MTMR-Net: Multi-task Deep Learning with Margin Ranking Loss for Lung Nodule Analysis

Lihao Liu[1(✉)], Qi Dou[1], Hao Chen[1,3], Iyiola E. Olatunji[2], Jing Qin[4], and Pheng-Ann Heng[1]

[1] Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Sha Tin, Hong Kong
`lhliu@cse.cuhk.edu.hk`
[2] Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong, Sha Tin, Hong Kong
[3] Imsight Medical Technology Co., Ltd., Shenzhen, China
[4] Center for Smart Health, School of Nursing,
The Hong Kong Polytechnic University, Kowloon, Hong Kong

Abstract. Lung cancer is the leading cause of cancer deaths worldwide. Early diagnosis of lung nodules is of great importance for therapeutic treatment and saving lives. Automated lung nodule analysis requires both accurate lung nodule benign-malignant classification and attribute score grading. However, this is quite challenging due to the considerable difficulty of nodule heterogeneity modelling and limited discrimination capability on ambiguous cases. To meet these challenges, we propose a **M**ulti-**T**ask deep learning framework with a novel **M**argin **R**anking loss (referred as MTMR-Net) for automated lung nodule analysis. The relatedness between lung nodule classification and attribute score regression is explicitly explored in our multi-task model, which can contribute to the performance gains of both tasks. The results of different tasks can be yielded simultaneously for assisting the radiologists in diagnosis interpretation. Furthermore, a siamese network with a novel margin ranking loss was elaborately designed to enhance the discrimination capability on ambiguous nodule cases. We validated the efficacy of our MTMR-Net on the public benchmark LIDC-IDRI dataset. Extensive experiments demonstrated that our approach achieved competitive classification performance and more accurate attribute scoring over the state-of-the-arts.

1 Introduction

Lung cancer has been the leading cause of cancer deaths worldwide. In the year 2018, the estimated death cases of lung cancer will account for approximately 26% of all cancer deaths in the United States [1]. Early diagnosis of lung cancer is crucial in the future treatment of lung cancer patient, because its five-year survival rate is lower than 20% when it promotes to a late stage. Lung cancer usually refers to small malignant lung nodules (with the diameter in the

range of 3–30 mm), which can be detected on the chest computed tomography (CT) scans. However, distinguishing the nodules between benign and malignant is quite difficult even for experienced radiologists [2]. Because there are various potential malignancy-related characteristics (e.g., spiculation), these characteristics should be taken into consideration during the diagnosis process.

Computer-aided diagnosis techniques have been proven to be helpful for radiologists in decision making and hold the potential to improve diagnostic accuracy in distinguishing small benign nodules from malignant ones [3]. With the powerful representation capability, deep neural networks are capable of learning more complicated diagnosis patterns from labeled data. Hence, it could assist the automated lung nodule analysis. Recently, several deep learning based methods have been proposed for computer-aided diagnosis of lung nodules. Xie et al. [6] proposed a multi-model ensemble method that considered overall appearance, nodule shape and voxel value of each nodule slice simultaneously to achieve high classification accuracy. Chen et al. [5] introduced a multi-task regression model to explore the internal relationship among the semantic features. Instead of considering these two tasks independently, Hussein et al. [13] proposed a 3D CNN-based multi-task model to implicitly explore the relationship between malignancy classification and attribute score regression tasks. Although achieving state-of-the-art performance, these previous methods either independently or "jointly but implicitly" tackled the benign-malignant classification and attribute score regression tasks, instead of jointly analyzing and explicitly exploring their correlations for more convincing and interpretable diagnosis.

In this paper, we propose a novel **M**ulti-**T**ask deep learning framework with a new **M**argin **R**anking loss (called MTMR-Net) for automated lung nodule analysis. We build a bi-branch model which not only predicts nodule malignancy but also outputs regressed scores of eight attribute characteristics. The relatedness between two highly-correlated tasks is explicitly learned in our model, and both tasks can benefit from each other through the proposed architecture. Furthermore, we propose a novel margin ranking loss based on siamese network architecture to perform comparison while scoring nodules to model their heterogeneity. This enables the network to be more accurate on recognizing marginal lung nodules by referring to lung nodules with different labels but close malignancy scores. We validated our proposed framework on the public LIDC-IDRI dataset and achieved competitive classification accuracy over the state-of-the-arts. In addition, compared with previous approaches which can only output a binary classification result, our proposed model can provide more cues and evidence for radiologists by simultaneously yielding the scores of the attributes when making diagnosis.

2   Method

Our proposed MTMR-Net consists of two components. First, we propose a multi-task deep learning model for nodule analysis, which is composed of lung nodule

classification task and attribute score regression task. Second, to further discriminate the marginal nodules, we present a new margin ranking loss to train the model in order to enhance the distinguishing capability among marginal cases.
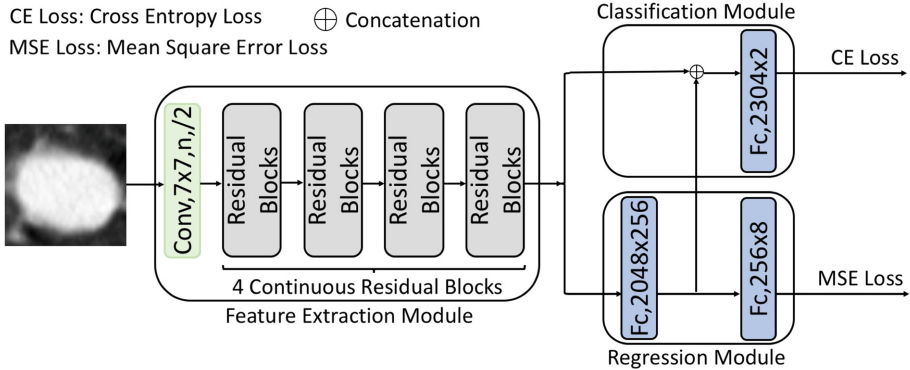

Fig. 1. Multi-task learning framework. Residual blocks used are exactly the same as the residual blocks in original 50-layer residual network [7]. Besides classification branch, an additional regression branch is added to predict 8 attributes scores. The "CE Loss" and "MSE Loss" denote cross entropy loss and mean square error loss, respectively.

2.1   Multi-task Learning for Lung Nodule Analysis

Benign-Malignant Classification. The multi-task model is fine-tuned from a 50-layer residual network [7]. We keep the feature extraction module of the original residual network. However, in the classification module, we concatenated the extracted feature maps with an additional feature map (feature map from regression module) before the last fully-connected layer, as shown in Fig. 1. We formulate the task as a classification problem rather than a regression problem, considering that a definite diagnosis can provide more intuitive information to experts. Therefore, we use cross entropy loss (CE Loss) for backward propagation in the classification module, which is defined as:

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_i log \ p_i^c \left(y_i^c | x_i; W_{cls}, W_s \right), \tag{1}$$

where x_i and p_i^c are the input image and output probability from the classification module, while $y_i^c \in \{0, 1\}$ is the ground truth of lung nodule classification label, W_s and W_{cls} are the weights of shared feature extraction path and nodule classification task, respectively. N is the total number of training samples.

Nodule Attribute Score Regression. Motivated by the clinical observation that radiologists analyze the characteristics of attributes for malignancy assessment, we hypothesize that exploring the correlation between malignancy

classification and attributes scoring would help to further improve the discrimination capability for lung nodule analysis. Therefore, besides the classification task, we also add a regression module for attributes score prediction in the network. Before the last fully-connected layer for final regression, we explicitly extract attributes features using another fully-connected layer following the shared feature extraction module, as shown in Fig. 1. In addition, rather than using these attributes features solely for regression task, we concatenate the malignant feature in the classification module with the attributes features. The concatenation between malignancy feature map and attributes feature map enables more attributes information guidance in the nodule classification task. For the attributes score regression task, we used mean square error loss (MSE Loss) during the training process, which is defined as:

$$\mathcal{L}_{reg} = \frac{1}{N} \sum_i ||\hat{y}_i^r(x_i; W_s, W_{reg}) - y_i^r||_2^2, \tag{2}$$

where $y_i^r \in \mathbb{R}^{1 \times n}$ is the output of regression task of network, while $\hat{y}_i^r \in \mathbb{R}^{1 \times n}$ is the ground truth of attribute scores. $n = 8$, for using eight semantic attributes.
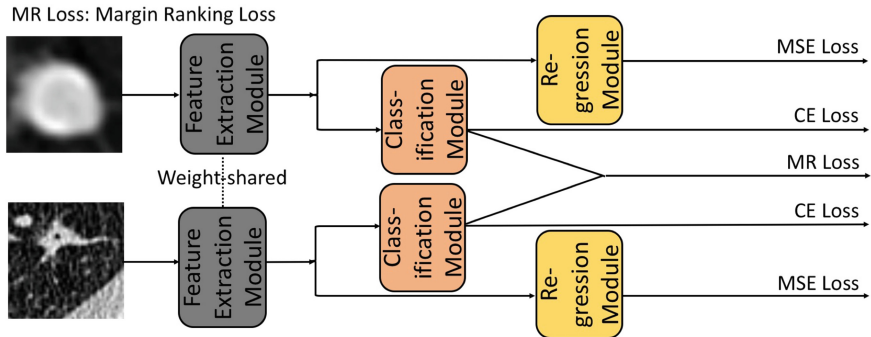

Fig. 2. Siamese model based on two shared-weight proposed multi-task model. "MR Loss" means margin ranking loss. All 3 modules (feature extraction, classification, regression) are weight-shared in two branches of siamese network.

2.2 Margin Ranking Loss for Discriminating Marginal Nodules

Despite multiple correlated supervision information is employed in our deep neural network, we still observe there exists misclassification on marginal lung nodules. To tackle the similar misclassification problem, Kong et al. [8] used siamese network to enhance model's discrimination capability on ambiguous cases. Inspired by Kong et al. [8], we perform the same architecture with a novel margin ranking loss while scoring nodules to model nodules' heterogeneity. Siamese network is well-known for using two shared-weight feature extraction

branches in its network architecture. It enables the network to train in a pairwise mode, see Fig. 2, which can enhance classification accuracy by applying comparison and referring. Besides, a novel margin ranking loss is designed for capturing the ranking relationship between different training samples:

$$\mathcal{L}_{rank} = \frac{1}{2N} \sum_{i,j} max\left(0, \gamma - \delta\left(p_i^c, p_j^c\right) * \left(t_i^c - t_j^c\right)\right), \tag{3}$$

$$\delta\left(p_i^c, p_j^c\right) = \begin{cases} 1, & p_i^c \geq p_j^c \\ -1, & p_i^c < p_j^c \end{cases}, \tag{4}$$

where $t_i^c \in [0,1], t_j^c \in [0,1]$ denotes the ground truth malignancy score for the ith, jth training sample, respectively. While $p_i^c \in [0,1], p_j^c \in [0,1]$ are the ith, jth training sample's predicted malignancy probability, respectively. $\delta\left(p_i^c, p_j^c\right)$ is the indicator function. γ is the margin parameter.

If the predicted scores' ranking is the same as ground truth scores' ranking (e.g., $t_i^c \geq t_j^c, p_i^c \geq p_j^c$), then the loss is 0. Otherwise, the loss is penalized during the training process (e.g., $t_i^c \geq t_j^c, p_i^c < p_j^c$). Applying this mechanism into a siamese network can easily explore and model the difference between marginal lung nodules by adjusting the margin parameter γ.

2.3   Joint Training of MTMR-Net

In summary, there are three not independent but rather complementary losses for our proposed MTMR-Net. Hence, the total minimization loss is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda\mathcal{L}_{reg} + \beta\mathcal{L}_{rank} + \eta(||W_s||_2^2 + ||W_{cls}||_2^2 + ||W_{reg}||_2^2), \tag{5}$$

where λ, β, η are hyper-parameters balancing $\mathcal{L}_{cls}$, $\mathcal{L}_{reg}$ and weight decay term.

In our experiments, Adam optimizer was used for training the entire network. Learning rate was initially set to 3e−3 for the shared feature extraction part and 3e−5 for both classification and regression module. Learning rate also periodically annealed by 0.1. We trained our model for 150 epochs using the pytorch. After using grid-search for finding hyper-parameters, we set 3 parameters for controlling the weights for λ, β, η as 1, 5e−1, 1e−3, respectively, and the marginal parameter γ was chosen as 1e−1.

3   Experiments

3.1   Dataset and Preprocessing

We validated the proposed MTMR-Net on the LIDC-IDRI dataset, which consisted of 1018 CT scans [9] and 1422 lung nodules (972 benign lung nodules and 450 malignant lung nodules). The nodules were rated from 1 to 5 by four experienced radiologists signifying the degree of malignancy in an increasing order. For benign-malignant classification task, nodules with average score less than

3 and greater than 3 were labeled as benign and malignant, respectively. Nodules with average score of 3 were left out in our experiments as all other works did [4–6]. Besides malignancy, eight semantic attributes (i.e., subtlety, calcification, sphericity, margin, spiculation, texture, lobulation and internal structure) were also scored in the LIDC-IDRI dataset. The higher the score is, the more obvious the characteristic is. Most features were rated in the range of 1–5, while the internal structure and calcification were given scores in the range of 1–4 and 1–6, respectively. We rescaled the average score labels from 1–5, 1–6, 1–4 to 0–1 for normalization before training.

We divided the dataset into training (90%) and testing (10%) sets following the setting in [4], which is well calculated so the sampled training and testing dataset has similar distribution. We cropped an adaptive patch region according to the diameter and position of the nodule and resized the patch to 224×224 using bilinear interpolation. In addition, we employed random cropping, horizontal flipping, and vertical flipping as data augmentations. In [12], Dou et al. employed 3D CNN to preserve more spatial information. Instead, we use 2D CNN to explore each slice's malignancy and semantic attribute score, and then averaged the probability scores of slices enclosing nodule to get the final results as mentioned in [6]. This method may lose some spatial information, but the average operation can effectively prevent overfitting.
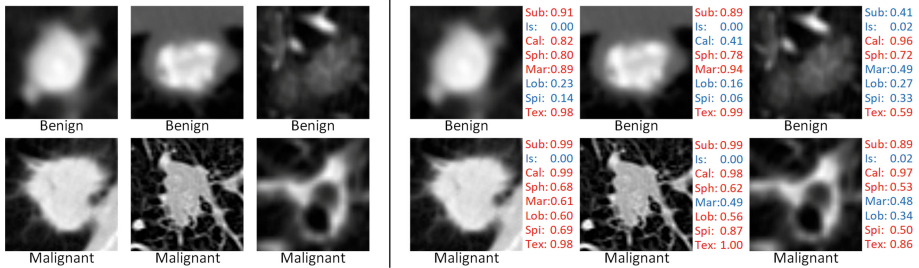

Fig. 3. Left part: classification outputs from previous work's model [4,6]. Right part: classification outputs with attribute score from MTMR-Net. Sub, Is, Cal, Sph, Mar, Lob, Spi, Tex denotes subtlety, internal structure, calcification, sphericity, margin, spiculation, lobulation and texture, respectively. Score for each attribute is rescaled to the range of 0–1. The higher the score is, the more obvious the characteristic is.

3.2 Results and Evaluation Comparison

Benign-Malignant Classification. We compared the proposed model with several state-of-the-art methods and performed an ablation analysis of the proposed model. The results are reported in Table 1. We employed four commonly used metrics for the comparison: accuracy, specificity, sensitivity and area under curve (AUC); the definitions of these metrics can be found in [6]. As shown

Table 1. Performance of lung nodule classification methods on LIDC-IDRI dataset

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Anand et al. 2015 [10]	86.3	89.6	86.7	–
Xie et al. 2016 [6]	93.4	**91.4**	94.1	**0.978**
Shen et al. 2017 [11]	87.1	77.0	93.0	0.930
Causey et al. 2017 [4]	93.2	87.9	**98.5**	0.971
50-layer Residual Net	90.1	83.1	97.0	0.950
MTMR-Net (w/o $\mathcal{L}_{reg}$)	91.6	84.6	**98.5**	0.957
MTMR-Net (w/o $\mathcal{L}_{rank}$)	92.3	86.2	97.0	0.946
MTMR-Net	**93.9**	89.2	**98.5**	0.957

in Table 1, our method achieved the best accuracy, sensitivity and comparable specificity, AUC when compared with state-of-the-art methods, demonstrating the effectiveness of exploiting the relatedness of classification task and attribute prediction task as well as the margin ranking loss in improving the classification accuracy. In order to carefully scrutinize the contributions of different components of the proposed model, we further compared the proposed original the 50-layer Residual Net, the MTMR-Net without MSE Loss, and the MTMR-Net without MR Loss. It is observed that both the MTMR-Net without MSE Loss and the MTMR-Net without MR Loss achieve better performance than the 50-layer Residual Net while the proposed model not only further improved the performance but also outperformed the 50-layer Residual Net by a great margin, further corroborating the effectiveness of the proposed multi-task learning scheme as well as the margin ranking loss.

Nodule Attribute Score Regression. We further compared the results of attribute score prediction of our model with two commonly used models, lasso regression model and elastic network, as well as a state-of-the-art method, MTR [5]. The results are shown in Table 2. We employed the metric of absolute distance error to evaluate the prediction results and its definition can be found in [5]. Compared with previous methods, our model achieved significantly lower absolute distance error on most of the features, demonstrating in our multi-task model trained based on the relatedness between these two tasks, while the attribute prediction task can improve the performance of the classification task, in turn, the classification task can also enhance the attribute prediction accuracy.

Figure 3 showed typical results of classification and the corresponding attribute prediction results. Inspiringly, we found our results are quite consistent with those of previous clinical studies. For example, the malignant cases usually have higher calcification, higher lobulation and lower spiculation while internal structure has no influence on malignancy diagnosis. The results also demonstrate that we cannot classify the nodules based solely on one or two attributes. However, we should comprehensively consider more attributes, which has also been stated in many clinical studies. Compared with previous methods without explicitly exploring the relatedness of two tasks, the proposed model can also provide more cues and evidence for diagnosis by simultaneously outputting the

Table 2. Performance of attribute scores prediction. MTR, LASSO, EN are multi-task regression model [5], lasso regression model and elastic network, respectively. Sub, Is, Cal, Sph, Mar, Lob, Spi, Tex shares the same definition as in Fig. 3. The score is calculated on the original unscaled data.

Methods	Features							
	Sub	Is	Cal	Sph	Mar	Lob	Spi	Tex
MTR [5]	0.75	0.04	**0.48**	0.81	0.86	0.87	0.80	0.58
LASSO	1.25	**0.02**	2.18	1.25	1.13	0.95	0.89	1.04
EN	1.20	0.14	1.44	1.09	0.98	0.96	0.86	1.24
MTMR-Net	**0.52**	0.03	0.62	**0.58**	**0.53**	**0.51**	**0.49**	**0.38**

attribute scores, besides better classification accuracy. The proposed method not only can be used in automated lung nodule diagnosis systems, but also it can be employed as a tool for the investigations which aim at revealing the underlying yet complicated relationship between the malignancy of a nodule and its attributes as shown in Fig. 3.

4   Conclusion

In this paper, we presented the MTMR-Net under a multi-task deep learning framework with margin ranking loss for automated lung nodule analysis. The relatedness between lung nodule classification and attribute score regression was explicitly explored with multi-task deep learning, which contributed to the performance gains of both tasks. Furthermore, a novel margin ranking loss was explored to model nodule heterogeneity and encourage the discrimination capability of ambiguous nodule cases. Extensive experiments on the benchmark dataset verified the efficacy of our method and achieved competitive performance over the state-of-the-arts.

References

1. Siegel, R.L., Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2018. CA Cancer J. Clin. **67**(1), 7–30 (2018)
2. del Ciello, A., Franchi, P., Contegiacomo, A., Cicchetti, G., Bonomo, L., Larici, A.R.: Missed lung cancer: when, where, and why? Diagn. Interv. Radiol. **23**(2), 118–126 (2017)
3. Kumar, D., Wong, A., Clausi, D.A.: Lung nodule classification using deep features in CT images. In: Computer and Robot Vision (CRV), pp. 133–138 (2015)
4. Causey, J., et al.: Highly accurate model for prediction of lung nodule malignancy with ct scans. Scientific Reports 8(1), 9286 (2018)

5. Chen, S., Qin, J., Ji, X., Lei, B., Wang, T., Ni, D., Cheng, J.Z.: Automatic scoring of multiple semantic attributes with multi-task feature leverage: a study on pulmonary nodules in ct images. IEEE Trans. Med. Imaging **36**(3), 802–814 (2017)
6. Xie, Y., Xia, Y., Zhang, J., Feng, D.D., Fulham, M., Cai, W.: Transferable multi-model ensemble for benign-malignant lung nodule classification on chest CT. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 656–664. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_75
7. 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)
8. Kong, S., Shen, X., Lin, Z., Mech, R., Fowlkes, C.: Photo aesthetics ranking network with attributes and content adaptation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 662–679. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_40
9. Armato III, S.G., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phys. **38**(2), 915–931 (2011)
10. Anand, S.V.: Segmentation coupled textural feature classification for lung tumor prediction. In: IEEE International Conference on Communication Control and Computing Technologies (ICCCCT). pp. 518–524 (2010)
11. Shen, W., et al.: Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recogn. **61**, 663–673 (2017)
12. Dou, Q.: 3d deeply supervised network for automated segmentation of volumetric medical images. Med. Image Anal. **41**, 40–54 (2017)
13. Hussein, S., Cao, K., Song, Q., Bagci, U.: Risk stratification of lung nodules using 3D CNN-based multi-task learning. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 249–260. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_20