2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313417

Hierarchical Attention-Based Multiple Instance Learning Network for Patient-Level Lung Cancer Diagnosis

Qingfeng Wang[1,#], Ying Zhou[2,#], Jun Huang[1,*], Zhiqin Liu[1], Ling Li[1], Weiyun Xu[2], and Jie-Zhi Cheng[3]

[1]*School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China*

[2]*Radiology Department, Mianyang Central Hospital, Mianyang, China*

[3]*Shanghai United Imaging Intelligence Co. Ltd., Shanghai, China*

* indicates the corresponding author (huangjuncs@swust.edu.cn), and # indicates the equal contributors

***Abstract*—Lung cancer is the leading cause of cancer-related deaths worldwide, while the risk factors for lung cancer mortality can be significantly reduced if the accurate early diagnoses for small malignant lung nodules are possible. In this paper, we propose a hierarchical attention-based multiple instance learning (HA-MIL) framework for patient-level lung cancer diagnosis by introducing two-level cascaded attention mechanisms, one at nodule level and the other at attribute level. The proposed HA-MIL framework is constructed by aggregating important attribute representation into nodule representation and then aggregating important nodule representation into lung cancer representation. The experiments on the public Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset showed that the HA-MIL model performed significantly better than the previous approaches such as the higher-order transfer learning, instance-space MIL and embedding-space MIL, which demonstrated the effectiveness of hierarchical multiple instance learning based on two-level attentions. The results analysis suggested that the HA-MIL model also found the key nodules and attributes by higher attention weights, which were more interpretable for the model decision making.**

***Index Terms*—Multiple instance learning, hierarchical attention mechanism, attribute-level attention, nodule-level attention, patient-level lung cancer diagnosis.**

I. Introduction

Since lung cancer often develops from small malignant lung nodules with diameter less than 30mm, the malignant diagnosis for the nodules has important clinical significance for the diagnosis of early lung cancer [1]. Computed tomography (CT) is widely used in lung cancer screening, which exhibits the lung nodules with strong contrast difference in intensity, texture and shape. Radiologists could subjectively diagnose whether the nodules are benign or malignant by observing the phenotypic characteristics of the nodules in CT images. As one patient can be often found with multiple nodules on the chest CT, radiologists usually diagnose lung cancer as positive if the patient has at least one malignant nodules; otherwise negative if and only if all nodules are benign. Therefore, the diagnosis of lung cancer at the patient level, say patient-level lung cancer diagnosis, can be formulated as a multi-instance learning (MIL) task based on multiple nodules [2].

As shown in Fig. 1 (a), radiologists diagnosed the patient case #187 (selected from the LIDC-IDRI dataset [3]) to be patient-level malignant according to the presence of one malignant nodule. In addition, attributes like malignancy (Mal), texture (Tex), sphericity (Sph), etc. are commonly used to describe phenotypic characteristics of nodules, which are also informative for lung cancer diagnosis [4]. As a nodule has several attributes, the assessment of nodule based on multiple attributes can also be formulated as a multi-instance learning (MIL) task. Thus, the MIL task based on multiple nodules can be extended to multiple attributes, which mirrors a hierarchical structure among patient, nodules and attributes. Accordingly, patient-level lung cancer diagnosis can be further formulated as a hierarchical MIL task based on multiple nodules and attributes.

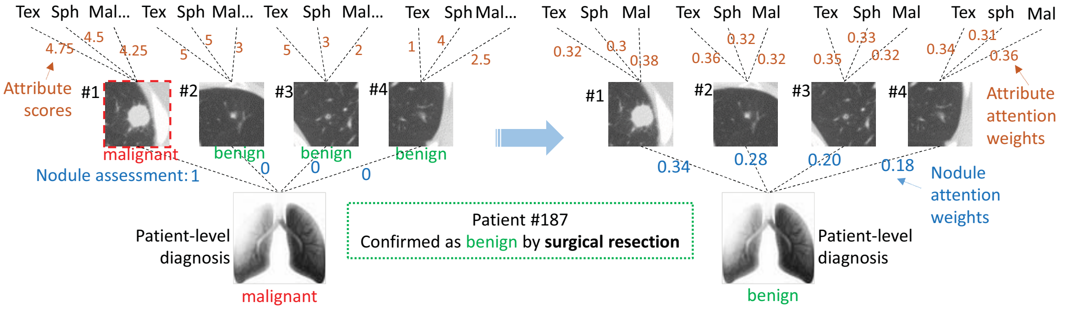
Due to subjective factors, nodules might be misclassified by radiologists, but can be further confirmed by biopsy or surgical resection. As shown in Fig. 1, the patient case #187 was diagnosed as malignant by radiologists while it was confirmed to be benign by surgical resection. The confirmed diagnosis by biopsy or surgical resection is the golden standard for clinical diagnosis of lung cancer. In the scenario of computer-assisted diagnosis for the patient-level lung cancer, it is more preferable to use the confirmed diagnosis as the target label rather than the subjective diagnosis given by radiologists [5]. However, two issues arise when using the confirmed diagnoses as the target labels: 1) the acquisition of the confirmed diagnoses is time-consuming and often invasive, leading to an issue of insufficient target labels; 2) the confirmed diagnosis labels are often directly given at the patient level where the confirmations for the individual nodules are unknown [6], and therefore the instance-space MIL network might be trained insufficiently and introduces additional error to the final bag classification.

To address the issue of insufficient confirmed labels, a MIL deep model transferring from semantic information domain to the confirmed diagnosis domain can be established to assist the diagnosis of patient-level lung cancer. Shen et al. [2] proposed a CNN-based MIL framework with an instance-

This work was supported partly by Doctor's Fund of Southwest University of Science and Technology (Nos. 19zx7143 and 20zx7137), and Sichuan Science and Technology Program, China (No. 2019JDRC0119)

978-1-7281-6215-7/20/$31.00 © 2020 IEEE

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on August 12,2021 at 08:26:22 UTC from IEEE Xplore. Restrictions apply.



(a) Diagnosed as malignant by radiologists **(b) Diagnosed as benign by our HA-MIL model**

Fig. 1. The illustration of patient-level decision making for the lung cancer diagnosis with multiple nodules. Patient case #187 was diagnosed as malignant by radiologists (a) while diagnosed as benign by our model (b), and this case was confirmed to be benign by surgical resection. The attributes texture, sphericity and malignancy are abbreviated as "Tex", "Sph" and "Mal", respectively.

space regression to predict the patient-level lung cancer by transferring from the semantic malignancy features to the confirmed diagnoses. In this study, we try to learn the knowledge extraction of radiologists from various attributes rather than just semantic malignancy.

To alleviate the additional error caused by unknown labels for the individual instances, Wang et al. [7] revised the MIL framework and proposed an embedding-space MIL network by representing each bag as a single feature vector, and trained the MIL network directly with a bag-level classifier instead of the instance-space classifier. Thus, the weak supervision of instance-space MIL has been transformed into a fully-supervised embedding-space MIL [7]. However, as the embedding-space MIL network determines a joint representation of a bag by aggregating all instance embedding features and then gives final decision based on the bag representation, it cannot provide the instance labels for interpretability [7]. To incorporate interpretability to the embedding-space MIL network and increasing its robustness, attention-based operator was introduced to provide insight into the contribution of each instance to the bag label prediction [8].

In this study, we introduce the attention operators into the hierarchical MIL framework and propose a hierarchical attention-based multiple instance learning (HA-MIL) model for the patient-level lung cancer diagnosis by incorporating two-level attention mechanisms, one at the nodule level and the other at the attribute level. The hierarchical attentions allow the HA-MIL model to pay more or less attention to the nodules and the attributes when constructing the representation of the patient-level lung cancer. In this way, the important attribute features can be aggregated into the nodule features and then the important nodule features can be aggregated into the patient-level lung cancer representation to improve the performance of computer-assisted diagnosis for the early lung cancer.

The contribution of this study can be summarized into three folds. First, to alleviate the problem of insufficient labels for confirmed diagnoses, we attempt to transfer from various attribute domains to confirmed diagnoses by learning radiologists' knowledge extraction for nodule assessments. Second, to the best of our knowledges, it is the first time that the HA-MIL model is proposed and the two-level cascaded attention mechanisms force the model to learn harder to seek the informative nodules and attribute features. Third, we quantitatively analyze the contribution of nodule and attribute features, and reveal the underlying relationship between the confirmed diagnosis and its highly-correlated attributes to demonstrate the interpretability of the proposed HA-MIL model.

II. Dataset and Preprocessing

In this study, the LIDC-IDRI dataset [3] was used to demonstrate the effectiveness of the proposed HA-MIL model. The dataset [3] collected thoracic CT scan series from 1010 patients. The lung nodules in the scan slices were reviewed and annotated by four experienced thoracic radiologists. A total of 2632 nodules were extracted, and the data pre-preprocess was the same as in [5], [9], [10]. According to the absence or presence of the confirmed diagnosis of lung cancer, the 2632 nodules were divided into a discovery set and a diagnosis set at the patient level as in [2], [5]. For simplicity, we denote the discover set as D_1 and the diagnosis set as D_2.

The dataset D_1 includes 2283 nodules, and each nodule is scored with 9 semantic attributes, i.e., texture, calcification, sphericity, spiculation, lobulation, subtlety, margin, internal structure and malignancy, which are abbreviated as "Tex", "Cal", "Sph", "Spi", "Lob", "Sub", "Mar", "Int" and "Mal", respectively. All attributes of each nodule were labeled with a range of [1, 5] by radiologists, except for "Cal" with a range of [1, 6]. The attribute scores from all radiologists were averaged as the ground truth for training and test [2], [5], [9], [10].

The dataset D_2 consists of 117 patients (31 benign cases and 86 malignant cases) with 349 nodules. For D_2, not only does each nodule be scored with 9 semantic attributes, but each patient who underwent biopsy or surgery was given a binary confirmed malignancy label. The confirmed malignancy label indicates whether the patient has been definitely diagnosed as lung cancer positive. For clarity, we denoted the binary label of each patient in D_2 as "pathologic malignancy", abbreviated

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on August 12,2021 at 08:26:22 UTC from IEEE Xplore. Restrictions apply.

as “Pmal”. Notably, there was no “Pmal” label for each patient in the discovery set D_1.

III. Methodology

A. Attribute-Specific Models for Semantic Knowledge Extraction

To learn knowledge extraction of radiologists on the assessment of nodules, we construct an attribute-specific model and train it with semantic score labels across various attributes S on the discovery set D_1. The attribute-specific model has a feature extractor and regression network architecture, where the feature extractor is trained to learn semantic knowledge extraction and the regression network is optimized to predict the semantic attribute scores of nodules. For each semantic attribute $s (s \in S)$, we jointly train the feature extractor ϕ_s and the regression network $\mathcal{R}_s$ in an attribute-specific model on D_1, and the loss function between the predictive scores and the radiologists' scores can be formulated as in (1).

$$\mathcal{L}_{reg} = \min_{\mathcal{R}_s, \phi_s} \frac{1}{|D_1|} \sum_{(x, y_s) \in D_1} (\mathcal{R}_s(\phi_s(x)) - y_s)^2 \tag{1}$$

where y_s denotes the score of attribute s for nodule x provided by radiologists. ResNet-18 [11] network pre-trained from ImageNet without the final classification layer was used as the backbone of the feature extractor. The regression network is designed as a sub-network structure that consists of three fully-connected layers, each containing 512 input neurons, 32 hidden neurons and 1 output neuron. The first two fully connected layers are applied with ReLU activation, and the output neuron is operated with mean squared error (MSE) loss. A total of nine attribute-specific models are trained for the nine semantic attributes, respectively.

B. HA-MIL Model for Patient-Level Lung Cancer Diagnosis

The proposed HA-MIL builds on the concepts of bags, instances and attributes. A bag has a number of instances and an instance also has various attributes, which implies a cascaded hierarchical structure with two-level insights into the instances and the attributes. We formulate the patient-level cancer diagnosis as a HA-MIL task, in which one patient can be defined as a bag, a nodule as an instance, and a semantic attribute as an attribute. Given a patient case X with a bag of nodules: $\{x_1, \ldots, x_i, \ldots, x_n\}$ and semantic feature extractors trained in section III-A: $\{\phi_1, \ldots, \phi_j, \ldots, \phi_m\}$, our goal is to classify the pathologic malignancy label $Y \in \{0, 1\}$ with the HA-MIL model. The HA-MIL network mainly includes attribute representation, attribute-level attention, nodule representation, nodule-level attention, and patient-level representation and diagnosis.

Attribute representation. For each nodule x_i in a patient case, the trained semantic feature extractor ϕ_j is used to obtain the representation of the j-th attribute of the nodule. We denote the j-th attribute representation of nodule x_i as u_{ij}:

$$u_{ij} = \phi_j(x_i) \tag{2}$$

Attribute-level attention. The sight of attribute-level attention mechanism is aimed to learn the importance of attribute features to the pathologic malignancy diagnosis by adaptively assigning different weights to the attribute representation. To learn the contribution of each attribute to nodule x_i, we construct an attention subnetwork $Att(.)$ with the attributes' representation as input, and normalize the output of the subnetwork into the range of [0, 1]. The attributes' weights of nodule x_i can be formulated as:

$$\alpha_{ij} = \frac{\exp(Att(u_{ij}))}{\sum_j \exp(Att(u_{ij}))} \tag{3}$$

where the subnetwork $Att(.)$ is constructed with 512 input neurons, 128 hidden neurons, 32 hidden neurons and 1 output neuron, and each fully-connected layer is applied with ReLU activation. The normalized outputs of $Att(.)$ are used as the attention weights of the attributes. Notably, the sum of the attention weights for the attributes of nodule x_i is equal to 1, i.e., $\sum_{j=1}^{m} \alpha_{ij} = 1$.

Nodule representation. We compute the nodule representation vector v_i as a weighted sum of each attribute representation with its corresponding contribution weight.

$$v_i = \sum_{j=1}^{m} \alpha_{ij} u_{ij} \tag{4}$$

Nodule-level attention. The goal of the nodule-level attention is to learn the importance of each nodule for the diagnosis of patient-level malignancy. Likewise, we use the representation of nodules as input to the attention subnetwork $Att(.)$, and the output is also normalized. The contribution of each nodule to the patient-level lung cancer diagnosis can be defined as:

$$\beta_i = \frac{\exp(Att(v_i))}{\sum_i \exp(Att(v_i))} \tag{5}$$

where the sum of the attention weights for the nodules in a bag is also equal to 1, i.e., $\sum_{i=1}^{n} \beta_i$=1.

Patient-level representation and diagnosis. Similarly, we compute the bag representation vector z as a weighted sum of each nodule representation v_i with its corresponding contribution weight β_i.

$$z = \sum_{i=1}^{n} \beta_i v_i \tag{6}$$

Since the pathologic malignancy bag label Y can be positive (1) or negative (0), the patient-level lung cancer diagnosis can be formulated as a binary classification problem. Therefore, the training loss between the prediction result $F(z)$ and the ground-truth malignancy label Y can be defined as:

$$\mathcal{L}_{cls} = -(1 - Y) * \log(1 - F(z)) - Y * \log(F(z)) \tag{7}$$

where the classification subnetwork $F(.)$ is composed of 512 input neurons, 32 hidden neurons and 1 output neuron, and each fully-connected layer is activated by ReLU.

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on August 12,2021 at 08:26:22 UTC from IEEE Xplore. Restrictions apply.

C. Training Scheme

We first train the 9 attribute-specific models by jointly optimizing the feature extractors and regression networks on D_1, and then train and test the HA-MIL model on D_2 with the optimized feature extractors. There exists a combinatorial explosion problem in the selection of 9 attributes to the HA-MIL model, but fortunately, the correlations between each attribute and pathologic malignancy can be highly different. To this end, we first rank the performance of the 9 semantic attributes in the independent classification of pathologic malignancy, and then train the HA-MIL models by selecting the top-k attributes to alleviate the combinatorial explosion problem.

IV. Experiments and Results

A. Experimental Settings

The 5-fold cross validation scheme based on the patient-level data partition was conducted on the training and test for HA-MIL models as in [2]. As all parts of the HA-MIL network are differentiable, we trained the HA-MIL model end-to-end by stochastic gradient descent (SGD) algorithm with a batch size of 1 (=1 bag) [8] and a fixed weight decay of 1e-4. The learning rate was set as 0.001, and the number of the training iteration was set as 2000. To tackle the imbalance between the negative and positive bags, a positive and a negative bag with nodules were alternately used as the input during training. For fair comparison, we reproduced the CNN-MIL [2] and MI-Net [7] methods with the backbone ResNet-18 [11] as the same as in the higher-order transfer model [5]. For the attribute-specific models, SGD algorithm was also adopted for optimization, with a batch size of 64 and a fixed weight decay of 1e-4. We train each attribute-specific model for 60 epochs, with a learning rate starting at 0.01 and a factor of 10 reduction per 20 epochs. All experiments in this study are conducted on a linux server with 4 NVIDIA Titan X GPUs.

B. Performance of Patient-Level Lung Cancer Diagnosis

Table I reports the diagnostic performance of the higher-order transfer [5], CNN-MIL [2], MI-Net [7] and our HA-MIL model on pathologic malignancy (Pmal), with each semantic attribute. Since the four models are evaluated with the 5-fold cross validation scheme, the mean±standard deviation statistics of accuracy, AUC and f1-score are reported. The HA-MIL method performs much better than the the higher-order transfer [5], CNN-MIL [2] and MI-Net [7] on most attributes, which demonstrates that the patient-level fully-supervised MIL model within attention mechanisms can further improve the diagnostic performance on the Pmal. This suggests that the attention-based MIL is more adaptive in feature learning than the instance-space MIL (CNN-MIL) and embedding-space MIL (MI-Net). To illustrate the correlations between pathologic malignancy and various attributes, we rank the 9 attributes according to the f1-score of the HA-MIL model, and list them in the column of "Source→Target" in Table I. The ranking of attributes implies the degree of importance of each attribute in providing semantic information for the Pmal diagnosis, and also alleviates the combinatorial explosion problem in the selection of 9 attributes.

TABLE I
Performance of the pathologic malignancy (Pmal) diagnosis with each semantic attribute, in terms of the models of higher-order transfer, CNN-MIL, MI-Net and the proposed HA-MIL.

Source → Target	Methods	Accuracy	AUC	F1-score
Tex→Pmal	Higher-Order [5]	0.768±0.066	0.729±0.068	0.709±0.066
	CNN-MIL [2]	0.820±0.075	0.768±0.087	0.787±0.080
	MI-Net [7]	0.829±0.028	0.837±0.026	0.791±0.031
	HA-MIL(ours)	**0.846±0.037**	**0.849±0.027**	**0.813±0.037**
Sph→Pmal	Higher-Order [5]	0.739±0.017	0.689±0.058	0.666±0.038
	CNN-MIL [2]	0.838±0.051	0.747±0.121	0.782±0.084
	MI-Net [7]	0.811±0.081	0.777±0.107	0.768±0.076
	HA-MIL(ours)	**0.848±0.059**	**0.831±0.080**	**0.803±0.073**
Mal→Pmal	Higher-Order [5]	0.702±0.050	0.669±0.045	0.644±0.035
	CNN-MIL [2]	0.820±0.044	0.793±0.059	0.780±0.054
	MI-Net [7]	0.820±0.034	0.800±0.043	0.783±0.031
	HA-MIL(ours)	**0.837±0.035**	**0.830±0.021**	**0.800±0.037**
Lob→Pmal	Higher-Order [5]	0.725±0.026	0.721±0.057	0.666±0.034
	CNN-MIL [2]	0.814±0.057	0.751±0.062	0.776±0.049
	MI-Net [7]	0.820±0.034	0.796±0.039	0.786±0.025
	HA-MIL(ours)	**0.822±0.065**	**0.818±0.050**	**0.797±0.058**
Spi→Pmal	Higher-Order [5]	0.742±0.045	0.656±0.043	0.651±0.035
	CNN-MIL [2]	0.820±0.070	0.779±0.083	0.789±0.076
	MI-Net [7]	0.820±0.034	0.796±0.041	0.779±0.046
	HA-MIL(ours)	**0.838±0.061**	**0.811±0.056**	**0.796±0.075**
Mar→Pmal	Higher-Order [5]	0.733±0.066	0.667±0.066	0.659±0.063
	CNN-MIL [2]	0.794±0.076	0.756±0.075	0.744±0.081
	MI-Net [7]	**0.822±0.035**	0.801±0.054	**0.789±0.031**
	HA-MIL(ours)	0.813±0.061	**0.819±0.038**	0.789±0.056
Cal→Pmal	Higher-Order [5]	0.739±0.046	0.689±0.106	0.675±0.056
	CNN-MIL [2]	0.810±0.106	0.758±0.108	0.767±0.116
	MI-Net [7]	**0.811±0.024**	0.756±0.050	0.764±0.013
	HA-MIL(ours)	0.803±0.044	**0.796±0.032**	**0.768±0.038**
Sub→Pmal	Higher-Order [5]	0.716±0.028	0.671±0.067	0.652±0.033
	CNN-MIL [2]	0.795±0.017	0.748±0.103	0.752±0.015
	MI-Net [7]	**0.803±0.020**	0.762±0.061	0.760±0.032
	HA-MIL(ours)	0.802±0.048	**0.775±0.086**	**0.767±0.057**
Int→Pmal	Higher-Order [5]	0.702±0.032	0.646±0.051	0.63±0.022
	CNN-MIL [2]	0.759±0.120	0.692±0.142	0.722±0.128
	MI-Net [7]	0.785±0.089	**0.726±0.085**	0.738±0.073
	HA-MIL(ours)	**0.794±0.076**	0.704±0.079	**0.743±0.064**

TABLE II
Performance of HA-MIL model with the top-k attribute sources.

Top-k sources→Target	Accuracy	AUC	F1-score
Tex→Pmal	0.846±0.037	0.849±0.027	0.813±0.037
Tex+Sph→Pmal	0.863±0.051	0.867±0.070	0.834±0.062
Tex+Sph+Mal→Pmal	**0.880±0.032**	**0.877±0.036**	**0.849±0.037**
Tex+Sph+Mal+Lob→Pmal	0.847±0.056	0.842±0.053	0.817±0.066
Tex+Sph+Mal+. . .+Spi→Pmal	0.846±0.020	0.82±0.0460	0.800±0.031
Tex+Sph+Mal+. . ..+Mar→Pmal	0.837±0.065	0.833±0.065	0.808±0.071
Tex+Sph+Mal+. . .+Cal→Pmal	0.829±0.028	0.821±0.056	0.791±0.036
Tex+Sph+Mal+. . .+Sub→Pmal	0.828±0.041	0.821±0.043	0.798±0.047
Tex+Sph+Mal+. . .+Int→Pmal	0.803±0.020	0.816±0.013	0.770±0.028

Table II reports the diagnostic performance of our HA-MIL model on Pmal with the top-k attribute sources, where $k = \{1, 2, \ldots, 9\}$. The top-k sources, which are selected from the ranking attributes in table I, are considered as the most effective combination among the k attributes. As can be observed in Table II, the performance of the HA-MIL model increases with the increment of the attribute dimension at the beginning, but decreases with the increase attribute dimension after reaching the performance peak at "Tex", "Sph" and "Mal". This may be because the increasing source attributes

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on August 12,2021 at 08:26:22 UTC from IEEE Xplore. Restrictions apply.

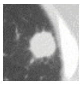
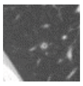
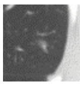
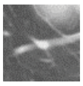
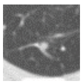
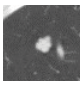
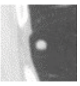
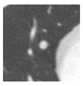
Patient-level prediction	Patient-benign case #187 (organizing pneumonia)												Patient-malignant case #182 (lung cancer)									Patient-malignant case #237 (Leiomyosarcoma)								
Nodule-level attention weights	Nodule #1			Nodule #2			Nodule #3			Nodule #4			Nodule #1			Nodule #2			Nodule #3			Nodule #1			Nodule #2			Nodule #3		
																														
	0.34			0.28			0.20			0.18			0.32			0.33			0.35			0.69			0.21			0.10		
Attribute-level attention weights	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal	Tex	Sph	Mal
	0.32	0.30	0.38	0.36	0.32	0.32	0.35	0.33	0.32	0.34	0.31	0.36	0.31	0.32	0.37	0.32	0.32	0.36	0.31	0.33	0.36	0.43	0.23	0.34	0.40	0.27	0.33	0.44	0.36	0.20
Rating scores	4.75	4.5	4.25	5	5	3	5	3	2	1	4	2.5	5	3.5	2	5	4	2	5	4.25	2.5	5	4	3.75	5	4.75	2.5	5	5	3

Fig. 2. The patient-level benign and patient-level malignant examples with the nodule-level and attribute-level attention weights adaptively assigned by our HA-MIL model, as well as the corresponding rating scores provided by radiologists.

may contain less complementary information for solving the target task. Meanwhile, the attribute source dimension is increased without increasing the nodule instances in the feature vector, the dimensionality of the feature space becomes sparser and sparser which forces the HA-MIL model to be overfitted by loosing generalizing capability.

C. Results Analysis

We show several patient-level benign and malignant examples with the nodule-level and attribute-level attention weights that were adaptively assigned by our HA-MIL model, as shown in Fig. 2. For patient-benign case #187, it can be observed that the radiologists had rated a high level of malignancy for nodule #1 (Mal=4.25), which was considered highly likely to be malignant. In fact, the case #187 was confirmed to be benign by surgical resection and turned out to be organizing pneumonia. Nevertheless, our HA-MIL model accurately diagnosed the case#187 to be benign, and assigned a relatively large attention weight to nodule #1.

The patient-malignant case #182 was confirmed as primary lung cancer, while it was diagnosed as relatively benign (Mal$\leq$2.5) by radiologists. Our model accurately diagnosed the case #182 to be malignant, with relatively uniform distribution of attention weights for both levels of nodules and attributes. The patient case #237 was diagnosed as malignant (Mal=3.75) by radiologists and turned out to be metastatic lung cancer. The nodule #1 in case #237 has a popcorn-like appearance and was distinctly malignant, to which our model adaptively assigned a larger attention (0.69). The results analysis demonstrated that the proposed HA-MIL model detected the key nodules and attributes that were more interpretable for the model decision.

V. Conclusion

In this study, we proposed a hierarchical attention-based multiple instance learning (HA-MIL) framework that incorporates the two-level cascaded attention sights into the nodules and attributes for improving the diagnostic performance of early lung cancer. The two-level cascaded attentions also quantitatively reflected the importance of the nodules and the attributes in the lung cancer diagnosis. Experimental results demonstrated that the HA-MIL model performs significantly better than the previous studies and our model could provide an interpretation for the decision-making by presenting the nodule-level and attribute-level attention weights, which are extremely important in practical medical applications.

References

[1] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, and W. Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest ct," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 991–1004, April 2019.

[2] W. Shen, M. Zhou, F. Yang, D. Dong, C. Yang, Y. Zang, and J. Tian, "Learning from experts: Developing transferable deep features for patient-level lung cancer prediction," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Cham: Springer International Publishing, 2016, pp. 124–131.

[3] S. Armato III, G. Mclennan, L. Bidaut, M. McNitt-Gray, C. R. Meyer, and A. P. Reeves, "The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans," *Medical Physics*, vol. 38, pp. 915–931, 2011.

[4] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, and J. Cheng, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in ct images," *IEEE Transactions on Medical Imaging*, vol. 36, no. 3, pp. 802–814, 2017.

[5] Q. Wang, J. Huang, Z. Liu, J. Cheng, Y. Zhou, Q. Liu, Y. Wang, X. Zhou, and C. Wang, "Higher-order transfer learning for pulmonary nodule attribute prediction in chest ct images," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019, pp. 741–745.

[6] B. Vendt and J. Kirby. (2020) IEEEtran homepage. [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI

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

[8] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018, pp. 2127–2136.

[9] Q. Wang, X. Zhou, C. Wang, Z. Liu, J. Huang, Y. Zhou, C. Li, H. Zhuang, and J. Cheng, "Wgan-based synthetic minority oversampling technique: Improving semantic fine-grained classification for lung nodules in ct images," *IEEE Access*, vol. 7, pp. 18 450–18 463, 2019.

[10] Q. Wang, J. Cheng, Z. Liu, J. Huang, Q. Liu, Y. Zhou, W. Xu, C. Wang, and X. Zhou, "Multi-order transfer learning for pathologic diagnosis of pulmonary nodule malignancy," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec 2018, pp. 2813–2815.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on August 12,2021 at 08:26:22 UTC from IEEE Xplore. Restrictions apply.