.

SA. Self attention attends attention to the input by using the input itself. This method learns global context information of input by computing the inner product between instances. Given a bag of instances as input $X = \{x_1, x_2, \ldots, x_N\} \in \mathbb{R}^{n\times d_i}$. Queries $Q \in \mathbb{R}^{n\times d_o}$, keys $K \in \mathbb{R}^{n\times d_o}$, values $V \in \mathbb{R}^{n\times d_o}$ are derived from different linear projections of the input. The formal description is

$$\mathrm{Att}(Q,K,V;\omega) = \omega(QK^T)V, \tag{4}$$

$$\mathrm{SA}(X;\tilde{\omega}) = \mathrm{Att}(XW^Q, XW^K, XW^V;\tilde{\omega}) \in \mathbb{R}^{n\times d_o}. \tag{5}$$

The linear projections are $W^Q \in \mathbb{R}^{d_i\times d_o}$, $W^K \in \mathbb{R}^{d_i\times d_o}$, $W^V \in \mathbb{R}^{d_i\times d_o}$, $d_o = d_i/2$. The matrix QK^T describes the dependencies among instances, and a value in the matrix gets greater if the corresponding instances have a closer relation. We use a scaled softmax $\omega(\cdot) = \mathrm{softmax}(\cdot/\sqrt{d})$ to compute weights. And the output is a weighted sum of the input.

Note that we also replace common softmax $\omega(\cdot)$ with the scaled α-entmax $\tilde{\omega}(\cdot)$ inspired by Correia *et al.* [19]. The α-entmax is an adaptive function which can learn sparsity of the mapping automatically with a learnable parameter α:

$$\alpha\text{-entmax}(z) = \underset{p\in\triangle^d}{\mathrm{argmax}}\, p^T z + \mathrm{H}^T_\alpha(p), \tag{6}$$

$$\mathrm{H}^T_\alpha(p) = \begin{cases} \frac{1}{\alpha(\alpha-1)}\sum_j (p_j - p_j^\alpha), & \alpha \neq 1,\\ -\sum_j p_j \log p_j, & \alpha = 1. \end{cases} \tag{7}$$

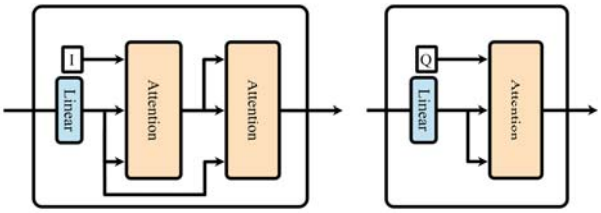


Fig. 2. Diagrams of ISA block (left) and MAP block (right).

where $\triangle^d := \{p \in \mathbb{R}^d : \sum_i p_i = 1\}$ is the *probability simplex*, and $\mathrm{H}_\alpha^T(p)$ is the Tsallis continuous family of entropies for $\alpha \geq 1$, (7).

ISA. In order to reduce the time consumption of the dot product, we introduce an Induced Self-Attention block (ISA) inspired by ISAB from Set Transformer. Unlike ISAB, while reducing computational complexity, ISA also reduces the number of trainable parameters. Experimental results show that sparse parameters can avoid the model overfitting on small samples. As is shown in Fig. 2, let $X \in \mathbb{R}^{n \times d_i}$ be a bag of n feature vectors, $I \in \mathbb{R}^{m \times d_o}$ be the inducing points:

$$\begin{aligned} &\mathrm{ISA}_m(X) = \mathrm{Att}(X, H, H; \tilde{\omega}) \in \mathbb{R}^{n \times d_o}, \\ &\text{where } H = \mathrm{Att}(I, XW, XW; \tilde{\omega}) \in \mathbb{R}^{m \times d_o}. \end{aligned} \tag{8}$$

The ISA_m has learnable parameters $\{I, W\}$, where $W \in \mathbb{R}^{d_i \times d_o}$. The transformation XW is implemented by a full connection layer without activation functions. d_i, d_o are the input dimension and output dimension of ISA_m layer, respectively. A typical choice for the dimension hyperparameters is $d_o = d_i/2$. Additionally, we use α-entmax instead of softmax.

MAP. Existing MIL pooling operators can be categorized into two groups: manual, non-trainable operators and adaptive, trainable operators. The former (e.g., the max/mean/lse operator [8]) computes faster, but lacks interpretability, and does not distinguish the importance of different instances. The latter (e.g., attention/gated attention operator [9]) is interpretable and can adapt to different tasks and data sets. PMAB is a trainable operator which aggregates vectors by multihead attention. We design a trainable operator which aggregates vectors by multiquery attention. As is shown in Fig. 2, let $X \in \mathbb{R}^{n \times d_i}$ be a bag of n feature vectors, $Q \in \mathbb{R}^{k \times d_o}$ be the queries:

$$\mathrm{MAP}_k(X) = \mathrm{Att}(Q, XW, XW; \tilde{\omega}) \in \mathbb{R}^{k \times d_o}. \tag{9}$$

Where $W \in \mathbb{R}^{d_i \times d_o}$ is the learnable parameters. The MAP_k can be regarded as the combination of a fully connected layer and a multiquery attention layer. Thus, the quantity, dimension of output can be controlled by hyperparameter k, d_o, respectively. We use only one query ($k = 1$) in our model which achieves a similar effect to attention operator.

D. The Overall Architecture.

In this paper, we propose an Induced Self-Attention based deep MIL method (Fig. 1), which includes a feature extractor that converts instances into embeddings with structure information, and a feature aggregator that merges the embeddings into the bag representation. We uses the proposed blocks (i.e., SA, ISA) as the feature extractor and MAP as the feature aggregator. In addition, to evaluate the Set Transformer, We embed ISAB and PMAB into this framework to classify medical images.

III. Experiments

In all experiments, we use the same evaluation methodology as baselines [8, 9]. In each experiment, we ran 5 times 10-fold CV with different random seeds, and report the average performance metrics over 5 trials. In the case of MNIST-bags, just like baselines, we adopt a fixed division for training and test set. In addition to the baselines, we compared various pipelines composed of different extractor and aggregator.

- MI-Net [8]: rFFs in extractor, max/mean/lse operator in aggregator.
- Att-Net & Gated Att-Net [9]: rFFs in extractor, attention operator or gated attention operator in aggregator.
- SetTrans [15]: ISAB in extractor, PMAB in aggregator.
- SA-MIL (ours): SA in extractor, MAP in aggregator.
- ISA-MIL (ours): ISA in extractor, MAP in aggregator.

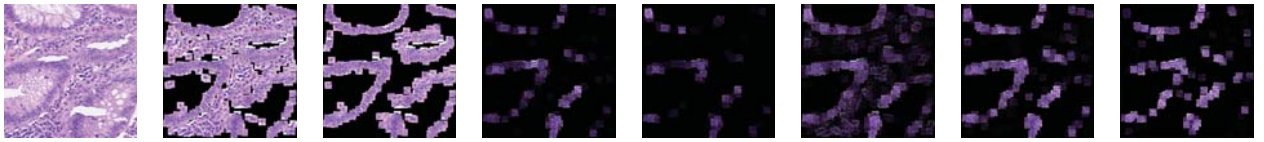
TABLE I
Accuracy of different methods on UCSB breast and Messidor.

Methods	UCSB breast	Messidor
MI-Net [8]	0.832	0.691
Att-Net [9]	0.867	0.697
Gated Att-Net [9]	0.874	0.698
SetTrans [15]	0.887	0.702
SA-MIL(ours)	**0.891**	**0.726**
ISA-MIL(ours)	0.871	0.718

A. UCSB breast and Messidor

Details: UCSB breast consists of 58 TMA images from 32 benign and 26 malignant breast cancer patients [20]. Each medical image of patient is treated as a bag of some instances. Messidor is an eye fundus image data set for diabetic retinopathy diagnosis [21]. This data set consists of 654 malignant images from diabetic patients and 546 benign images without manual annotations.

Results and discussion: Compared with baselines, the self-attention based approaches (i.e., SA-MIL, ISA-MIL, SetTrans) achieve the best results on both UCSB breast and Messidor data sets, which indicates that the structure information can effectively improve the classification ability of the model. By comparing three self-attention based approaches, it can be found that multihead scheme improve less the classification performance of the model, the inducing point scheme reduces the training time while causing a slight performance loss on UCSB breast. Under the same experimental conditions, the experiment on UCSB breast shows that SetTrans method is 0.12 seconds /epoch slower than SA-MIL, and SA-MIL is slightly slower than ISA-MIL.



(a) Original (b) Patches (c) Ground Truth (d) ISA-MIL (e) SA-MIL (f) SetTrans (g) Att-Net (h) Gated Att-Net

Fig. 3. The last five columns are heatmaps from different models. Heatmap: each patch from the second columns multiplied by its attention weight, we recalled the attention weight using $a'_k = (a_k - \min(\mathbf{a}))/(\max(\mathbf{a}) - \min(\mathbf{a}))$.

TABLE II
RESULTS OF DIFFERENT METHODS ON COLON CANCER.

Methods	Accuracy	Precision	Recall	F-score	AUC
Att-Net [9]	0.872	0.885	0.885	0.870	0.873
Gated Att-Net [9]	0.868	0.870	0.887	0.860	0.883
SetTrans [15]	0.900	0.912	0.906	0.900	0.900
SA-MIL(ours)	0.910	0.926	0.911	0.914	0.912
ISA-MIL(ours)	**0.920**	**0.928**	**0.931**	**0.927**	**0.914**

B. Colon cancer

Details: Colon cancer involves 100 H&E histology images [14]. In each image, most of the nuclei are marked as one of four class labels, i.e., epithelial, inflammatory, fibroblast, and miscellaneous. Similar to UCSB breast, an image is divided into a bag of blocks. Assume that the bag is malignant when the corresponding image contains one or more epithelial, the data set is then divided into 51 bags labeled malignant and 49 bags labeled benign. Due to the limited amount of data samples, we perform data augmentation to avoid overfitting.

Results: We present results in Table II for Colon cancer. First, the results show that our approaches outperform all other methods. The structure information captured by self-attention makes task easier to deal with. Besides, compared to baselines, the proposed method is not only most accurate but it also received the highest recall, an important metric in the medical field. Finally, we notice that ISA performs better than SA in all metrics.

Interpretability: In order to show the effectiveness of different attention mechanism in providing ROIs, we plot heatmaps of the same test image in different methods (Fig. 3). Patches are the small areas centered around all marked nuclei, ground truth is the patches that belong to the class epithelial. Compared to ground truth, we notice that the heatmap in ISA-MIL highly matches the ground truth. SA-MIL tends to select a subset of positive patches while SetTrans considers many invalid, negative patches. Compared to the baselines, we find that ISA-MIL attends less attention on negative patches. The background (black areas) of the heatmap in ISA-MIL is purer than of Att-Net. The experimental results show the proposed method not only has good classification performance but also provides ROIs which assists in pathological diagnosis.

C. MNIST-bags

Details: We construct a synthetic MIL data using MNIST image data set. Each bag consists of a random quantity of grayscale images from MNIST data set. The quantity of images follows a Gaussian-distribution centered on a given value. The mean and variance of the bag size are set to 10, 2 respectively. Assume that a bag containing both the digits '7' and '9' is positive and the others are negative. We choose '7' and '9' because they are similar. We use varying numbers of bags for training and 1000 bags for testing.

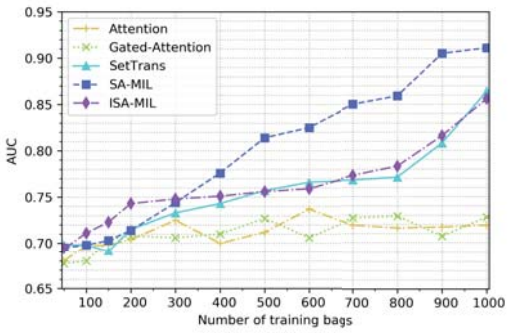


Fig. 4. The AUC curves of different methods on MNIST-bags.

Results: The AUC curves of different methods are shown in Fig. 4. The figure indicates that: (i) The proposed method achieves significantly higher AUC than other methods. Self-Attention mechanism learn well about the complex relation between different digits; (ii) ISA-MIL performs as well as SetTrans, and SA-MIL performs much better when the size of training bags is greater than 300. (iii) The attention operator does not work well in the task based generalized MIL assumption, especially on the large samples data.

How to Self-Attention: As shown in the Fig. 5, the colors of the heatmaps are distributed in stripes rather than diagonals, and the key, positive images (i.e., digit '7', '9') correspond to the darker color. We believe that the attention matrix measures the dependence rather than the structure similarity among images in a bag. Additionally, all queries are high corresponding to the positive images, not to themselves, which indicates that self-attention mechanism can locate key instances, and separate positive and negative instances.

We also do another visualization work (Fig. 6). In the top right, we find that the 3D-vectors labeled in the same class have similar orientation, and the positive vectors have longer length, just like the "capsule". This phenomenon explains why the inner product of the positive image and other images is greater than of the other images themselves. In the bottom, all digits '7' and '9' are highlighted and the sum is 0.875. Three

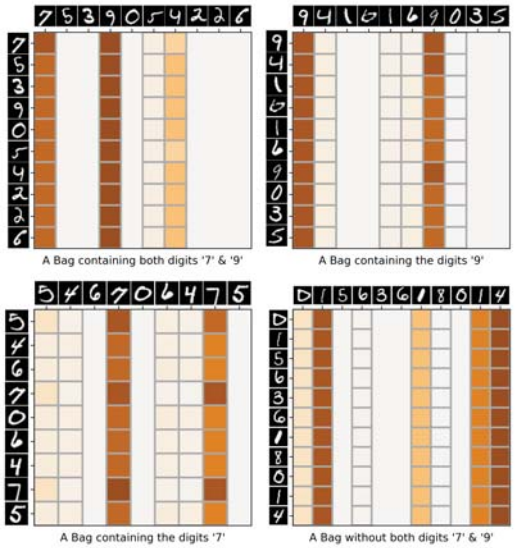


Fig. 5. The heatmaps of self-attention matrices of 4 different MNIST-bags. X axis (top) indicates the input (a bag of images) and Y axis (left) indicates the query items (the mapped images). Digits '7' and '9' are positive instances. The colormap values represent the dependency between images. In order to make the figure more contrasting, we adjust the range of colormap values.

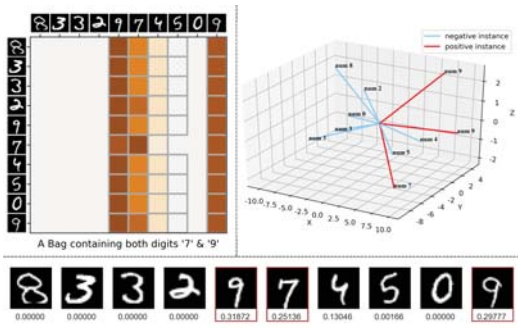


Fig. 6. Three visualizations of proposed ISA-MIL for a MNIST-bag. The top left is the heatmap of attention matrix of **ISA.**; the top right is visualization of the 3-D vectors, which are the results from the PCA for the embedding XW^V in **ISA.**; the bottom is the global weights which provide the contribution of the corresponding image to the final decision. The global weights are multiplied by the weights in **ISA.** and the weights in **MAP.**.

figures reveal that the induced self-attention mechanism can learn very discriminative and different features for target and non-target instances within a bag.

IV. Conclusion and Future Work

In this paper, we proposed induced self-attention based deep multi-instance learning for medical image classification. The inducing point based scheme is introduced into the self-attention block for learning the global structure information. We showed empirically that proposed method outperforms other deep MIL methods. Besides, it can handle medical tasks based on the generalized MIL assumption. In the future, we will apply our method on more challenging medical problems under the generalized MIL assumption, such as protein superfamily identification.

References

[1] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature medicine*, vol. 24, no. 10, pp. 1559–1567, 2018.

[2] C. D. Malon and E. Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *Journal of pathology informatics*, vol. 4, 2013.

[3] K. Nguyen, A. K. Jain, and B. Sabata, "Prostate cancer detection: Fusion of cytological and textural features," *Journal of pathology informatics*, vol. 2, 2011.

[4] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[5] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.

[6] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, I. Eric, and C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 1626–1630.

[7] Y. Chen, J. Bi, and J. Z. Wang, "Miles: Multiple-instance learning via embedded instance selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1931–1947, 2006.

[8] X. Wang, Y. Yan, P. Tang, X. Bai, and W. Liu, "Revisiting multiple instance neural networks," *Pattern Recognition*, vol. 74, pp. 15–24, 2018.

[9] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," *arXiv preprint arXiv:1802.04712*, 2018.

[10] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, and W. Zhang, "Accurate screening of covid-19 using attention based deep 3d multiple instance learning," *IEEE Transactions on Medical Imaging*, 2020.

[11] M. Tu, J. Huang, X. He, and B. Zhou, "Multiple instance learning with graph neural networks," *arXiv preprint arXiv:1906.04881*, 2019.

[12] Z.-H. Zhou and J.-M. Xu, "On the relation between multi-instance learning and semi-supervised learning," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 1167–1174.

[13] Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li, "Multi-instance learning by treating instances as non-iid samples," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 1249–1256.

[14] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[15] J. Lee, Y. Lee, J. Kim, A. Kosiorek, S. Choi, and Y. W. Teh, "Set transformer: A framework for attention-based permutation-invariant neural networks," in *International Conference on Machine Learning*, 2019, pp. 3744–3753.

[16] J. R. Foulds and E. Frank, "A review of multi-instance learning assumptions," 2010.

[17] S. Scott, J. Zhang, and J. Brown, "On generalized multiple-instance learning," *International Journal of Computational Intelligence and Applications*, vol. 5, no. 01, pp. 21–35, 2005.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[19] G. M. Correia, V. Niculae, and A. F. Martins, "Adaptively sparse transformers," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 2174–2184.

[20] M. Kandemir, C. Zhang, and F. A. Hamprecht, "Empowering multiple instance histopathology cancer diagnosis by cell graphs," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 228–235.

[21] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed image database: the Messidor database," *Image Analysis and Stereology*, pp. 231–234, 2014.