

多分支卷积神经网络肺结节分类方法及其可解释性



张佳嘉 张小洪

重庆大学大数据与软件学院 重庆 400000

(gagazhang@cqu.edu.cn)

摘要 肺结节 CT 图像表征复杂且多样,导致对肺结节进行分类较为困难。虽然越来越多的深度学习模型被应用到计算机辅助肺癌诊断系统的肺结节分类任务中,但这些模型的“黑盒”特性无法解释模型从数据中学习到哪些知识,以及这些知识是如何影响决策的,导致诊断结果缺乏可信性。为此,文中提出了一种可解释的多分支卷积神经网络模型来判别肺结节的良恶性。该模型利用医生诊断时所用的肺结节语义特征信息来辅助诊断肺结节的良恶性,并将这些特征与肺结节良恶性判别网络融合成多分支网络,在完成肺结节良恶性诊断任务的同时,得到肺结节相关语义特征的预测结果,为医生提供可信的诊断依据。在 LIDC-IDRI 数据集上的实验结果表明,与现有方法相比,所提模型不仅可以得到可解释的诊断结果,而且实现了更好的肺结节良恶性分类效果,其准确率可达 97.8%。

关键词: 计算机辅助诊断;卷积神经网络;多分支;可解释性;肺结节恶性程度分类

中图法分类号 TP183

Multi-branch Convolutional Neural Network for Lung Nodule Classification and Its Interpretability

ZHANG Jia-jia and ZHANG Xiao-hong

School of Big Data & Software Engineering, Chongqing University, Chongqing 400000, China

Abstract The characteristics of lung nodules are complex and diverse, which make it difficult to classify lung nodules. Although more and more deep learning models are applied to the lung nodule classification task of computer-aided lung cancer diagnosis systems, the “black box” characteristics of these models cannot explain what knowledge the model has learned from the data and how the knowledge influences the decision, leading to a lack of reliability in the diagnosis results. To this end, an interpretable multi-branch convolutional neural network model is proposed to identify the benign and malignant lung nodules. The model uses the semantic features of the pulmonary nodules which radiologists use in diagnosis to assist identifying the benign and malignant lung nodules. These characteristics are combined with the branch of malignancy classification into a multi-branch network. Then beyond the malignancy classification, the model can predict nodule attributes, which could potentially explain the diagnosis result. Experimental results on the LIDC-IDRI dataset show that, compared with the existing methods, the proposed model can not only obtain interpretable diagnostic results, but also achieve better classification of lung nodules with an accuracy rate of 97.8%.

Keywords Computer-aided diagnosis, Convolutional neural network, Multi-branch, Interpretable, Classification of malignant degree of pulmonary nodules

1 引言

肺癌是全球死亡率最高的癌症^[1-2]。肺癌的早期症状在医学影像上大多表现为肺结节,而且肺结节有很大概率转换为恶性肺结节。由于恶性肺结节的早期症状不明显,大多数患者在发现时已是晚期,错过了最佳治疗时机。因此,对肺部疾病患者进行及早的肺结节检测和治疗是降低肺癌死亡率的关键措施。美国的全国肺癌筛查试验(NLST)^[3]显示,相对

于使用普通胸片进行筛查,使用计算机断层扫描技术(Computer Tomography, CT)进行筛查将高风险肺癌受试者的死亡率降低了 20%。因此,最佳的肺癌诊断方法是低剂量 CT。医生面对越来越多的 CT 影像诊断任务,难免面临以下问题: 1) 肺结节生长环境复杂多变、形态各异,给医生肉眼诊断带来了极大困难; 2) 肉眼阅片的工作繁重,容易造成漏诊和误诊; 3) 肉眼阅片存在一定的主观性和不稳定性。

目前,已经有一些用来协助放射科医师诊断肺结节良恶

到稿日期:2019-07-29 返修日期:2020-03-12 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61772093);重庆市重大主题专项项目(cstc2018jszx-cyztzxX0017)

This work was supported by the National Natural Science Foundation of China (61772093) and Chongqing Major Theme Projects (cstc2018jszx-cyztzxX0017).

通信作者:张小洪(xhongz@cqu.edu.cn)

性的计算机辅助诊断 (Computer-Aided Diagnosis, CAD) 系统^[4-9], 这些系统提高了医生的工作效率和诊断准确率。一般来说, 传统的计算机辅助诊断系统的诊断过程包括以下步骤: 分割肺结节, 从候选结节中提取图像底层特征 (如纹理、形态、灰度、形状等), 然后把把这些特征输入传统分类器 (如 SVM、K 近邻、神经网络) 中进行学习, 最后得出候选结节良恶性的分类结果。这样的肺结节分类方法存在以下问题: 1) 将提取图像特征作为肺结节分割的后继步骤, 导致提取的特征很容易受分割结果好坏的影响^[10-12]; 2) 底层特征的提取是通过计算机对图像像素之间的灰度关系进行计算得到的, 底层特征虽然可以被计算机理解, 但是与医生描述肺结节的语言有很大差异。该方法很难得到医生易理解的分类型结果, 也无法为医生的临床诊断提供合理性依据。

同时, 深度学习方法, 特别是卷积神经网络, 也已经被应用到肺结节分类任务中, 并取得了良好的效果。Dou 等^[13]提出一种融合多尺度肺结节上下文语义特征的 3D 卷积神经网络, 用于提取肺结节的细节特征和全局特征进行学习, 完成肺结节的假阳性去除任务。Shen 等^[11]也利用 Multi-crop 卷积神经网络学习不同尺度的肺结节特征, 并对肺结节进行恶性程度分类。这些方法只需要整张 CT 图像作为输入, 然后使用卷积神经网络提取深度特征来进行肺结节分类, 虽然解决了传统分类方法提取的特征容易受分割结果影响的问题, 但是也存在以下问题。首先, 这些深度学习模型的“黑盒性”无法解释模型到底学到了哪些图像特征以及模型是怎么根据学习到的特征给出预测结果的; 其次, 模型可解释性的缺乏使得放射科医师无法了解模型的工作机制, 从而阻碍了模型的临床推广使用。

据研究, 肺结节的语义特征主要有: 分叶征 (肺结节的边缘轮廓不是纯粹的圆形, 连接部分较为凹陷), 毛刺征 (肺结节边缘呈现尖刺状或放射状), 对比度 (肺结节与周围环境的对比度), 球形度, 边缘, 纹理等。放射科医师一般根据这些语义特征的具体情况来进行良恶性诊断, 如恶性结节常表现为边缘不规则或伴有毛刺、针尖状或偏心性钙化; 良性结节主要表现为边缘光滑, 无明显分叶毛刺等。图 1 给出了一些不同语义特征的结节示例 (带边框的图片表示恶性结节, 每一列从上到下语义特征明显度增加)。可以较明显地看出, 肺结节毛刺征程度、分叶征程度、对比度等特征明显度越高, 该结节是恶性结节的可能性就越大^[14]。因此, 肺结节语义特征不仅对评估肺结节良恶性有帮助^[15-16], 而且这些特征对于放射科医师来说是直观的, 有助于医生理解诊断结果。

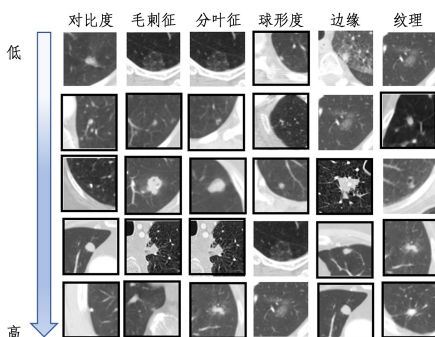


图 1 良恶性结节的语义特征

Fig. 1 Semantic characteristics of malignant and benign nodules

基于此, 本文将肺结节语义特征结合到卷积神经网络的设计中, 提出了一种新颖的可解释的多分支特征卷积神经网络模型 (Interpretable Multi-branch Attributes Convolutional Neural Network, IMACNN) 来评估肺结节的恶性程度。本文的主要贡献如下:

(1) 采用放射科医师诊断肺结节良恶性所依赖的语义特征信息来辅助判断肺结节的良恶性, 形成多分支卷积神经网络, 实验证明肺结节良恶性分类性能有所提高;

(2) 构建了具有可解释性的卷积神经网络模型, 该模型不仅可以得到肺结节良恶性诊断结果, 还可以得到相关的语义特征预测结果, 为良恶性诊断结果提供可解释性依据。

2 数据集分析与处理

2.1 数据集

美国肺部图像数据联盟 LIDC-IDRI^[17] 是全世界最大的公开肺部影像数据集, 用于高危人群早期肺癌的检测诊断。该数据集包含 1018 名患者的肺部 CT 扫描数据, 每个患者的数据包括该患者肺部所有 CT 扫描图像序列, 以及 4 位胸部放射科医师的诊断结果 XML 文件。医师针对每个结节给出 9 个医学语义特征 (对比度、分叶征、毛刺征、球形度、边缘、纹理、内部结构、钙化程度和恶性程度) 的具体分级 (1-6 级), 等级越高, 语义特征越明显。

2.2 肺结节语义特征选择

肺结节语义特征可以促进肺结节恶性程度分类的准确率, 然而, 并不是所有的特征都对肺结节的恶性程度有促进作用。因此, 本文使用 LIDC-IDRI 数据集计算各肺结节语义特征不同程度下恶性结节所占比例的方差, 分析肺结节语义特征对恶性程度的敏感性。

$$S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (1)$$

其中, n 为肺结节各语义特征等级的等级, X_i 为恶性结节的比例, \bar{X} 为平均比例。敏感性较高的肺结节语义特征代表着“当该特征出现时更容易增强或者减弱分类模型对肺结节类别的判断”。将计算出的敏感性进行排序, 排序结果如表 1 所列。可以发现, 前两个敏感属性是“内部结构”和“钙化程度”, 但是这两个属性的样本分布非常不平衡, 导致在单独训练这两个特征网络分支时产生不好的预测结果, 从而破坏肺结节诊断系统的整体性能。因此, 我们除去这两个肺结节语义特征, 选择敏感性靠前的 3 个肺结节特征 (分叶征、毛刺征和对比度) 作为肺结节分类的辅助特征信息。

表 1 各属性下各级恶性结节的比例

Table 1 Proportion of malignant nodules in each level

属性	恶性结节比例/%						敏感度	排序结果
	1	2	3	4	5	6		
内部结构	18.90	47.62	50.00	0.00	0.00	—	0.060	1
钙化程度	0.00	0.00	0.00	1.80	10.81	21.88	0.057	2
分叶征	6.77	26.67	54.55	56.12	13.79	—	0.052	3
毛刺征	7.47	31.39	58.70	57.39	41.94	—	0.045	4
对比度	2.56	5.92	5.53	17.23	44.61	—	0.030	5
球形度	0.00	17.28	23.00	20.79	8.60	—	0.009	6
边缘	20.56	23.83	28.36	21.80	9.78	—	0.004	7
纹理	11.56	20.35	20.86	23.57	18.41	—	0.002	8

2.3 数据集样本划分

本文使用 LIDC-IDRI^[17] 中有放射科医师标注语义特征等级的 2635 个肺结节 CT 影像数据来对本文提出的方法进行评估。本文使用文献^[11,18] 中正负样本定义的方法划分结节各语义特征的正负样本,结果如表 2 所列。其中,等级 1,2,3 的数据为负样本,表示特征不明显;等级 4,5 的数据为正样本,表示特征很明显。

表 2 肺结节语义特征的正负样本划分

Table 2 Binary labels for nodule characteristics

语义特征	Label0	Label1
恶性程度	Level1-3	Level4-5
分叶征	Level1-3	Level4-5
毛刺征	Level1-3	Level4-5
对比度	Level1-3	Level4-5

2.4 数据预处理

(1)数据归一化:LIDC-IDRI 数据集中的 CT 图像放射密度强度值的范围是 $[-1\ 024, 2\ 000]$,其中放射密度强度在 $[-1\ 000, 400]$ 区间的是处理肺结节需要考虑的,所以将 CT 图像中的放射密度强度剪切到 $[-1\ 000, 400]$ HU(Hounsfield

Unit)单位之间,并将它们归一化到 $(0,1)$ 。

(2)构造三维肺结节图像输入数据:将含有肺结节的 CT 图像剪切成 $40\text{ mm} \times 40\text{ mm}$ (最大的肺结节为 30 mm) 大小,并采用双线性插值法将图像大小调整为 $20\text{ mm} \times 20\text{ mm}$,以降低网络计算量。以候选结节 CT 图像为中心,在 CT 图像序列中选取该图像前后各 5 张 CT 图像,构成 $20 \times 20 \times 10$ 的三维肺结节立方体作为卷积神经网络的输入。

(3)数据增强:由于标注的肺结节数据量较少,本文对数据样本进行旋转增强,对肺结节质心坐标每个体素沿着每个轴在横向平面做 $90^\circ, 180^\circ$ 和 270° 旋转,最后得到 10 540 个数据样本。

3 模型的网络结构

IMACNN 模型的网络结构如图 2 所示。该模型的网络结构与传统的卷积神经网络的区别在于,该模型包含多个独立的 CNN 分支模块,各分支单独训练,分别学习肺结节恶性程度特征和其他 3 个语义特征,然后再将 4 个分支训练的结果融合起来对肺结节的恶性程度进行分类。

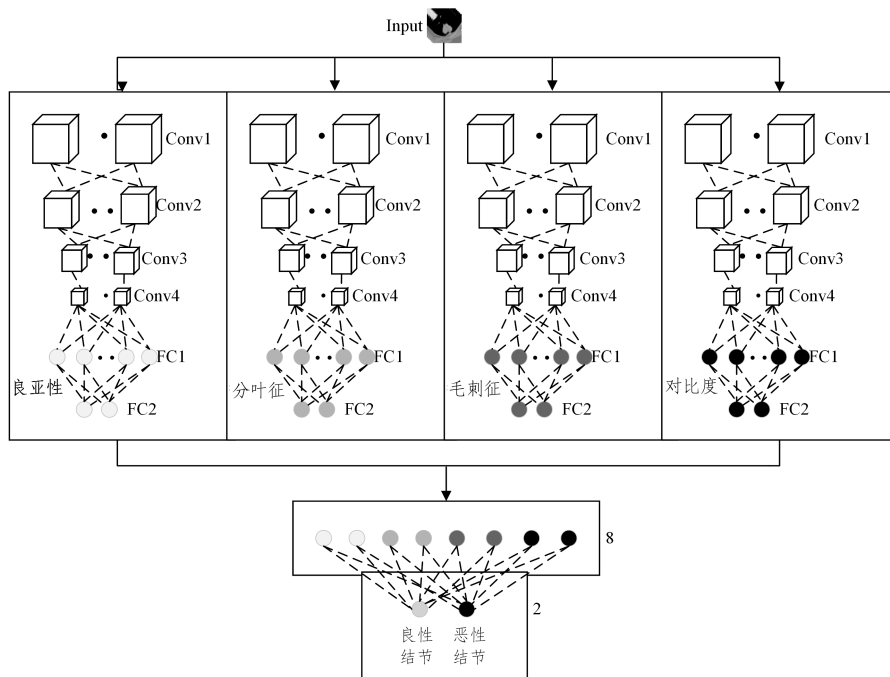


图 2 IMACNN 模型的网络结构

Fig. 2 Structure of IMACNN

3.1 单分支卷积神经网络结构

每个分支网络的输入是处理好的 $20 \times 20 \times 10$ 的三维 CT 图像立方体,每个分支的网络结构相同,都由多个卷积层、池化层、全连接层和 softmax 层组成。经过多次实验,设计各个 CNN 分支的具体网络参数,如表 3 所列。

(1)卷积层(Conv1, Conv2, Conv3, Conv4)。输入图像首先被送入卷积层,与三维卷积核进行如下卷积操作:

$$\mathbf{X}_j^l = \sigma \left(\sum_{i \in M_j} \mathbf{X}_i^{l-1} * \mathbf{W}_{ij}^l + b_j^l \right) \quad (2)$$

其中, \mathbf{X}_j^l 和 \mathbf{X}_i^{l-1} 分别表示第 l 层的第 j 个三维特征体和第 $l-1$ 层的第 i 个三维特征体; \mathbf{W}_{ij}^l 是连接 \mathbf{X}_j^l 和 \mathbf{X}_i^{l-1} 的三维卷积核; b_j^l 是偏置项; $\sigma(\cdot)$ 是非线性激活函数(本文使用 ReLU 函数)。

(2)全连接层(FC1, FC2)。输入的图像经过 4 次卷积之后,得到 256 个三维特征体,将其输入两个全连接层:

$$\mathbf{X}^f = \sigma(b^f + \mathbf{W}^f \mathbf{X}^{f-1}) \quad (3)$$

式(3)表示,全连接层将特征体展平为神经元向量 \mathbf{X}^{f-1} ,然后与权重矩阵 \mathbf{W}^f 逐元素相乘,并添加偏置项 b^f ,最后应用非线性激活函数 $\sigma(\cdot)$ (ReLU 函数),输出第 f 层的神经元向量 \mathbf{X}^f 。两个全连接层的神经元数目分别是 200 和 2。

(3)softmax 层。把全连接层输出的神经元向量 \mathbf{X}^f 输入 softmax 层,就可以预测每一类别 c 的概率分布。其定义为:

$$P_c(\mathbf{X}^f) = e^{X_c^f} / \sum_{c=0}^{C-1} e^{X_c^f} \quad (4)$$

其中, C 表示类别总数(这里是一个二分类问题), X_c^f 表示神

经元向量的第 c 个元素。

(4) 损失函数。每个分支使用的是交叉熵损失函数, 损失函数定义如下:

$$Loss = -\frac{1}{N} \sum_{j=1}^N \sum_{c=0}^{C-1} y_{j,c} \log \hat{y}_{j,c} \quad (5)$$

其中, $y_{j,c}$ 和 $\hat{y}_{j,c}$ 分别表示第 j 个数据样本属于 c 类的真实标签和网络预测标签 (softmax 函数输出的概率分布), N 为数据样本数。在网络训练过程中, 通过最小化损失来优化模型参数。

表 3 IMACNN 模型各分支的具体网络参数

Table 3 Parameters of single-branch convolution neural network

网络层	卷积核	通道
Input	—	1
Conv1	$5 \times 5 \times 3$	64
Conv2	$5 \times 5 \times 3$	128
Conv3	$5 \times 5 \times 3$	256
Conv4	$1 \times 1 \times 1$	256
FC1	—	200
FC2	—	2
Softmax	—	2

3.2 多分支卷积神经网络融合

IMACNN 模型将恶性程度分支和语义特征分支训练的结果融合起来对肺结节的恶性程度进行分类。具体融合方式如图 2 所示, 在得到各分支的全连接层的 2 维输出向量后, 将各分支的输出向量通过级联操作融合成 8 维向量, 再输入两

表 5 IMACNN 模型的对比实验结果

Table 5 Results of comparative experiments

模型	准确率	灵敏度	特异性	AUC 值	参数量/百万	FLOPs/亿
IMACNN(恶性程度+6 特征)	0.938	0.845	0.979	0.942	434.88	23.94
IMACNN(恶性程度+3 特征)	0.978	0.957	0.993	0.976	248.5	13.68
IMACNN(恶性程度)	0.866	0.696	0.999	0.845	62.13	3.42
NoduleX ^[19]	0.932	0.879	0.985	0.971	11.23	13
Fuse-TSD ^[20]	0.895	0.842	0.920	0.966	16.85	4.21
MC-CNN ^[11]	0.871	0.770	0.930	0.930	40.63	8117.49

由表 5 可知, “IMACNN(恶性程度+3 特征)”模型准确率、灵敏度和 AUC 值分别为 97.8%, 95.7% 和 97.6%, 分别比“IMACNN(恶性程度)”高出 11.2%, 26.1% 和 13.1%, 两者特异性相差不多。该实验结果说明, 卷积神经网络在数据量较少的情况下对大量特征不能充分学习, 而融合的肺结节语义特征分支可以弥补恶性程度分支对大量特征学习的不充分性, 帮助恶性程度分支更好地进行分类。

本文还训练了融合恶性程度和 6 个语义特征分支 (除了“内部结构”和“钙化程度”外, 包括分叶征、毛刺征、对比度、球形度、边缘、纹理 6 个语义特征) 的模型“IMACNN(恶性程度+6 特征)”作为对比实验, 两者各分支网络结构一致, 训练数据和训练方法一致。如表 5 所列, 经过特征选择后的“IMACNN(恶性程度+3 特征)”模型的各项指标皆高于“IMACNN(恶性程度+6 特征)”模型, 该结果证明了并不是所有的肺结节语义特征对肺结节恶性程度分类都有促进作用, 部分语义特征可能会对分类结果产生抑制作用, 因此对肺结节语义特征进行敏感性分析非常重要。

如表 5 所列, 与现有的恶性程度分类模型相比, 模型

个全连接层, 得到肺结节良恶性的二分类结果。

4 实验

4.1 实验结果

4.1.1 单分支网络训练结果

各个肺结节语义特征网络单独训练和验证, 训练过程使用五折交叉验证, 采用 Momentum 为优化器进行反向传播更新, 动量系数为 0.9, 学习率为 0.01, mini-batch 为 64。每个分支的训练结果如表 4 所列, 分叶征、毛刺征和对比度的分类准确率分别达到 0.923, 0.916 和 0.709, 同时恶性程度分支单独训练的诊断准确率达 0.866。

表 4 单分支网络的分类结果

Table 4 Classification performance for each single network

语义特征分支	准确率	灵敏度	特异性	AUC 值
分叶征	0.923	0.609	0.985	0.875
毛刺征	0.916	0.618	0.986	0.883
对比度	0.709	0.687	0.849	0.591
恶性程度	0.866	0.696	0.999	0.845

4.1.2 多分支融合分类结果

把 3 个肺结节语义特征 (分叶征、毛刺征和对比度) 分支训练得到的预测结果融合到卷积神经网络中来形成 IMACNN 网络, 再使用该网络对肺结节恶性程度进行分类。IMACNN 网络和单分支网络的训练方式一致。得到肺结节恶性程度分类的实验结果如表 5 所列。

“IMACNN(恶性程度+3 特征)”在分类评估各项指标上都取得了不错的结果。除此之外, 按照文献[18]中对网络模型复杂度的计算方法, 使用“参数量”和“FLOPs”分别对本文提出的模型和现有模型进行复杂度评估。由表 5 可知, NoduleX^[19] 和 Fuse-TSD^[20] 模型的复杂度远远小于 MC-CNN^[11] 和 IMACNN 的复杂度, 因为 MC-CNN^[11] 和 IMACNN 都是三维模型, 所需网络计算更多。同时, 本文提出的 IMACNN 模型的“FLOPs”指标远远低于同是三维模型的 MC-CNN^[11], 这说明本文提出的 IMACNN 模型能在模型复杂度较小的情况下产生更好的肺结节分类结果。

4.1.3 IMACNN 模型的过程可解释性

为了更好地解释该模型, 本文将各个语义特征分支卷积神经网络的卷积过程可视化: 在每个网络分支中, 把第三层卷积层卷积操作之后得到的肺结节中心切片的子特征图输出, 再将各子特征图合在一起, 形成总的特征图。每个分支的总特征图可视化结果如图 3 所示 (以 3 个结节为例)。由图 3 可知, 对于同一个结节, 每个分支卷积过后得到的特征图均存在差异, 这是因为每个分支提取的肺结节语义特征不同。比如,

肺结节的对比度分支主要提取肺结节与周围环境的对比度信息,特征图中,结节区域与非结节区域亮点对比明显;分叶征分支主要提取肺结节的轮廓信息,特征图中,亮点集中在结节轮廓周围;毛刺征分支主要提取结节的边缘信息,特征图中,亮点集中在结节周围,且亮点呈放射状;而恶性程度分支不仅提取结节区域信息,还提取结节周围环境信息,特征图中,亮点分布较均匀。

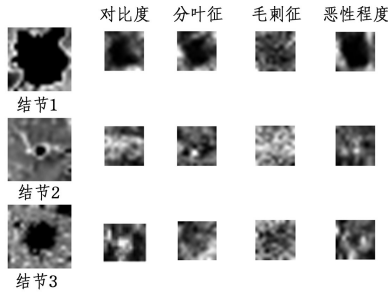


图3 神经网络特征图的可视化

Fig.3 Visualization of neural network feature map

4.1.4 IMACNN 模型的结果可解释性

如表4所列,本文模型在诊断肺结节分叶征、毛刺征和对比度上分别达到了92.3%,91.6%和70.9%的准确率。该模型能在完成肺结节恶性程度分类任务的同时,给出肺结节各语义特征的诊断结果,为放射科医师提供可解释的诊断结果。

图4展示了IMACNN模型具有可解释性的诊断结果(以两个肺结节为例)。图4(a)中,IMACNN模型预测结节1为恶性结节,并且结节1与其周围环境有很高的对比度,结节周围有明显毛刺状和分叶状,结节1的这些语义特征预测结果与医师所认知的恶性结节的语义特征是一致的。因此,与普通的3D卷积神经网络模型相比,IMACNN模型在解释它的预测结果方面更具说服力。图4(b)中,结节2和结节1的预测结果不同,IMACNN模型预测结节2与周围环境无明显对比度,且结节周围无明显毛刺状和分叶状。该语义特征预测结果可以解释为什么IMACNN模型预测结节2为良性结节,这与放射科医师诊断结节良恶性的过程是一致的。

	真实标签	预测标签
分叶征	1	1
毛刺征	1	1
对比度	1	1
良恶性	1	1

(a)结节1

	真实标签	预测标签
分叶征	0	0
毛刺征	0	0
对比度	0	0
良恶性	0	0

(b)结节2

图4 IMACNN模型的可解释性诊断结果

Fig.4 Interpretable diagnostic results of IMACNN model

图5展示了IMACNN模型对肺结节语义特征分类或者良恶性分类不准确的两个具有代表性的例子。图5(a)中,IMACNN模型对肺结节良恶性产生错误预测,却能够正确预

测所有语义特征;图5(b)中,IMACNN模型能够正确预测肺结节4的良恶性,却错误预测部分语义特征(分叶征)。这两个实验结果说明,对于小部分结节,良恶性的正确判断与这3个语义特征的正确判断不一致。出现这种情况的原因有:对于大多数结节来说,IMACNN模型选择的3个语义特征对判断良恶性是有帮助的;但是也存在小部分结节,其良恶性的判断不仅仅依赖IMACNN模型中的3个语义特征,还依赖其他语义特征。因此,在未来的工作中,我们将会研究如何通过融合更多语义特征来对模型进行改进,同时让模型针对不同结节学习不同语义特征的权重,以增强模型的可变性。

	真实标签	预测标签
分叶征	0	0
毛刺征	0	0
对比度	1	1
良恶性	1	0

(a)结节3

	真实标签	预测标签
分叶征	1	0
毛刺征	1	1
对比度	1	1
良恶性	0	0

(b)结节4

图5 IMACNN模型分类不准确的例子

Fig.5 Examples where IMACNN model incorrectly classifies

结束语 本文提出了一种可解释的多分支卷积神经网络模型,用于肺结节分类,选择对肺结节良恶性敏感性较高的语义特征与肺结节分类卷积神经网络相融合形成多分支网络,每个分支提取肺结节不同的特征,同时得到肺结节特征的预测和恶性程度的诊断结果,使得诊断结果可解释。实验证明,IMACNN模型不仅可以得到可解释的肺结节分类结果,而且实现了更好的肺结节分类性能。

由于CNN的学习能力强,数据集数据样本少,因此所有肺结节分类网络模型都存在过度拟合的风险。另外,本文提出的模型对小部分结节存在语义特征的不完整性,未来我们将研究如何融合更多语义特征以对模型进行改进,同时让模型针对不同结节学习不同语义特征的权重,以增强模型的可变性。此外,现有的模型方法中,很少有用到患者的其他病史信息,如吸烟史和家族史,这与医生临床诊断过程有所不同,我们未来也将尝试把患者的病史等信息融合到CNN模型中,以提高其泛化能力。

参考文献

[1] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2016 [J]. CA: A Cancer Journal for Clinicians, 2016, 66(1): 7-30.

[2] SHEN S, HAN S X, PETOUSIS P, et al. A Bayesian model for estimating multi-state disease progression [J]. Computers in Biology and Medicine, 2017, 81: 111-120.

[3] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening [J]. New England Journal of Medicine, 2011, 365(5): 395-409.

[4] ARMATO III S G, ALTMAN M B, WILKIE J, et al. Automa-

- ted lung nodule classification following automated nodule detection on CT: A serial approach[J]. *Medical Physics*, 2003, 30(6): 1188-1197.
- [5] SHEN S, BUI A A T, CONG J, et al. An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy[J]. *Computers in Biology and Medicine*, 2015, 57: 139-149.
- [6] DUGGAN N, BAE E, SHEN S, et al. A technique for lung nodule candidate detection in CT using global minimization methods [C]// *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. Cham: Springer, 2015: 478-491.
- [7] FIRMINO M, ANGELO G, MORAIS H, et al. Computer-aided detection (CADE) and diagnosis (CADx) system for lung cancer with likelihood of malignancy[J]. *Biomedical Engineering Online*, 2016, 15(1): 2.
- [8] AMIR G J, LEHMANN H P. After detection: The improved accuracy of lung cancer assessment using radiologic computer-aided diagnosis[J]. *Academic Radiology*, 2016, 23(2): 186-191.
- [9] HUANG P, PARK S, YAN R, et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study[J]. *Radiology*, 2017, 286(1): 286-295.
- [10] SHEN W, ZHOU M, YANG F, et al. Multi-scale Convolutional Neural Networks for Lung Nodule Classification[J]. *Inf Process Med Imaging*, 2015, 24: 588-599.
- [11] SHEN W, ZHOU M, YANG F, et al. Multi-crop Convolutional Neural Networks for Lung Nodule Malignancy Suspiciousness Classification[J]. *Pattern Recognition*, 2016, 61(61): 663-673.
- [12] PIEDRA E A R, TAIRA R K, EL-SADEN S, et al. Assessing variability in brain tumor segmentation to improve volumetric accuracy and characterization of change[C]// *IEEE-EMBS International Conference on Biomedical & Health Informatics*. IEEE, 2016.
- [13] DOU Q, CHEN H, YU L, et al. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection[J]. *IEEE Transactions on Biomedical Engineering*, 2017, 64(7): 1558-1567.
- [14] LIANG C H, YUE J Y, HAN D M, et al. Diagnostic value of lobulation sign in solitary pulmonary nodule by CT to determine benignity or malignancy[J]. *Medical Information (Section of Operative Surgery)*, 2007(10): 94-95.
- [15] HUSSEIN S, CAO K, SONG Q, et al. Risk stratification of lung nodules using 3d cnn-based multi-task learning[C]// *International Conference on Information Processing in Medical Imaging*. Cham: Springer, 2017: 249-260.
- [16] KIM H, PARK C M, GOO J M, et al. Quantitative Computed Tomography Imaging Biomarkers in the Diagnosis and Management of Lung Cancer[J]. *Investigative Radiology*, 2015, 50(9).
- [17] ARMATO S, MCLENNAN G, M MCNITT-GRAY, et al. W-EB-201B-02: The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Public Database of CT Scans for Lung Nodule Analysis[J]. *Medical Physics*, 2010, 37(6Part6): 3416-3417.
- [18] MOLCHANOV P, TYREE S, KARRAS T, et al. Pruning Convolutional Neural Networks for Resource Efficient Inference[J]. *International Conference on Learning Representations*, 2017, 2: 324-332.
- [19] CAUSEY J, ZHANG J Y, MA S Q, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans[J]. *Scientific Reports*, 2018, 8(1): 9286.
- [20] XIE Y, ZHANG J, XIA Y, et al. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT[J]. *Information Fusion*, 2018, 42: 102-110.



ZHANG Jia-jia, born in 1994, postgraduate. Her main research interests include medical image analysis, deep learning and so on.



ZHANG Xiao-hong, born in 1973, professor. His main research interests include data mining of software engineering, topic modeling, image semantic analysis and video analysis.