

# 基于胸部 X 射线影像的常见疾病预测方法研究

王江峰<sup>1</sup>, 刘利军<sup>1,2\*</sup>, 黄青松<sup>1</sup>, 刘骊<sup>1</sup>, 付晓东<sup>1</sup>

<sup>1</sup>昆明理工大学信息工程与自动化学院, 云南 昆明 650500;

<sup>2</sup>云南大学信息学院, 云南 昆明 650091

**摘要** 利用 X 射线影像进行胸部疾病诊断是一种常用的诊断方法, 具有重要的临床诊断价值。随着大规模可用数据集的发布, 已经提出了几种利用胸部 X 射线图像预测常见疾病的方法。然而大多数现有的预测模型大都仅考虑单个视图, 忽略了多视图影像对于临床医生诊断的支持作用。此外使用单个模型进行影像特征抽取时, 存在有效特征提取不全的问题, 进而导致疾病预测准确率较低。为此, 提出了一种新的深度相关多级特征融合方法 (DFFM), 该方法融合不同模型提取的不同视图的视觉特征, 以提高疾病预测的准确性。并在目前最大的胸部 X 射线数据集 MIMIC-CXR 上进行了验证, 实验结果表明, 所提方法的 area under the receiver operating characteristic curve (AUC) 值达 0.847, 与现有的单视图及简单进行特征拼接的多视图模型相比, AUC 值分别提升了 12.6 个百分点和 5.3 个百分点, 验证了所提多级融合方法的有效性。

**关键词** 医用光学; 疾病预测; 模型融合; 深度相关; 多视图; 特征抽取

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP202259.1817001

## Prediction Method for Common Diseases Based on Chest X-Ray Images

Wang Jiangfeng<sup>1</sup>, Liu Lijun<sup>1,2\*</sup>, Huang Qingsong<sup>1</sup>, Liu Li<sup>1</sup>, Fu Xiaodong<sup>1</sup>

<sup>1</sup>School of Information Engineering and Automation, Kunming University of science and technology, Kunming 650500, Yunnan, China;

<sup>2</sup>School of Information, Yunnan University, Kunming 650091, Yunnan, China

**Abstract** X-ray imaging is a commonly used diagnostic method with important clinical value in chest-disease diagnosis. Exploiting the release of large-scale available datasets, several methods have been proposed for predicting common diseases using chest X-ray images. However, most of the existing predictive models are limited to single-view inputs, ignoring the supportive role of multiview images in clinical diagnosis. Additionally when image features are extracted using a single model, the effective features are incompletely extracted and the accuracy of disease prediction decreases. The present study proposes a new depth-dependent multilevel feature fusion method (DFFM) that combines the visual features of different views extracted via different models to improve the accuracy of disease prediction. DFFM was verified using MIMIC-CXR, the largest available chest X-ray dataset. Experimental results show that the area under the receiver operating characteristic curve was 0.847, 12.6 and 5.3 percentage points higher than the existing single-view and multiview models with simple feature splicing, respectively. These results confirm the effectiveness of the proposed multilevel fusion method.

**Key words** medical optics; disease prediction; fusion model; depth correlation; multi view; feature extraction

## 1 引言

胸部疾病是影响现代社会公众健康安全的主要威胁之一, 据世界卫生组织 (WHO) 相关报告, 每年约有 4.5 亿例肺炎病例记录在案, 约有 400 万人死于该疾病, 占全球患病死亡总人数的 (5700 万人) 7%<sup>[1]</sup>。胸部

X 射线是诊断胸部疾病的最常见的方法<sup>[2]</sup>。国内外专家主要通过胸部病症的早诊早治来降低死亡率<sup>[3-4]</sup>。然而, 在临床医学中阅读胸部 X 射线光片需要极强的专业知识, 即使对有经验的医生来说阅读 X 光片也很费时费力, 并且由于胸部病理特征的多样性及放射科医生诊断的主观性, 在实际医学诊断过程中会存在误

收稿日期: 2021-06-02; 修回日期: 2021-07-12; 录用日期: 2021-07-12

基金项目: 国家自然科学基金 (81860318, 81560296)

通信作者: \*cloneiq@126.com

诊漏诊的问题<sup>[5]</sup>。另一方面,胸部疾病预测作为临床医学诊断报告生成的前期工作,针对胸部 X 射线疾病预测准确率的提升,对于提升诊断报告的完备性、准确性有着极其重要的意义<sup>[6]</sup>。因此,亟需开发高准确率的胸部 X 射线辅助诊断算法以支持放射科医师的临床工作,从而有效地对患者进行准确、快速的疾病预测。

近年来,医学人工智能作为一种新兴技术,在全球范围内引起了极大的关注<sup>[7]</sup>。计算机辅助诊断技术的迅速发展在许多医学领域产生了重要影响,如皮肤病变分割<sup>[8]</sup>、糖尿病检测<sup>[9]</sup>、肺结节检测<sup>[10-11]</sup>、医学诊断报告生成<sup>[12]</sup>等。目前,已有学者将计算机辅助诊断技术应用于胸部 X 射线影像疾病预测任务。文献<sup>[13]</sup>针对 ChestX-Ray14 数据集,训练了一个多标签 Deep Convolutional Neural Network (DCNN)分类模型来解决问题,在模型中省去了全连接层和最终的分层,在最后插入一个过渡层、一个全局池化层、一个预测层和一个损失层,通过组合过渡层、深度激活层和预测层的权重确定疾病的空间位置。文献<sup>[14]</sup>针对 ChestX-Ray14 数据集,提出了 CheXNet 模型,该模型通过调用在 ImageNet 数据集上预训练的 DenseNet-121 变体实现了疾病预测。随后文献<sup>[15]</sup>针对 ChestX-Ray14 数据集提出了一种新算法,该算法通过压缩激励模块实现自适应特征重标定,以提升网络的细粒度分类能力。在 ChestX-Ray14 官方数据集上,该算法对 14 种常见疾病的预测效果最好,平均 area under the receiver operating characteristic curve (AUC) 值达到 0.83。文献<sup>[5]</sup>针对 MIMIC-CXR 数据集,分别抽取了正视图影像与侧视图影像,提出了一种新的 DualNet 架构,该架构通过同时处理放射学检查中获得的正面和侧面 Chest-X ray (CXR) 图像来模拟临床实践。文献<sup>[16]</sup>针对 MIMIC-CXR 数据集,提出了一种结合不同类型

标签的算法,该算法引入标签上的注意力,动态抽取更好的标签集作为训练数据。该算法在 MIMIC-CXR 数据集上获得了比其他模型更为优秀的结果,对于胸部常见的 14 种疾病的预测,平均 AUC 值达到了 0.794。

虽然上述算法取得了较好的结果,但是仍然存在一些不足。1)不同于自然图像,在临床医学诊断过程中,医生会按照患者情况安排 X 射线的拍摄,有时单单一正视图的 X 射线影像不足以满足医生的诊断需求,可能还需要更多视图的 X 射线影像相互结合起来观察,才能对患者做出准确全面的疾病诊断<sup>[17]</sup>。2)对于同一张影像而言,不同的模型提取的特征信息存在差异,如果只是采用相同的模型提取影像特征,之后对其进行简单的拼接,无法将不同视图、不同模型之间的特征深度关联起来<sup>[18]</sup>。3)上述算法均使用二元交叉熵作为损失函数,无法解决正负样本不平衡的问题<sup>[19]</sup>。

基于此,本文提出了一种基于深度学习的胸部常见疾病预测方法——深度相关多级特征融合模型 (DFFM)。在 ResNet<sup>[20]</sup>和 DenseNet<sup>[21]</sup>的基础上引入了子图特征提取模块,通过卷积注意力模块<sup>[22]</sup>以提升模型对于重要特征的选择能力,使其关注于重要信息。另外,加入了一个深度强相关的多级影像特征融合模块使得不同视图的影像特征能深度相互关联,采用焦点损失<sup>[19]</sup>降低易分类样本的权重使模型能够更多地去关注难以分类的样本,以实现胸部常见疾病的预测。

## 2 基本原理

### 2.1 DFFM 总体框架

传统的现代医学诊断不仅会拍摄正面的胸部 X 光片,为了更准确的诊断,必要时也会拍摄其他体位的 X 光片以辅助诊断,图 1 为所提 DFFM 的总体框架。

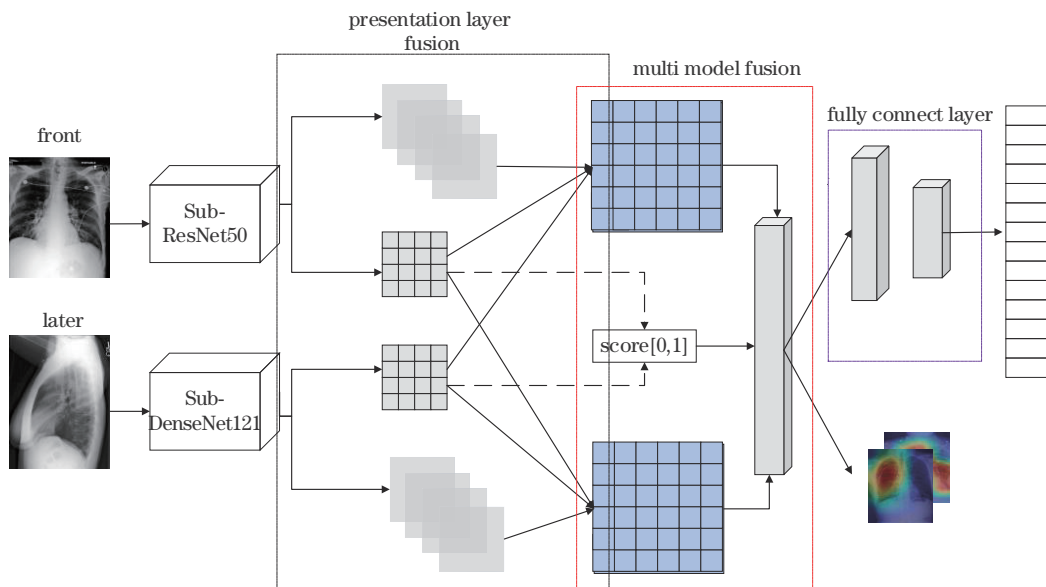


图 1 DFFM 总体框架

Fig. 1 General framework of DFFM

DFFM 同时使用 ResNet 和 DenseNet, 首先在小型数据集上使用 ResNet 和 DenseNet 模型针对医学影像进行预训练, 微调模型以在小型数据集上达到最优效果, 保存好训练好的模型参数。将之前训练好的模型作为提取正面胸部影像和侧面胸部影像的特征提取器, 分别提取不同影像的特征: Positive ResNet Feature (P-RF)、Later DenseNet Feature (L-DF)。对于提取到的图像特征不是简单对其进行串联或是并联, 而是进行多级融合使不同模型、不同视图间的特征深度相关, 从而利用不同视图、不同模型的信息来细化医学影像的表示, 最后将完全融合之后的特征输入全连接网络, 获得关于胸部影像的常见疾病的预测结果。

### 2.2 图像块特征提取

本实验组在 ResNet50 和 DenseNet121 的基础上加入了 Convolutional Block Attention Module (CBAM), 使模型学习到的特征覆盖到影像的更多部位, 提高判别不同病变区域的概率。除此之外, 还替换了预训练模型最后的全连接层, 并对模型进行了微调, 将输入图像的大小调整为  $448 \times 448$ , 并且将原始的医学影像划分为 4 个区域。具体来说, 对原始医学图像进行平均裁剪, 划分为 4 个  $224 \times 224$  大小的子区域, 并将每个子区域的大小重新调整为  $448 \times 448$ , 裁剪的意义在于输入模型的原始影像是经过压缩的图像, 会造

成病变区域不明显、有效信息缺失等问题, 对于原始图像进行裁剪放大可以使得病变区域更为明显, 对于模型提取有效特征具有积极意义。

对每个子区域进行特征提取, 获得的每个子区域的影像特征可表示为

$$\mathbf{V}_{R-sub-i} = \text{ResNet}(\mathbf{I}_i), \quad (1)$$

$$\mathbf{V}_{D-sub-i} = \text{DenseNet}(\mathbf{I}_i), \quad (2)$$

式中:  $\mathbf{I}_i$  表示不同子区域的图像,  $i=0, 1, 2, 3$ ;  $\mathbf{V}_{R-sub-i}$ 、 $\mathbf{V}_{D-sub-i}$  分别表示不同子区域利用不同模型提取得到的影像特征。将所有子区域图像特征的均值作为图像的引导矢量, 有

$$\mathbf{V}_{R-mean} = \sum_{i=1}^N \mathbf{V}_{R-sub-i} / N, \quad (3)$$

$$\mathbf{V}_{D-mean} = \sum_{i=1}^N \mathbf{V}_{D-sub-i} / N, \quad (4)$$

式中:  $N$  表示子区域的数量;  $\mathbf{V}_{R-mean}$ 、 $\mathbf{V}_{D-mean}$  分别表示不同子模型得到的影像特征的均值。

所用图像块特征提取模块的通道注意力子模块结构如图 2 所示, 为了有效计算通道注意力, 压缩了输入特征图的空间维数, 为了聚集空间信息, 同时使用平均池化和最大池化特性生成两个不同方向上的上下文描述  $\mathbf{F}_{cavg}$ 、 $\mathbf{F}_{cmax}$ , 分别表示平均池化特征和最大池化特征, 之后将两类特征输入一个共享网络中, 通过激活函数将特征图还原到原始大小, 即获得添加了通道注意力的特征。

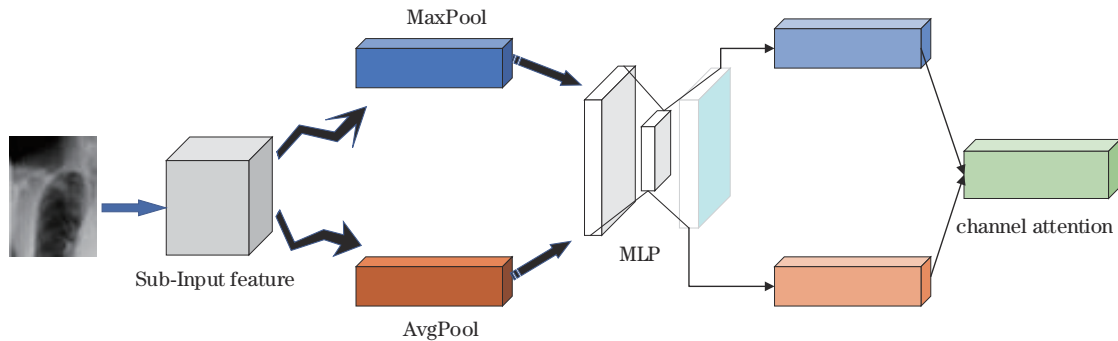


图 2 通道注意力子模块  
Fig. 2 Channel attention sub module

添加了通道注意力的图像特征进入空间注意力模块后, 首先用平均池化和最大池化将其连接起来, 生成一个有效的特征描述符; 之后沿通道轴应用池化操作显示出有效区域, 并在之前产生的特征描述符上采用

卷积操作生成空间注意力编码, 对图中的有效区域进行加强或是抑制, 最终得到通道注意力和空间注意力共同作用的子图特征。图像块特征提取模块的空间注意力子模块结构如图 3 所示。

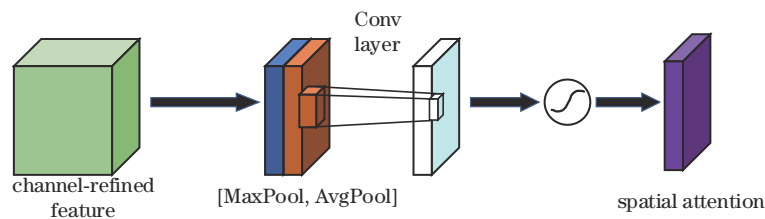


图 3 空间注意力子模块  
Fig. 3 Spatial attention sub module

### 2.3 表示层特征融合

表示层  $i$  的输入为从 ResNet 特征提取子模型中得到的总体特征、特征均值及从 DenseNet 模型中得到的输入影像特征的均值。首先, 将从 ResNet 中得到的总体特征分别与 ResNet 模型提取的特征的均值与 DenseNet 模型提取的总体特征拼接得到一个特征列:

$$\mathbf{V}_{R\text{-sum}} = \sum_{i=1}^N \text{ResNet}(\mathbf{I}_i), \quad (5)$$

$$\mathbf{V}_{D\text{-sum}} = \sum_{i=1}^N \text{DenseNet}(\mathbf{I}_i), \quad (6)$$

$$\mathbf{V}_{R\text{-mean}} = \mathbf{V}_{R\text{-sum}}/N, \quad (7)$$

$$\mathbf{V}_{D\text{-mean}} = \mathbf{V}_{D\text{-sum}}/N, \quad (8)$$

式中:  $\mathbf{V}_{R\text{-sum}}$ 、 $\mathbf{V}_{D\text{-sum}}$  分别表示从不同模型中得到的总体特征;  $\mathbf{V}_{R\text{-mean}}$ 、 $\mathbf{V}_{D\text{-mean}}$  分别表示不同模型提取出子图特征的均值。

紧接着, 将从 ResNet 中得到的总体特征分别与 DenseNet 模型提取的特征的均值拼接得到另一个特征列:

$$\mathbf{V}_{1i} = \mathbf{V}_{\text{ResNet-sum}} + \text{ResNet}(\mathbf{I}_i), \quad (9)$$

$$\mathbf{V}_{2i} = \mathbf{V}_{\text{DenseNet-sum}} + \text{DenseNet}(\mathbf{I}_i), \quad (10)$$

式中:  $\mathbf{V}_{1i}$ 、 $\mathbf{V}_{2i}$  分别表示针对不同模型的总体特征与特征均值拼接的结果。

其次, 在  $\text{dim}$  为 0 时分别将这两列特征值拼接起来, 对拼接之后的结果进行归一化得到不同的权重值。最后, 将得到的不同的权重值作用在总体特征上之后求得表示层的输出, 即

$$\mathbf{w}_1 = \sigma[\mathbf{v}_{11}, \mathbf{v}_{12}, \mathbf{v}_{13}, \mathbf{v}_{14}], \quad (11)$$

$$\mathbf{w}_2 = \sigma[\mathbf{v}_{21}, \mathbf{v}_{22}, \mathbf{v}_{23}, \mathbf{v}_{24}], \quad (12)$$

$$\mathbf{F} = (\mathbf{w}_1 + \mathbf{w}_2) \cdot \mathbf{V}_{R\text{-sum}}/2, \quad (13)$$

式中:  $\sigma$  为 Sigmoid 激活函数;  $\mathbf{w}_1$ 、 $\mathbf{w}_2$  表示的是不同视图影像特征的权重得分;  $\mathbf{F}$  表示最终要进行分类的特征向量。

### 2.4 多模型特征融合

对从子模型特征融合阶段得到的 ResNet-Image-Vector ( $\mathbf{F}_R$ ) 及 DenseNet-Image-Vector ( $\mathbf{F}_D$ ) 进行多次线性变化, 得到不同模型提取的特征对应的不同得分; 其次, 将两个模型的得分拼接、归一化之后得到一个归一化的得分列表; 然后, 分别对子模型特征提取阶段获取的特征进行维度调整, 以满足最终疾病预测阶段的输入; 最后, 按照不同模型特征的不同得分获得最终的融合特征。

$$\mathbf{s} = \text{Softmax}[\mathbf{R}_{\text{score}}, \mathbf{D}_{\text{score}}], \quad (14)$$

$$\mathbf{F}_u = \mathbf{F}_R \times \mathbf{s}[0] + \mathbf{F}_D \times \mathbf{s}[1], \quad (15)$$

式中:  $\mathbf{R}_{\text{score}}$ 、 $\mathbf{D}_{\text{score}}$  是通过  $\mathbf{F}_R$ 、 $\mathbf{F}_D$  进行线性变换得到的;  $\mathbf{s}$  表示得分;  $\mathbf{F}_u$  表示融合之后的图像特征。

### 2.5 焦点损失

在 ChestX-Ray14 和 MIMIC-CXR 数据集中, 每张图像都定义了一个 14 维标签向量  $\mathbf{Y} = [y_1, y_2, \dots, y_c]$ ,

每个维度代表一种疾病, 值为 1 表示患有该疾病, 值为 0 表示未患有该疾病。在之前的研究中, 绝大多数算法都选择二元交叉熵损失作为损失函数:

$$L = -y \log_a y' - (1-y) \log_a (1-y') = \begin{cases} -\log_a y, & y = 1 \\ -\log_a (1-y'), & y = 0 \end{cases}, \quad (16)$$

式中:  $y$  为疾病类别;  $y_c \in \{0, 1\}$  为真实病理标签, 包含该疾病则值为 1, 反之则值为 0;  $y' \in (0, 1)$  为模型的预测结果。

在医学图像中, 往往重要的病理信息只存在于少量的患病样本中, 正常样本过多, 而正类样本过少的现实问题使得模型无法充分学习到患病样本的信息, 另外因为疾病的种类不同, 有些疾病易于观察到而有些疾病则难以观察, 之前的交叉熵损失函数并不能解决这些问题, 所以本实验组在交叉熵损失函数的基础上增加了平衡因子  $\alpha$  和权重因子  $\beta$ ,  $\alpha = 0.25$ ,  $\beta = 2$ ,  $\alpha$  用来平衡正负样本的比例,  $\beta$  用来平衡分类正确和分类错误的样本的损失权重:

$$L = \begin{cases} -\alpha(1-y')^\beta \log_a y', & y = 1 \\ -(1-\alpha)y'^\beta \log(1-y'), & y = 0 \end{cases}. \quad (17)$$

由此, 整个模型会根据损失函数调整每种疾病的损失值, 使得难以识别疾病乘以一个较大的权重, 易于识别的疾病乘以一个较小的权重, 以改善模型对于不同识别难度的疾病的分类能力。

## 3 分析与讨论

### 3.1 数据集描述

为验证所提算法的有效性, 使用公开的数据集 MIMIC-CXR 数据集和 ChestX-ray14 数据集进行对比实验。ChestX-ray14 是由美国国立卫生研究院 (NIH) 整理并公开的大型多标签数据集, 由 65240 名患者的 224316 张胸片组成, 其中有 14 个观察结果被标记为阳性、阴性或不确定。MIMIC-CXR 是由斯坦福大学发布的大型公共胸片预处理数据集, 与 ChestX-ray14 数据集的不同之处在于 MIMIC-CXR 数据集不仅公布了病人的医学影像, 同时也发布了不同影像对应的医学诊断报告, MIMIC-CXR 由 227835 份报告及 377110 张医学影像构成, 类似的 MIMIC-CXR 也有 14 个观察结果被标记为阳性、阴性或不确定。ChestX-ray14 数据集和 MIMIC-CXR 数据集中的标签来自于人工设计的贴标机, 贴标机通过自动检测放射报告中是否存在疾病标签, 从而生成医学影像对应的标签列表。图 4 中列举了在 MIMIC-CXR 数据子集中不同标签的分布情况。

### 3.2 数据集预处理

因为硬件设备与下游工作的需求, 本实验组对原始的数据进行了挑选, 在 MIMIC-CXR 数据集选择

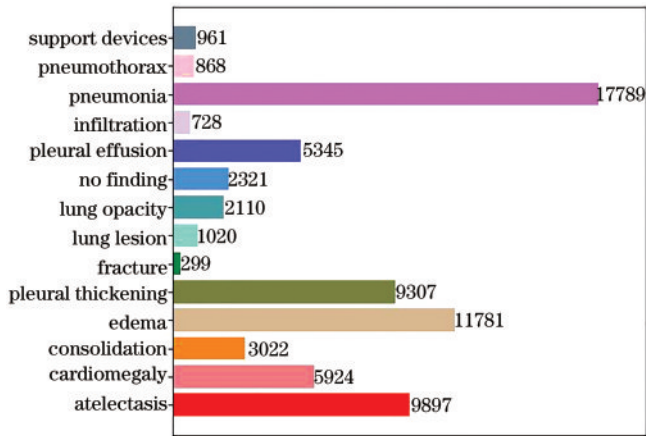


图4 MIMIC-CXR数据集中不同标签的分布情况

Fig. 4 Distribution of different tags in the MIMIC-CXR

了具有至少两张不同视图且对应的文本报告中包含了病人病史信息的病人影像,该子数据集包含了10000张医学影像。而对于ChestX-ray14数据集,由于原始数据集中并没有区分视图信息,所以本实验对于ChestX-ray14中的图像没有进行视图方面的处理。

**影像:**在提取输入影像特征之前,为了减少计算量同时匹配网络模型,将输入图像的大小调整为 $448 \times 448$ ;将resize后的图像平均切分为4份不重叠的子图;并将子图的大小调整为 $448 \times 448$ 以获得和原始图像相同大小的输入。

**标签:**因为MIMIC-CXR数据集有positive、negative、uncertain三种不同的原始标签,所以对于该数据集一般有以下5种训练模式:

- 1) U-Ignore:在训练期间忽略了不确定的标签。
- 2) U-Zeroes:将不确定标签的所有实例映射到0。
- 3) U-Ones:将不确定标签的所有实例映射到1。
- 4) U-SelfTrained:首先使用U-Ignore方法训练模型进行收敛,然后使用该模型进行预测,利用模型输出的概率预测重新标记每个确定性标签。

5) U-MultiClass:将不确定性标签视为自己的类别。

本实验组使用模式1)对模型进行训练。

### 3.3 实验装置

实验使用的硬件配置是Inter(R) Xeon(R) Silver4110 CPU、NVIDIA TITAN XP GPU,所用的操作系统为Ubuntu 16.04,采用Python3.6作为编程语言,使用Pytorch机器学习框架及基于Pytorch的深度学习库。在迭代过程中,输入一批医学影像进行批处理,采用随机梯度下降算法更新网络权重, batch的大小设置为64。在ImageNet数据集上进行预训练的最大迭代次数设置为10,初始的学习率设置为0.001。当在MIMIC-CXR数据集和ChestX-ray14数据集上进行微调时,ResNet特征提取子模型的全连接层由fc-

1000替换为fc-2048,DenseNet特征提取子模型的全连接层由fc-1000替换为fc-2048,最终融合模型的Softmax分类器的输出类别由1000改为14,训练的最大迭代次数为100。

原始X射线影像尺寸为 $3056 \times 2544$ ,本实验组将图像缩放至 $256 \times 256$ 作为网络输入,训练过程中应用随机翻转的数据增强方式,采用Adam为优化器。初始学习率设置为0.001,设置batchsize为32,每次验证损失在稳定之后学习率衰减10倍,选择验证损失最低的模型保存其参数。为了方便地与当前其他同类型的诊断算法进行比较,采用国际上常用的受试者操作特征曲线(ROC)作为每种疾病的评价指标,根据ROC曲线下面积对每种算法进行定量分析,ROC曲线越靠近左上角,其AUC值越接近1,说明算法的预测性能越好。

### 3.4 实验结果

为了验证所提算法的性能,将所提算法与文献[5,14-17]中的算法在两个不同的数据集上进行了比较,其中文献[14,16]针对ChestX-ray14数据集进行了测试,文献[5,17]针对MIMIC-CXR数据集进行了测试。

图5的ROC曲线描绘了所提算法在MIMIC-CXR数据上对胸部常见的14种疾病的分类表现,曲线整体位于左上角,表明算法整体分类性能良好。

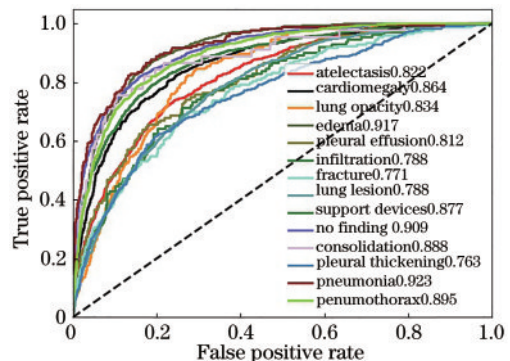


图5 MIMIC-CXR数据集上的ROC曲线和AUC值

Fig. 5 ROC curves and results on MIMIC-CXR dataset

图6为不同算法在不同数据集上对每种疾病类别的AUC值及所有类别的平均AUC值情况。

表1、表2更为详细地列出了在不同数据集上每种疾病的AUC值。表1将所提算法的分类AUC值与前人的工作(Rajpurkar等<sup>[14]</sup>,2017;张驰名<sup>[15]</sup>,2020)在ChestX-ray14数据集上的结果进行了比较,所提算法的平均AUC值达到了0.833,与之前的算法相比提升效果较低的原因可能是ChestX-ray14原始数据集中包含的不同视图影像的比例差距过大导致的。

表2中将所提算法的分类AUC值与前人工作(Rubin等<sup>[5]</sup>,2018;Wang等<sup>[16]</sup>,2020;基础ResNet;基

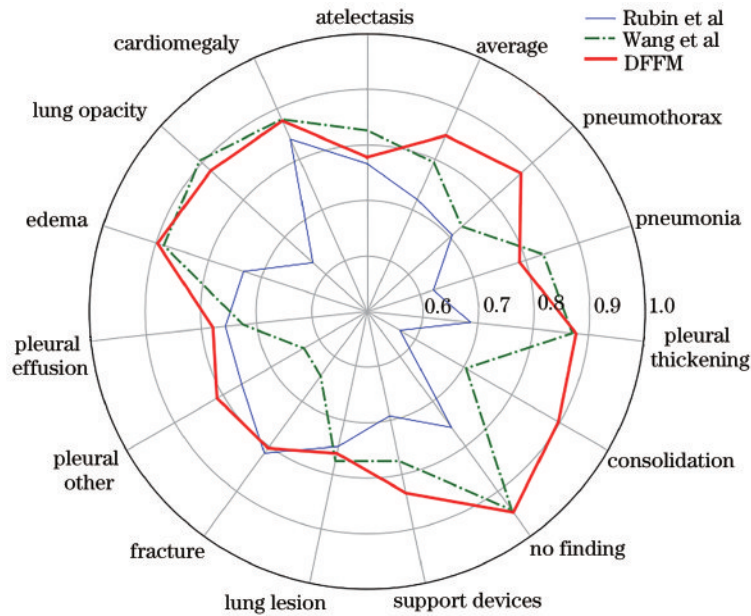


图 6 在 MIMIC-CXR 数据集中不同算法各种疾病的 AUC 值对比

Fig. 6 Comparison of AUC values of various diseases in different documents of MIMIC-CXR dataset

表 1 不同算法在 ChestX-ray14 数据集上的 AUC 值比较

Table 1 Comparison of AUC values of different algorithms on ChestX-ray14 dataset

Pathology	Rajpurkar et al	Zhang C M	DFFM
Atelectasis	0.700	0.789	0.820
Cardiomegaly	0.814	0.900	0.871
Effusion	0.759	0.838	0.808
Infiltration	0.661	0.708	0.906
Tumor	0.693	0.839	0.795
Nodules	0.669	0.798	0.839
Pneumonia	0.658	0.736	0.788
Pneumothorax	0.799	0.877	0.721
Consolidation	0.703	0.763	0.890
Edema	0.805	0.851	0.899
Emphysema	0.833	0.938	0.916
Fibrosis	0.786	0.832	0.707
Pleural Thickening	0.684	0.804	0.830
Fracture	0.871	0.944	0.877
Average	0.745	0.830	0.833

表 2 不同算法在 MIMIC-CXR 数据集上的 AUC 值比较

Table 2 Comparison of AUC values of different algorithms on MIMIC-CXR dataset

Pathology	ResNet	DenseNet	Rubin et al	Wang et al	DFFM
Atelectasis	0.695	0.687	0.766	0.826	0.822
Cardiomegaly	0.742	0.727	0.840	0.879	0.864
Lung Opacity	0.767	0.732	0.632	0.906	0.834
Edema	0.901	0.880	0.734	0.885	<b>0.917</b>
Pleural Effusion	0.726	0.646	0.757	0.725	<b>0.812</b>
Infiltration	0.510	0.468	0.761	0.632	<b>0.788</b>
Fracture	0.735	0.617	0.815	0.643	0.771
Lung Lesion	0.701	0.704	0.748	0.775	0.788
Support Devices	0.782	0.780	0.692	0.775	<b>0.877</b>
No Finding	0.878	0.872	0.758	0.942	<b>0.909</b>
Consolidation	0.632	0.615	0.568	0.705	<b>0.888</b>
Pleural Thickening	0.617	0.609	0.687	0.871	<b>0.763</b>
Pneumonia	0.859	0.839	0.625	0.833	0.923
Pneumothorax	0.808	0.891	0.706	0.729	<b>0.895</b>
Average	0.739	0.719	0.721	0.794	<b>0.847</b>

础 DenseNet) 在 MIMIC-CXR 数据集上进行结果比较。所提算法对 14 种常见的胸部疾病的识别准确率的平均值超过了对比算法,平均 AUC 值达到 0.847, 有较大的提升。渗透、肺病变、骨折、肺炎等几种疾病的 AUC 值较低,这是由于肺炎的样本数量过大,网络对其产生了一定程度的过拟合,而另外三种疾病则是由于样本数量过小,网络对于这三种疾病的特征学习不充足。而虽然气胸的样本数量同样较小,但是由于在医学上气胸多会造成纵隔、心脏向健侧移位,并且会伴有明确的气胸线,胸部正视影像与侧视影像相结

合可以更容易判断出该患者是否患有气胸,所以对于气胸的识别也较高。所提算法在其余疾病的识别分类上准确率均高于其他算法,相较于 Wang 等单模型的特征提取算法, AUC 值提升了 0.053,较 Rubin 等的基础双模型提取特征算法, AUC 值提升了 0.126。综上所述,所提算法能较好地胸部常见的 14 种疾病进行预测分类,且总体来看,所提算法对于胸部疾病识别的平均 AUC 值有一定的提升,总体分类性能较好。

在医学辅助诊断中,网络的可解释性也尤为重要,为了更为直观地反应所提模型对于胸部 X 射线检测的结果,通过 Grad-CAM 生成病灶区域热图来表现

所提模型对于胸部常见疾病检测的有效性,具体结果如图 7 所示,影像下方为与原始影像相对应的医学报告。

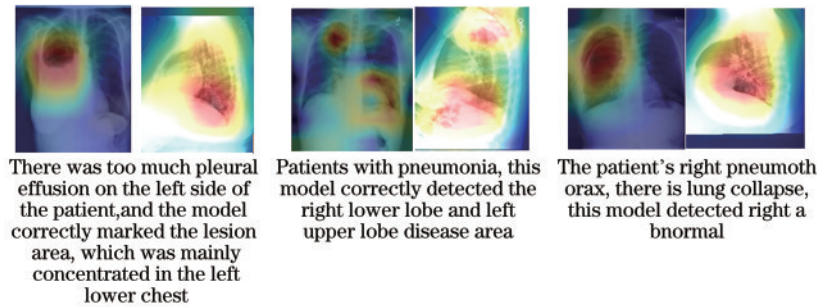


图 7 通过 Grad-CAM 生成的病灶区域热图

Fig. 7 Heat maps of lesion area generated by Grad-CAM

## 4 结 论

提出了一种针对胸部 X 射线影像,预测 14 种常见胸部疾病的模型。通过融合来自不同视图、不同模型的影像特征有效提高了对于胸部常见疾病预测的准确性。将所提融合模型与其他预测模型进行了对比,实验结果表明,多视图、多模型的多级融合算法在 MIMIC-CXR 数据集上的性能优于单模型、单视图的算法。所提算法的不足在于,在临床医学中病人的病史信息、年龄、性别等文本信息对于医生诊断疾病都起着十分重要的作用,后续的工作可以考虑将图像信息和文本信息结合,进行不同模态信息的融合以实现更高效的疾病预测。

### 参 考 文 献

- [1] Ruuskanen O, Lahti E, Jennings L C, et al. 2011. Viral pneumonia. *Lancet*[EB/OL]. (2011-03-23)[2021-02-04]. [http://biblioteca.asmn.re.it/allegati/lancet2011\\_viralpneumonia\\_120114093922.pdf](http://biblioteca.asmn.re.it/allegati/lancet2011_viralpneumonia_120114093922.pdf).
- [2] Sekuboyina A, Oñoro-Rubio D, Kleesiek J, et al. A relational-learning perspective to multi-label chest X-ray classification[EB/OL]. (2021-03-10)[2021-02-03]. <https://arxiv.org/abs/2103.06220>.
- [3] 刘若愚, 刘立波. 基于改进全卷积网络模型的肺结节检测[J]. *激光与光电子学进展*, 2020, 57(16): 161015.  
Liu R Y, Liu L B. Detection of pulmonary nodules based on improved full convolution network model[J]. *Laser & Optoelectronics Progress*, 2020, 57(16): 161015.
- [4] 高大川, 聂生东. 结合深度卷积神经网络与影像学特征的肺结节良恶性鉴别方法[J]. *光学学报*, 2020, 40(24): 2410002.  
Gao D C, Nie S D. Method for identifying benign and malignant pulmonary nodules combing deep convolutional neural network and hand-crafted features[J]. *Acta Optica Sinica*, 2020, 40(24): 2410002.
- [5] Rubin J, Sanghavi D, Zhao C, et al. Large scale automated reading of frontal and lateral chest X-rays using dual convolutional neural networks[EB/OL]. (2018-04-20)[2021-02-05]. <https://arxiv.org/abs/1804.07839v2>.
- [6] Jing B Y, Xie P T, Xing E. On the automatic generation of medical imaging reports[EB/OL]. (2017-11-22)[2021-05-06]. <https://arxiv.org/abs/1711.08195>.
- [7] 潘亚玲, 王哈琦, 陆勇. 人工智能在医学影像 CAD 中的应用[J]. *国际医学放射学杂志*, 2019, 42(1): 3-7.  
Pan Y L, Wang H Q, Lu Y. The application of computer aided diagnosis with artificial intelligence in medical imaging[J]. *International Journal of Medical Radiology*, 2019, 42(1): 3-7.
- [8] 蒋新辉, 李喆. 基于 U 型结构上下文编码解码网络的皮肤病变分割研究[J]. *激光与光电子学进展*, 2021, 58(12): 1210006.  
Jiang X H, Li Z. Skin lesion segmentation based on U-shaped structure context encoding and decoding network [J]. *Laser & Optoelectronics Progress*, 2021, 58(12): 1210006.
- [9] Rathi T. Variable weights neural network for diabetes classification[EB/OL]. (2021-02-22)[2021-06-08]. <https://arxiv.org/abs/2102.12984v1>.
- [10] 阮宏洋, 陈志澜, 程英升, 等. C-3D 可变形卷积神经网络模型的肺结节检测[J]. *激光与光电子学进展*, 2020, 57(4): 041013.  
Ruan H Y, Chen Z L, Cheng Y S, et al. Detection of pulmonary nodules based on C-3D deformable convolutional neural network model[J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 041013.
- [11] Park S, Hwang W, Jung K H. Integrating reinforcement learning to self training for pulmonary nodule segmentation in chest X-rays[EB/OL]. (2018-11-21)[2021-06-02]. <https://arxiv.org/abs/1811.08840v1>.
- [12] Syeda-Mahmood T, Wong K, Gur Y, et al. Chest X-ray report generation through fine-grained label learning[M]// Martel A L, Abolmaesumi P, Stoyanov D, et al. *Medical image computing and computer assisted intervention-MICCAI 2020. Lecture notes in computer science*. Cham: Springer, 2020, 12262: 561-571.
- [13] Kumar P, Grewal M, Srivastava M M. Boosted cascaded convnets for multilabel classification of thoracic diseases in chest radiographs[M]// Campilho A, Karray

- F, Romeny B T H. International conference image analysis and recognition. Lecture notes in computer science. Cham: Springer, 2018, 10882: 546-552.
- [14] Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning[EB/OL]. (2017-11-14) [2021-06-03]. <https://arxiv.org/abs/1711.05225>.
- [15] 张驰名. 基于深度学习方法的胸部常见病变 X 线诊断研究[D]. 绵阳: 西南科技大学, 2020.  
Zhang C M. X-ray diagnosis of common chest lesions based on deep learning method[D]. Mianyang: Southwest University of Science and Technology, 2020.
- [16] Wang X S, Xu Z Y, Yang D, et al. Learning image labels on-the-fly for training robust classification models [EB/OL]. (2020-09-22)[2021-05-06]. <https://arxiv.org/abs/2009.10325v1>.
- [17] Lenga M, Schulz H, Saalbach A. Continual learning for domain adaptation in chest X-ray classification[C]//Medical Imaging with Deep Learning, MIDL 2020, July 6-8, 2020, Montreal, QC, Canada. Cambridge: PMLR, 2020: 413-423.
- [18] Ning G H, He Z H. Dual path networks for multi-person human pose estimation[EB/OL]. (2017-10-27)[2021-03-06]. <https://arxiv.org/abs/1710.10192>.
- [19] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [20] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [21] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [22] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.