# Novelty Detection Based Discriminative Multiple Instance Feature Mining to Classify NSCLC PD-L1 Status on HE-Stained Histopathological Images

Rui Xu[1,2], Dan Yu[1], Xuan Yang[1], Xinchen Ye[1,2], Zhihui Wang[1,2], Yi Wang[1,2], Hongkai Wang[3](✉), Haojie Li[4], Dingpin Huang[5,7], Fangyi Xu[5,7], Yi Gan[6], Yuan Tu[5,7], and Hongjie Hu[5,7](✉)

[1] DUT School of Software Technology & DUT-RU International School of Information Science and Engineering, Dalian University of Technology, Dalian, China
[2] DUT-RU Co-Research Center of Advanced ICT for Active Life, Dalian, China
[3] Faculty of Medicine, Dalian University of Technology, Dalian, China
wang.hongkai@dlut.edu.cn
[4] Shandong University of Science and Technology, China
[5] Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China
[6] Department of Pathology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China
[7] Medical Imaging International Scientific and Technological Cooperation Base of Zhejiang Province, Hangzhou, China
hongjiehu@zju.edu.cn

**Abstract.** It is crucial to analyze HE-stained histopathological whole slide images (WSIs) to classify PD-L1 status for non-small cell lung cancer (NSCLC) patients, due to the expensive immunohistochemical examination performed in practical clinics. Usually, a multiple instance learning (MIL) framework is applied to resolve the classification problems of WSIs. However? existing MIL methods cannot perform well in PD-L1 status classification, due to unlearnable instance features and challenging instances that contain weak visual differences. To address this problem, we propose a novelty detection based discriminative multiple instance feature mining method. It contains a trainable instance feature encoder, learning effective information from the on-hand dataset to reduce the domain difference problem, and a novelty detection based instance feature mining mechanism, selecting typical instances to train the encoder for mining more discriminative instance features. We evaluate the proposed method on a private NSCLC PD-L1 dataset and the widely used public Camelyon16 [1] dataset that is targeted for breast cancer identification. The experimental results show that the proposed method is not only effective in predicting the status of NSCLC PD-L1 but also generalized well in the public data set.

**Keywords:** Multiple instance learning · Novelty detection · Feature mining · Histopathological images · Lung cancer.

## 1 Introduction

Immunotherapy has made significant progress in the treatment of non-small cell lung cancer (NSCLC) in recent years [2, 5, 7, 8, 16]. Before immunotherapy, an immunohistochemical (IHC) examination is necessary to determine cancer biomarkers for patients, such as the status of programmed death ligand 1 (PD-L1) (positive or negative). However, IHC examination is expensive, increasing the economic burden on patients. Histopathological imaging stained with hematoxylin and eosin (HE-stained) is a routine medical examination for the diagnosis of lung cancer, and its cost is significantly lower compared to the IHC examination. Although areas related to PD-L1 cannot be highlighted in HE-stained images compared to IHC images, information related to PD-L1 is potentially embedded in them. Thus, it is possible to analyze the HE-stained images to classify the status of PD-L1, instead of using the expensive IHC examination.

HE-stained histopathological images are whole slide images (WSIs) that contain tens of thousands of pixels in horizontal and vertical directions. Usually, a multiple instance learning (MIL) framework [4, 14] is applied to solve the WSI classification problem. A WSI is treated as a bag and divided into a bunch of smaller image patches, which are called instances. A bag is considered positive only if it contains at least one positive instance, and negative if it contains no positive instances. Although current MIL methods have achieved good results in some WSI classification tasks [11, 12, 18, 20], such as cancer identification, they cannot perform well on our task.

Current MIL methods for WSI classification mainly focus on aggregating bag-level features, while neglecting the design of instance-level feature mining mechanisms. This leads to poor PD-L1 classification results. On the one hand, they often directly utilize pre-trained networks on pathology image patches to extract instance features. Due to that instance features are frozen during training, and domain difference is inevitably involved in MIL. On the other hand, PD-L1 status information is latently displayed in HE-stained images, and negative and positive instances have weak visual differences, unlike in the case of cancer identification, showing visually distinct instances. Thus, in our task, it is difficult for the network to find typical instances during the training, which is only supervised by bag labels [3, 13]. To address these problems, we propose a novelty detection based discriminative multiple instance feature mining method in this paper.

The proposed method contains an instance feature encoder, consisting of a frozen pre-trained network followed by a learnable network module. This means that the instance encoder can learn effective information from the on-hand dataset, reducing the problem of domain difference. We exploit the fact that negative bags contain only negative instances to select typical instances to drive the learnable module to extract more discriminative instance features. Specifically, all instances in negative bags are treated as normal samples, while possible positive instances in positive bags are considered novelty targets. Thus, we select all instances in negative bags as typical negative ones and exploit them to learn a more compact feature embedding, which has a tighter hypersphere for negative

instance features. Besides, this hypersphere can be exploited to select typical instances on positive bags, by measuring the distances between the hypersphere centroid and all instances' features. On positive bags, typical positive instances are the ones that are far away from the centroid, while typical negative ones are those that are near the centroid. Then, the selected typical instance features are further mined by a contrastive learning loss. Due to the carefully designed instance feature mining mechanism, more discriminative instance features can be learned, leading to better results for MIL. The proposed method is evaluated by using a private NSCLC PD-L1 dataset. Experimental results show that our method outperforms the existing leading MIL methods. In addition, it is also validated using a public Camelyon16 [1] data set, which is aimed at identifying breast cancer and widely used in the MIL literature. It is found that our method performs equivalently or even better than the state-of-the-art methods on the public dataset, exhibiting good generalization.

The contributions of this paper are summarized as follows: 1) We propose a method that analyzes HE-stained histopathological images to classify NSCLC PD-L1 status, aiming to replace the expensive IHC examination. 2) We design a novelty detection based instance feature mining mechanism that can select typical instances for mining more discriminative instance features. 3) We demonstrate that our method is not only effective in predicting NSCLC PD-L1 status and generalized well for breast cancer identification.

## 2 Novelty Detection Based Discriminative Multiple Instance Feature Mining

### 2.1 Overview of the Proposed Method

Fig. 1 provides an overview of the proposed method. Given that there are total $N$ WSIs, in MIL, any input WSI, denoted as $X_i$, is considered as a bag of multiple instances, represented as $X_i = \{x_{ij}\}_{j=1}^{N_i}$. $x_{ij}$ is the $j$-th image patch obtained from the $i$-th bag by using a pre-processing step to remove background areas followed by splitting, and $N_i$ gives the number of instances in the $i$-th bag. Each bag has a bag-level label, denoted $Y_i \in \{0, 1\}$, indicating that the bag is positive (i.e. $Y_i = 1$) or negative (i.e. $Y_i = 0$). It should be noted that instance-level labels are unknown. The goal of MIL is to learn a mapping $\mathcal{M}(\cdot)$, which predicts a label for a given bag $\hat{Y}_i \leftarrow \mathcal{M}(X_i)$. To this end, the popular solution is to input all instances into an instance feature encoder $\mathcal{F}(\cdot)$ to obtain $D$-dimensional instance level features $z_{ij} \in \mathbb{R}^D \leftarrow \mathcal{F}(x_{ij})$, and then produce the corresponding bag-level feature $Z_i \in \mathbb{R}^D$ from these instance level features based on an aggregation step. Finally, a classifier $\mathcal{C}(\cdot)$ is trained to predict the bag-level label $\hat{Y}_i \leftarrow \mathcal{C}(Z_i)$. In this paper, the bag-level feature is obtained by a widely-used attention-based aggregation method [9], which can be expressed by $Z_i = \frac{1}{N_i} \sum_{j=1}^{N_i} a_j z_{ij}$, where $a_j$ is a learnable scalar weight. These attention weights together with the classifier parameters are trained by a cross entropy

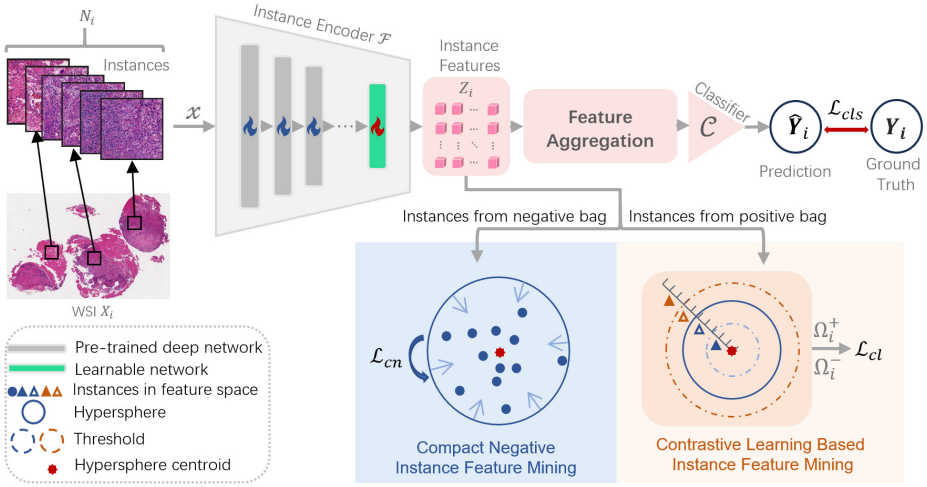


**Fig. 1.** Overview of the proposed novelty detection based discriminative multiple instance feature mining method.

loss, whose definition is given by Eq. 1.

$$\mathcal{L}_{cls} = \frac{1}{N}\sum_{i=1}^{N}[Y_i \log \hat{Y}_i + (1 - Y_i)\log(1 - \hat{Y}_i)] \tag{1}$$

Usually, the existing MIL methods directly apply a pre-trained deep network as the instance feature encoder, and freeze the instance-level features during the training. Due to the difference between the pre-trained and on-hand datasets, a domain gap is inevitably introduced into MIL. To address this, we adopt a partially trainable instance feature encoder, consisting of a frozen pre-trained network followed by a learnable network module. Compared to a fully trainable encoder, this design suits the classification task on WSIs. Due to the huge data of WSIs, GPU memory is too limited to be allocated for a fully trainable encoder with a large number of learnable parameters. Besides, more parameters make the network training hard. Thus, we adopt a partially trainable encoder whose learnable network module is a fully connected layer. This design leverages the knowledge embedded in the pre-trained network. In addition, it allows the encoder to learn information from the on-hand dataset, weakening the effect of the domain gap. The trainable parameters of the encoder are governed by not only the classifier loss ($\mathcal{L}_{cls}$) but also two carefully designed instance feature mining losses. They are the compact negative instance feature mining loss ($\mathcal{L}_{cn}$) and the contrastive learning based instance feature mining loss ($\mathcal{L}_{cl}$). Both of them are designed by integrating novelty detection with MIL, and their details are described in the following two subsections.

### 2.2 Compact Negative Instance Feature Mining

According to the definition of MIL, negative bags do not contain positive instances. This allows us to safely treat all instances in the negative bag as typical negative instances to find a better feature embedding for negative instances. An ideal feature embedding should be compact, which motivates us to exploit a deep one-class classification method [17] to constrain the features of typical negative instances inside a smaller hypersphere. This can be formulated as a compact negative instance feature mining loss $\mathcal{L}_{cn}$, whose definition is given by Eq. 2.

$$\mathcal{L}_{cn} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}[Y_i = 0] \frac{1}{N_i}\sum_{j=1}^{N_i} \|z_{ij} - \mu\|^2 \tag{2}$$

where $\mathbb{1}[Y_i = 0]$ is an indicator function, which gives 1 for a negative bag ($Y_i = 0$) and 0 for a positive bag ($Y_i = 1$), and $\mu$ is the hypersphere centroid calculated from all instance features and updated during training.

It can be seen that the loss function $\mathcal{L}_{cn}$ compresses typical negative instances into a feature embedding, where all of them are brought closer to each other. Thus, a compact negative feature mining can be achieved. Besides, this feature embedding can also be used to select typical instances on positive bags and motivate us to design the contrastive learning-based instance feature mining, whose details are described in the next subsection.

### 2.3 Contrastive Learning Based Instance Feature Mining

The MIL definition also tells us that there are both positive and negative instances on a positive bag. Due to that different instances show similar visual information for the PD-L1 classification problem, we take advantage of the compact negative instance feature embedding to find typical positive and negative instances on positive bags. We treat this as a novelty detection problem. Here, the negative instances are considered normal samples, and positive instances are novelty targets that should be detected. Remembering that the compact negative instance feature mining has already taken effect, features of normal samples (negative instances) should be constrained inside a small hypersphere, and features of novelty targets (positive instance) should be located at the boundary of the hypersphere. Taking advantage of this assumption, we can select typical positive and negative instances on positive bags by measuring distances between instance features and the hypersphere centroid. For all instance features on a positive bag, we calculate the distances between these features and the hypersphere centroid and normalize the distances inside the range from 0 to 1. Typical instance selection can be formulated as in Eq. 3.

$$\begin{cases} z_{ij} \in \Omega_i^+, & \text{if } ||z_{ij} - \mu|| > T_p \\ z_{ij} \in \Omega_i^-, & \text{if } ||z_{ij} - \mu|| < T_n \end{cases} \tag{3}$$

where $||z_{ij} - \mu||$ is the normalized distance between the instance feature $z_{ij}$ and the hypersphere centroid $\mu$, $\Omega_i^+$ is a set of typical positive instances whose

normalized distance is greater than the threshold value of $T_p$, and $\Omega_i^-$ is another set of typical negative instances whose normalized distance is smaller than the threshold value of $T_n$. The two threshold values are hyper-parameters determined in experiments.

After obtaining the two typical instance sets, we exploit contrastive learning to make features of the two sets to be far away from each other. Inspired from the InforNCE loss [15], we define the contrastive learning based instance feature mining loss $\mathcal{L}_{cl}$ as Eq. 4.

$$\mathcal{L}_{cl} \propto \sum_{i=1}^{N} \mathbb{1}[Y_i = 1] \sum_{k \in \Omega_i^+} \sum_{j^+ \in \Omega_i^+} -\log \frac{e^{\cos(z_{ik}, z_{ij^+})}}{e^{\cos(z_{ik}, z_{ij^+})} + \sum_{j^- \in \Omega_i^-} e^{\cos(z_{ik}, z_{ij^-})}} \quad (4)$$

where $\mathbb{1}[Y_i = 1]$ is an indicator function whose value is 1 for positive bags ($Y_i = 1$) and 0 for negative bags ($Y_i = 0$), $\cos(\cdot)$ is a cosine similarity function.

In summary, the total loss of the proposed method is formulated as Eq. 5

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \alpha \mathcal{L}_{cn} + \beta \mathcal{L}_{cl} \quad (5)$$

where $\alpha$ and $\beta$ are hyper-parameters used to balance the various losses.

## 3 Experiments

### 3.1 Dataset and Implementation Details

**Datasets and Preprocessing.** The proposed method was evaluated using two datasets: 1) The PD-L1 dataset is used to predict the status of PD-L1 from histopathology images stained with HE in lung cancer. It consists of 401 lung cancer HE-stained pathology images, with 191 negative and 210 positive. 2) Camelyon16 [1] is a publicly available dataset for detecting breast cancer metastasis. It includes 270 training images and 129 test images. The WSIs in both the PD-L1 dataset and the Camelyon16 [1] dataset are divided into image patches by applying a sliding window at 20× magnification level. In the PD-L1 dataset, the patch size is $1792 \times 1792$, resulting in a total of 0.16 million patches. The Camelyon16 [1] dataset consists of patches of size $512 \times 512$, totaling 3.85 million patches. During this process, patches that have a foreground area of less than 10% after binarization will be discarded as background.

**Evaluation Protocol.** The data in the PD-L1 dataset is divided into three sets: training, validation, and test, with a ratio of 4:1:1. For Camelyon16 [1], the model was trained using 80% of the training images for the training set and 20% for the validation set, with testing conducted on the official test set. The proposed model was evaluated based on the average results obtained from five random dataset partitions. We utilize five metrics to evaluate the classification performance of the model: the area under the receiver operating characteristic curve (AUC), accuracy (ACC), sensitivity (SE), specificity (SP), and F1 Score (F1).

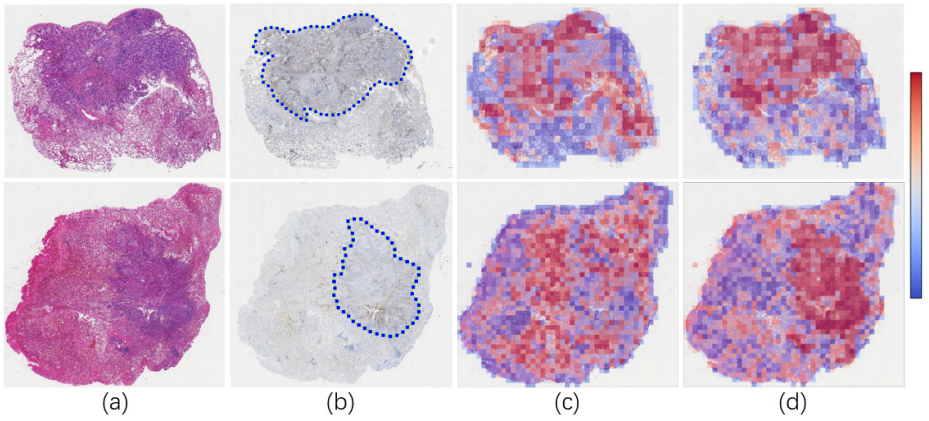


**Fig. 2.** (a) HE-stained images with limited visual information regarding PD-L1 expression. (b) IHC slides with the annotation of cancerous areas. The cells stained in a brownish-yellow color within the annotated regions exhibit PD-L1 positive expression. (c) The attention maps generated by the baseline method indicate dispersed attention. (d) The attention maps of the proposed method show focused attention on annotated regions. The color red indicates areas of high attention, while blue indicates low.

**Table 1.** Ablation study of the proposed method with different losses.

| Methods | Losses | | | PD-L1 dataset | | | | | Camelyon16 [1] dataset | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{L}_{cls}$ | $\mathcal{L}_{cn}$ | $\mathcal{L}_{cl}$ | AUC | ACC | SE | SP | F1 | AUC | ACC | SE | SP | F1 |
| baseline | ✓ | | | 0.668 | 0.621 | **0.949** | 0.252 | 0.727 | 0.894 | 0.881 | 0.727 | 0.975 | 0.820 |
| method-a | ✓ | ✓ | | 0.708 | 0.685 | 0.857 | 0.490 | **0.744** | 0.925 | 0.898 | 0.767 | 0.978 | 0.850 |
| method-b | ✓ | ✓ | ✓ | **0.724** | **0.691** | 0.840 | **0.523** | 0.742 | **0.947** | **0.923** | **0.820** | **0.985** | **0.889** |

**Experimental Setup.** To extract features from image patches, we utilize a pre-trained encoder. The CTransPath [19] is used for feature extraction in the PD-L1 dataset. For the Camelyon16 [1] dataset, we use ResNet50 [6] for feature extraction. For implementation details, the entire framework is implemented by Python 3.10.9 and Pytorch 2.1.0 on NVIDIA TITAN RTX. The network is initialized randomly and directly trained by using an Adam [10] optimizer with a learning rate of 0.0001. An early stopping strategy is employed to prevent overfitting. The training process is stopped if the loss in the validation set does not decrease for 20 consecutive epochs. Due to the varying number of instances obtained in each bag, the minibatch size is set to 1. The hyperparameters $\alpha = 2$, $\beta = 0.1$, $T_p = 0.7$, $T_n = 0.2$.

### 3.2 Ablation Study

To demonstrate the proposed method's effectiveness, we perform the ablation study. Table 1 presents the classification results of different methods on the PD-L1 dataset and the Camelyon16 [1] dataset. The baseline method refers to

**Table 2.** Comparison of the proposed method with other advanced methods on PD-L1 and Camelyon16 [1] dataset.

| Methods | Datasets | | | | | |
|---|---|---|---|---|---|---|
| | PD-L1 dataset | | | Camelyon16 [1] dataset | | |
| | AUC | ACC | F1 | AUC | ACC | F1 |
| MEAN_POOLING | 0.544 | 0.549 | 0.556 | 0.661 | 0.664 | 0.571 |
| MAX_POOLING | 0.536 | 0.552 | 0.695 | 0.858 | 0.823 | 0.794 |
| ABMIL [9] | 0.633 | 0.606 | 0.711 | 0.878 | 0.888 | 0.834 |
| CLAM [12] | 0.668 | 0.621 | 0.727 | 0.894 | 0.881 | 0.820 |
| DSMIL [11] | 0.646 | 0.652 | 0.653 | 0.917 | 0.899 | - |
| TransMIL [18] | 0.591 | 0.567 | 0.695 | 0.931 | 0.884 | - |
| DTFD-MIL [20] | 0.649 | 0.646 | 0.721 | 0.946 | 0.908 | 0.882 |
| Proposed | **0.724** | **0.691** | **0.742** | **0.947** | **0.923** | **0.889** |

the basic MIL approach supervised with cross-entropy loss. In this paper, the CLAM method is employed, it is worth noting that other MIL methods can be freely used. In method-a, a compact negative instance feature mining loss is applied to the negative bags in addition to the baseline method. Method-b includes both the compact negative instance feature mining loss for negative bags and the contrastive learning based instance feature mining loss for typical negative and positive instances within positive bags. The results demonstrate that the proposed method significantly improves the classification capability of the model on both datasets. The proposed method achieved a 5.6% increase in the AUC metric compared to the baseline on the PD-L1 dataset. In the baseline method, the network tends to classify the majority of images as positive, resulting in a high SE score but a deficient SP score. Our method obtains more discriminative instance features to assist the network in classification, improving its classification ability for different classes of images. As a result, we achieve a more balanced SE and SP score, leading to a slight decrease in the SE score while significantly improving the SP score. The attention visualization results of the model on the PD-L1 dataset are shown in Fig. 2. In the IHC image, cells stained brownish-yellow in the cancerous area (marked with blue circles) exhibit positive expression of PD-L1. This staining effectively expresses the presence of PD-L1 information. In the visualization results, we can observe that trained on HE-stained images, the baseline method fails to identify the area of interest, and its attention is scattered. In contrast, our proposed method is more focused and accurate.

### 3.3 Comparison with Other Advanced Methods

The experimental results in Table 2 verify the effectiveness of the proposed method. We compare the proposed method with other advanced methods, including deep models that use traditional MIL pooling operators such as max-pooling and mean-pooling, as well as recent deep MIL models [9, 11, 12, 18, 20], for the task of WSI classification. ABMIL [9] suggests the use of attention for

the aggregation of features. CLAM [12] proposes the addition of an instance prediction branch based on attention aggregation. DSMIL [11] suggests the design of a special MIL after self-supervised training of the feature extraction network. TransMIL [18] employs the Transformer architecture to aggregate instance features. DTFD-MIL [20] proposes a double-tier MIL training mechanism. These methods mainly focus on designing MIL aggregation strategies, overlooking information mining in HE images, which limits their ability to effectively express latent PD-L1-related expression. The proposed method outperforms the other methods in terms of the AUC, ACC, and F1 Score metrics. The AUC metric is independent of the threshold. The higher AUC reflects that the proposed method trains the model with better classification ability. The experimental results demonstrate that the proposed method effectively extracts PD-L1-related information from HE-stained images, resulting in discriminative features and improved classification performance.

## 4 Conclusion

This paper proposes a discriminative multiple instance feature mining method based on novelty detection to extract PD-L1-related information from HE-stained histopathological images. The mechanism that can select typical instances for mining more discriminative instance features. The experimental results confirm that the proposed method achieves a significant classification performance.

**Acknowledgments.** This work was supported by the Fundamental Research Funds for the Central Universities of China under grant DUT23YG225 and grant 226-2024-00185, the funding of Dalian Key Laboratory of Digital Medicine for Critical Diseases, and the Taishan Scholar Program of Shandong Province (tstp20221128).

**Disclosure of Interests.** All authors declare that they have no conflicts of interest.

## References

1. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama **318**(22), 2199–2210 (2017)
2. Borghaei, H., Langer, C.J., Paz-Ares, L., Rodríguez-Abreu, D., Halmos, B., Garassino, M.C., Houghton, B., Kurata, T., Cheng, Y., Lin, J., et al.: Pembrolizumab plus chemotherapy versus chemotherapy alone in patients with advanced non–small cell lung cancer without tumor pd-l1 expression: A pooled analysis of 3 randomized controlled trials. Cancer **126**(22), 4867–4877 (2020)
3. Dehaene, O., Camara, A., Moindrot, O., de Lavergne, A., Courtiol, P.: Self-supervision closes the gap between weak and strong supervision in histology. arXiv preprint arXiv:2012.03583 (2020)
4. Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artificial intelligence **89**(1-2), 31–71 (1997)

5. Gandhi, L., Rodríguez-Abreu, D., Gadgeel, S., Esteban, E., Felip, E., De Angelis, F., Domine, M., Clingan, P., Hochmair, M.J., Powell, S.F., et al.: Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. New England journal of medicine **378**(22), 2078–2092 (2018)
6. 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. pp. 770–778 (2016)
7. Hellmann, M.D., Paz-Ares, L., Bernabe Caro, R., Zurawski, B., Kim, S.W., Carcereny Costa, E., Park, K., Alexandru, A., Lupinacci, L., de la Mora Jimenez, E., et al.: Nivolumab plus ipilimumab in advanced non–small-cell lung cancer. New England Journal of Medicine **381**(21), 2020–2031 (2019)
8. Herbst, R.S., Giaccone, G., de Marinis, F., Reinmuth, N., Vergnenegre, A., Barrios, C.H., Morise, M., Felip, E., Andric, Z., Geater, S., et al.: Atezolizumab for first-line treatment of pd-l1–selected patients with nsclc. New England Journal of Medicine **383**(14), 1328–1339 (2020)
9. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2021)
12. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
13. Lu, M., Chen, R., Wang, J., Dillon, D., Mahmood, F.: Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding. arxiv 2019. arXiv preprint arXiv:1910.10825 (2020)
14. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. Advances in neural information processing systems **10** (1997)
15. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
16. Reck, M., Rodríguez-Abreu, D., Robinson, A.G., Hui, R., Csőszi, T., Fülöp, A., Gottfried, M., Peled, N., Tafreshi, A., Cuffe, S., et al.: Pembrolizumab versus chemotherapy for pd-l1–positive non–small-cell lung cancer. New England Journal of Medicine **375**(19), 1823–1833 (2016)
17. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: International conference on machine learning. pp. 4393–4402. PMLR (2018)
18. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
19. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical image analysis **81**, 102559 (2022)
20. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)