

基于机器学习的肺结节良恶性分类研究进展

杨 杨, 李晓琴*, 侯智超, 高 斌

北京工业大学环境与生命学部, 北京

Email: *lxq0811@bjut.edu.cn

收稿日期: 2021年2月3日; 录用日期: 2021年3月16日; 发布日期: 2021年5月6日

摘 要

简要分析了计算机断层扫描(computed tomography, CT)与其他肺结节成像方式的优缺点。随后介绍了目前研究者们运用最多的两个肺部计算机断层扫描影像数据库及其文件组成。最后从传统人工特征、深度学习、数据集的选择以及多分类这四个角度出发,重点介绍了肺结节良恶性分类的具体应用及进展,并加以讨论和展望。

关键词

机器学习, 深度学习, 计算机辅助诊断, 肺结节, 金标准, 结节分类

Advances in the Classification of Benign and Malignant Pulmonary Nodules Based on Machine Learning

Yang Yang, Xiaoqin Li*, Zhichao Hou, Bin Gao

Faculty of Environment and Life, Beijing University of Technology, Beijing

Email: *lxq0811@bjut.edu.cn

Received: Feb. 3rd, 2021; accepted: Mar. 16th, 2021; published: May 6th, 2021

Abstract

The advantages and disadvantages of computed tomography (CT) and other imaging methods of pulmonary nodules are analyzed. Then, the two CT image databases and their file composition that researchers use most at present are introduced. Finally, from the four perspectives of traditional artificial features, deep learning, database selection and multi-classification, the specific

*通讯作者。

application and progress of benign and malignant classification of pulmonary nodules are mainly discussed and prospected.

Keywords

Machine Learning, Deep Learning, Computer Aided Diagnosis, Pulmonary Nodule, Golden Standard, Nodule Classification

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

肺癌是发病率和死亡率增长最快, 对人群健康和生命威胁最大的恶性肿瘤之一[1]。据世界卫生组织报告, 中国在过去 40 年中, 肺癌死亡率增加了 4 倍[2], 未来可能还会进一步增加。肺癌患者的生存率一般与肺癌类型、治疗方式、癌症分期等因素有关系, 据统计, 肺癌患者在早期发现并及时治疗 5 年生存率可以接近 90%, 而晚期肺癌患者在进行治疗后的生存率仅为 2%~5%, 所以肺癌生存率与首次治疗时诊断的疾病阶段有非常重要的联系。相对于其他癌症, 肺癌患者在早期多无明显临床症状, 大约 75% 的肺癌患者被发现时已是中晚期, 此阶段治疗费用高且效果不佳, 因此肺癌的早期发现对于患者的生存和康复至关重要。临床上, 肺癌根据原发部位的不同分为原发性支气管肺癌和转移性肺癌, 原发性肺癌又分为小细胞肺癌和非小细胞肺癌, 小细胞肺癌疾病分期划分为局限期和广泛期两个主要阶段, 非小细胞肺癌疾病分期划分为 I (早期)、II (中期)、III (中晚期)、IV (晚期)。肺部是恶性肿瘤容易转移的脏器, 约有 30% 的肿瘤患者在病程的某个阶段发生肺转移, 因此转移性肺癌在肺癌中的占比也很大。

传统的肺癌早期筛查是医生通过肉眼直接观察肺部计算机断层扫描图像(computed tomography, CT), 来评估患者肺部的肺结节的良恶性程度。但一位患者的 CT 图像往往有上百张切片, 对这些图像采用人工方式进行观察、判断, 工作量巨大, 且一些 CT 图像上较小的肺结节由于不易观察可能会造成误诊漏诊[3]。所以迫切需要开发出辅助工具来帮助医生诊断这些可疑性肺结节, 提升医生对肺结节良恶性的诊断效率。

2. 肺结节的成像方式

一般肺部的影像学检查包括 X 线检查、计算机断层扫描、磁共振成像(magnetic resonance imaging, MRI)、正电子发射断层显像(positron emission tomography CT, PET-CT) [4]。X 线操作简单且价格便宜, 缺点是发现隐蔽部位难; CT 是横断扫描, 避免了器官的重叠, 能够发现细小病灶, 组织分辨率高, 密度分辨率相较于 X 线检查高, 其中的高分辨率 CT (high resolution CT, HRCT)相对于普通 CT, 可精准定位, 观察到肺结节的具体位置, 且具有更高的分辨率; 增强 CT 是在 CT 平扫基础上, 对发现的可疑部位, 在静脉注射造影剂后重点进行检查, 增加正常组织与可疑部位之间的对比, 提高病灶检出率; 低剂量螺旋 CT 相对于普通 CT 辐射量更小; MRI 拥有更好的软组织密度分辨率以及各方位成像, 对显示肿瘤较好, 更适用于中晚期、术后和放疗后患者的检查, 较 CT 图像采集时间长, 更易导致伪影, 且膨胀的肺内质子密度较低, 导致信噪比低, 图像质量差[5]; PET 是一种进行功能代谢显像的分子影像学设备, PET 检查采用正电子核素作为示踪剂, 通过病灶部位对示踪剂的摄入了解病灶功能代谢状态, 从而对疾病正确

诊断, 但需要注射对比剂, 对病灶的空间定位及定性准确率不高, 价格昂贵; PET-CT 是将 PET 获取的信息与 CT 图像相结合对病灶进行综合分析, 尤其在对诊断孤立肺结节时, 其灵敏度要比单独的 PET 或 CT 更高[6], 然而 PET-CT 费用较高昂, 辐射剂量大, 操作较复杂, 对肺结节评估受限, 也存在一定的假阳性及假阴性, 并不普及。

国家肺癌筛查试验(national lung screening trial, NLST)发现, 使用 CT 对肺癌进行筛查, 肺癌死亡率比使用 X 线投影筛查降低了 20% [7], CT 是最适合用于肺癌早期筛查的手段[8]。图 1 展示了良性结节、原发性恶性结节与转移性恶性结节在肺部 CT 中的区别。

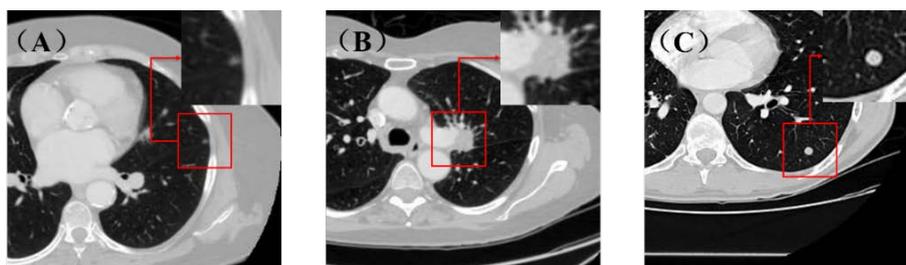


Figure 1. Representative CT images of pulmonary nodules. (A) Benign; (B) Primary lung cancer; (C) Metastatic lung cancer

图 1. 有代表性的肺结节 CT 图像。(A) 良性; (B) 原发性肺癌; (C) 转移性肺癌

3. 肺部图像数据库介绍

目前比较权威并且能够提供质量较好的胸部螺旋 CT 扫描图像的数据库有 LIDC-IDRI (lung image database consortium and image database resource initiative, LIDC-IDRI) [9]数据库和基于 LIDC-IDRI 开源数据集的美国 2016 年肺结节分析(lung nodule analysis 2016, LUNA16) [10]挑战赛开源数据集。

3.1. LIDC-IDRI 数据库

LIDC-IDRI 数据库(网址为: <https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI>), 由 7 个学术中心和 8 个医学影像公司共同合作建立, 为肺结节领域的研究者提供了一个权威且公开的标准[9]。这个数据库包含了 1010 个不同患者的 1018 套胸部螺旋 CT 扫描图像[10], 扫描使用了不同的扫描仪器和模型, 扫描层厚 0.6~5.0 mm, 512 × 512 像素, 结节大小 2.032~68.431 mm。每一个病例主要提供了 DICOM 和 XML 两类信息, 其中 DICOM 信息是每一个病例的若干张临床胸部螺旋 CT 扫描图像, XML 文件(可扩展标记语言)包含了对这个病例的注释信息, 该文件记录了四位经验丰富的胸部放射科医生执行的第二阶段图像注释过程的结果[11], 第一阶段为盲读阶段, 每个放射科医生独立地审查了每一个病例, 并标记了属于三种类型之一的病变, 三种类型分别为: 结节直径 ≥ 3 mm, 结节直径 < 3 mm, 非结节(直径 ≥ 3 mm)。第二阶段为非盲读阶段, 每位放射科医生独立检查自己的标记以及其他三名放射科医生的匿名标记, 然后提供最终结果。这样分两阶段读胸部螺旋 CT 的目的是: 在每次 CT 扫描中尽可能完整识别所有的肺结节, 不强迫医生标注一致性。

3.2. LUNA16 数据库

LUNA16 数据库(网址为 <https://luna16.grand-challenge.org/>), 基于 LIDC-IDRI 数据库, 将切片厚度 > 2.5 mm 的 CT 图像去除, 同时去除切片间距不连续和缺失切片的病例, 最终数据库包含 888 套 CT, 扫描层厚 0.45~2.5 mm, 结节大小 3.25~32.27 mm, 其中被 1~4 名医生同时标注的结节数量分别为 2290、1602、

1186 和 777。LUNA16 数据库将至少 3 位专家标注且直径 ≥ 3 mm 的结节定义为正样本，共有 1186 个。官网提供了 LUNA16 数据集的 10 个 subsets，每个 subset 中，CT 图像以 MetaImage (mhd/raw) 格式存储。每个 .mhd 文件都存储在一个单独的 .raw 二进制文件中。LUNA16 挑战赛分为两个赛道，结节检测和假阳性减少，官网并为两个赛道准备了所需的文件，annotations.csv 共有 1186 组数据，包含了病患病历号、结节的世界质心坐标以及结节的大小，candidates.csv 文件包含了病患病历号、候选者的世界质心坐标及候选者的类别。

4. 肺结节分类研究进展

图像分类中，传统的图像分类需要人工提取特征，再将特征输入分类器中从而对不同类别的图像进行分类，典型的人工特征有：方向梯度直方图(histogram of oriented gradient, HOG)特征、Gabor 特征和 Haar 特征，针对肺结节的人工特征有：结节的大小、强度、形状、纹理以及位置[12]。基于深度学习的图像分类能通过深层架构自动学习更多抽象层次的数据特征，无需对特定的图像数据或分类方式设计具体的人工特征，能够以图像特征的识别和分类来训练模型，并利用模型的准确率、特异性等多重评价标准来反向指导特征的提取[13]，最重要的深度学习的方法有：自动编码器(autoencoder, AE)、深度信念网络(deep belief network, DBN)、深度玻尔兹曼机(deep Boltzmann machines, DBM)以及卷积神经网络(convolutional neural network, CNN)，但是目前用于肺结节分类任务中的较多为 AE、DBN 和 CNN。当然，肺结节数据集的选择和最终分类的类别也很重要，这两点决定了最终实现的模型是否具有临床意义和价值。

4.1. 基于传统人工特征的肺结节分类研究

传统的图像分类需要人工提取特征，再将特征输入分类器中从而对不同类别的图像进行分类，处理流程如图 2 所示。Liu [14] 等通过提取肺部低剂量 CT 中 24 个放射学特征，当包含基于尺寸的特征时，利用线性分类器进行训练及验证，对于预测肺结节的恶性肿瘤最高准确率可达 74.3%。Samuel [15] 等在 CT 筛查时从不确定的肺结节中提取放射组学特征子集，利用分类器进行训练及验证，预测癌症或非癌症的后续发生的总体准确率为 80%。Samantha [16] 等提取肺部 CT 图像中结节定量特征、肺实质特征以及全局特征，通过统计分析以及逐步向前选择法对提取的特征进行选择，最后将最好的特征集利用人工神经网络(artificial neural network, ANN)，通过留一法，进行训练及验证，得到的结节良恶性分类最高的准确率可达 96%。Hugo [17] 等在分析不同恶性程度肿瘤的放射组学特征、临床数据和基因表达有何关系时，从肺部 CT 中提取四种放射组学特征，分别是肿瘤的大小、肿瘤的形状、肿瘤的纹理以及小波特征，将这四种特征与它们相应的临床数据和基因表达相对比，最终发现放射组学特征对肺癌的预后和验证都有很强的作用，并且放射组学特征与基因表达谱相关。Wang [18] 等使用 LIDC-IDRI 数据库，从 593 个病人的 CT 图像中提取了 150 个放射学特征，利用随机森林算法(random forest, RF)从 150 个特征中选择最终特征集，将最终 15 个具有预测价值的定量放射组学特征作为支持向量机(support vector machine, SVM)的输入，进行训练验证，最终在肺肿瘤的良恶性分类上可达到的最好的准确率为 76.1%。

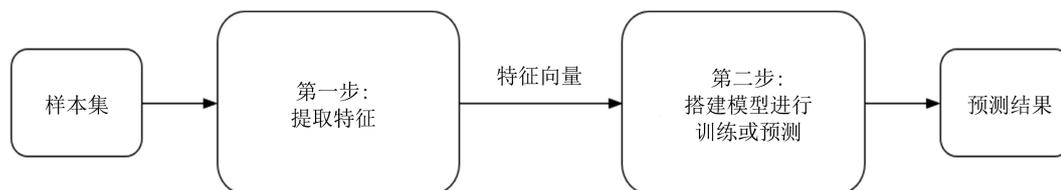


Figure 2. Processing flow of machine learning algorithm using artificial features

图 2. 采用人工特征的机器学习算法处理流程

4.2. 基于深度学习的肺结节分类研究

4.2.1. 基于 SDAE 的肺结节分类研究

自编码器是一种无监督学习的特殊神经网络，用于特征提取和数据降维。最简单的自动编码器由一个输入层、一个隐含层和一个输出层组成，如图 3 所示，该网络的前一部分是一个编码器 $h = f(x)$ ，后半部分是一个生成重构的解码器 $r = g(h)$ ，最后使得 x 约等于 $g(f(x))$ ，训练完成后，只使用编码器而不再需要解码器，编码器的输出结果被进一步使用。栈式自编码器(stacked AE, SAE)由多层 AE 堆叠构成，每一层都以前一层的表达特征为基础，抽取出更加抽象、更加适合的复杂特征。在 SAE 的输入层引入噪声，基于加噪数据来重构原始数据，可构成泛化能力更优，鲁棒性更高的去噪自动编码器(stack denoising AE, SDAE)。Chen [19]等人开发了三个多任务学习模型，从 LIDC-IDRI 数据集中随机选取 2400 个结节作为样本，对每个样本手工提取 Haar 特征和 Hog 特征，利用 SDAE 和 CNN 对每个样本提取深度特征，将四种特征进行异构作为多任务学习模型的输入，为了方便 SDAE 的输入，将样本尺寸从 28×28 的二维矩阵拉伸为 1×784 的一维矩阵，通过 SDAE 的无监督学习，最终将 100 维的矩阵作为每个样本的特征，SDAE 在此研究中作为一个特征提取器，其第一、二隐藏层对应的神经元个数分别为 200 和 100。

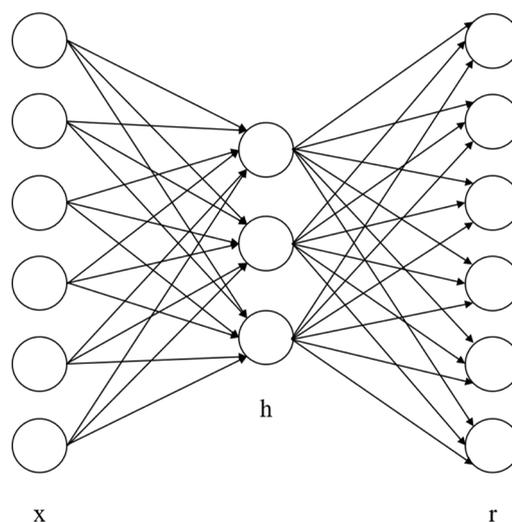


Figure 3. A simple autoencoder network
图 3. 一种简单的自动编码器网络

4.2.2. 基于 DBN 的肺结节分类研究

DBN 的单元为受限玻尔兹曼机(restricted Boltzmann machines, RBM)，DBN 的思想是把多个 RBM 串联堆叠起来形成一个更深的神经网络，它既可以作为一个自编码器用于非监督学习，也可以作为分类器用于监督学习。杨佳玲[20]等将 DBN 引入肺结节的良恶性诊断，选取 LIDC-IDRI 数据库中 2096 个结节作为训练样本，每个样本的特征集是由 4 个方面(形状、灰度、纹理和空间位置)提取的 87 个不同的特征构成，研究中 DBN 分类器由 3 个 RBM 组成，依次对三个 RBM 训练得到整个 DBN 的全局参数，通过大量实验，当 DBN 的隐含层数为 3，隐含层节点数依次为 90、120 和 90 个节点，RBM 训练学习率为 0.12 时网络可以达到最佳性能，在测试集上的分类准确率高达 95.3%。

4.2.3. 基于 CNN 的肺结节分类研究

卷积神经网络是一种监督学习的判别模型，一般是由输入层、卷积层、池化层、全连接层和输出层

构成, 具有局部链接、权值共享、下采样及挖掘数据局部特征, 对图像的平移、缩放和旋转等表现出较高的稳健性, 卷积层是由多个滤波器构成, 滤波器即指卷积核, 不同大小的卷积核可以提取不同的特征信息, 低层卷积易提取边缘和曲线底层特征, 高层卷积易提取抽象特征, 而肺结节的分类仅靠底层特征是难以分类的, 因此自从 CNN 问世以来, 众多肺结节的分类研究都会用到 CNN, 权值共享是为了防止过拟合、降低模型复杂度以及提高泛化能力, CNN 网络的收敛能力是评判此网络好坏的一个关键指数, 网络在训练时的过拟合、梯度消失以及梯度爆炸都会造成网络的收敛能力差, 因此研究者们就提出了三种解决办法, 分别是运用随机失活(dropout)方法, 批量归一化处理(batch normalization, BN), 以及用预处理的网络进行参数初始化加速学习的过程来增强网络的泛化能力。

在肺结节分类中, 常用的深度学习方法有二维卷积神经网络(two dimension convolutional neural network, 2D-CNN)、三维卷积神经网络(three dimension convolutional neural network, 3D-CNN)、多流多尺度卷积神经网络(multi-stream multi-scale convolutional neural network, MMCNN)。

Lecun [21] [22]等在 1989 年首次提出卷积神经网络, 由于当时计算能力有限以及可用于训练和测试的标记样本不足, 导致当时的卷积神经网络没有得到广泛的应用和发展。随着标记数据的增长和计算机计算能力的快速提升, Krizhevsky [23]等在 2012 年创新性地提出了卷积神经网络模型 AlexNet, 并且该模型在当年的 ImageNet 竞赛中获得第一名, 使卷积神经网络在图像领域得到了广泛的应用和快速的发展。之后, 不断有新的卷积神经网络被提出, 比如 VGG NET [24]、ResNet [25]、和 GooleNet [26]。因为 2D-CNN 在图像识别领域的优秀表现, 使得计算机视觉领域得到进一步的发展, 越来越多的科研人员将 2D-CNN 应用于肺结节的分类中。Hua [27]等在 2015 年将 2D-CNN 应用于肺结节的分类中, 解决了传统计算机辅助诊断系统性能不够稳定、由人工制定的诊断规则不够全面和无法持续提高准确率等问题。该方法的数据集来自于 LIDC-IDRI, 由 1010 名患者的已注释肺部 CT 图像组成, 从数据集中筛选出 2545 个直径大于 3 mm 的肺结节进行实验。最终, 该方法的灵敏度为 73.3%, 特异度为 78.8%, 工作效率高于传统的计算机辅助诊断系统。Rekka [28]等利用两个 CNN 网络分别提取特征, 再通过外矩阵乘积对每个图像、相同位置的特征进行合并, 利用 SVM 分类器进行肺结节与非结节的二分类, 最终得到的最好准确率可达 91.99%。吴世洋[29]等首先对 LIDC-IDRI 数据集中的肺部 CT 图像进行分割获得肺结节图像, 然后利用 CNN 进行特征提取, 将每张肺结节图像转化为一个 192 维的特征图, 最后利用 Logistic 分类器进行模型构建与测试, 取得了 84.4%的分类结果。Shen [30]等提出了一个多尺度卷积神经网络, 通过从不同尺度的肺结节图像中提取结节特征来捕获良恶性结节之间的异质性, 然后将提取的同一个结节的三种不同尺度的抽象特征进行融合, 输入到一个 SVM 或 RF 分类器来对结节进行分类, 准确率达 86.8%。实验结果证明, 深度学习在肺结节分类领域有广泛的发展前景。

但是, 2D-CNN 会丢失肺结节的立体信息, 未来解决此问题, Liu [31]等在 2017 年提出将 3D-CNN 应用于肺结节分类的想法, 3D-CNN 结构如图 4 所示, Liu [31]等设计了 2 个不同结构的 3D-CNN 在相同的数据集进行训练和测试, 通过对实验结果所生成的 ROC 曲线进行评估, 选出性能较好的网络结构。实验结果表明第 1 个 3D-CNN 的 AUC 为 0.732, 第 2 个 3D-CNN 的 AUC 为 0.698, 该实验继续设计 2D-CNN 用于对比 3D-CNN 的效果, 实验结果表明, 2D-CNN 的 AUC 为 0.688, 小于 3D-CNN 的 0.732, 故 3D-CNN 的效果更好。实验证明, 3D-CNN 的效果较好, 并且不需要依靠特定的专业知识, 通过不断学习便有助于系统的进一步完善, 提高了肺结节的诊断效率和准确性。但是 3D-CNN 存在特征较多、数据量较大和计算时间较长等问题。

Ciampi [32]等提出了一种应用于肺部检测的基于多流多尺度卷积神经网络的方法。此多流多尺度卷积神经网络不需要进行分割等预处理操作, 可以直接处理原始数据。相比于提取单一尺度特征的卷积神

神经网络, Ciompi [32]等通过分析任意数量二维界面图来学习肺结节的三维表示,能够提取到多尺度的特征,从而更加有利于肺结节的分类。Zheng [33]等提出了一种包含尺度转移模块和多特征融合运算的深度卷积神经网络 STM-Net。该网络可以放大小目标,适应不同分辨率的图像。评估数据来自复旦大学中山医院(ZSDB)提供的 CT 数据库,此数据集是肺腺癌数据集,所有资料均有病理标签。该方法对肺腺癌肺结节的精度和 AUC 分别为 95.455%和 0.987。Shen [34]等采用多裁剪卷积神经网络对卷积特征图中裁剪不同区域,来获得肺结节的主要信息,实验对 LIDC-IDRI 数据集进行了测试,良恶性分类准确率达到 87.14%, AUC 达到 0.93。

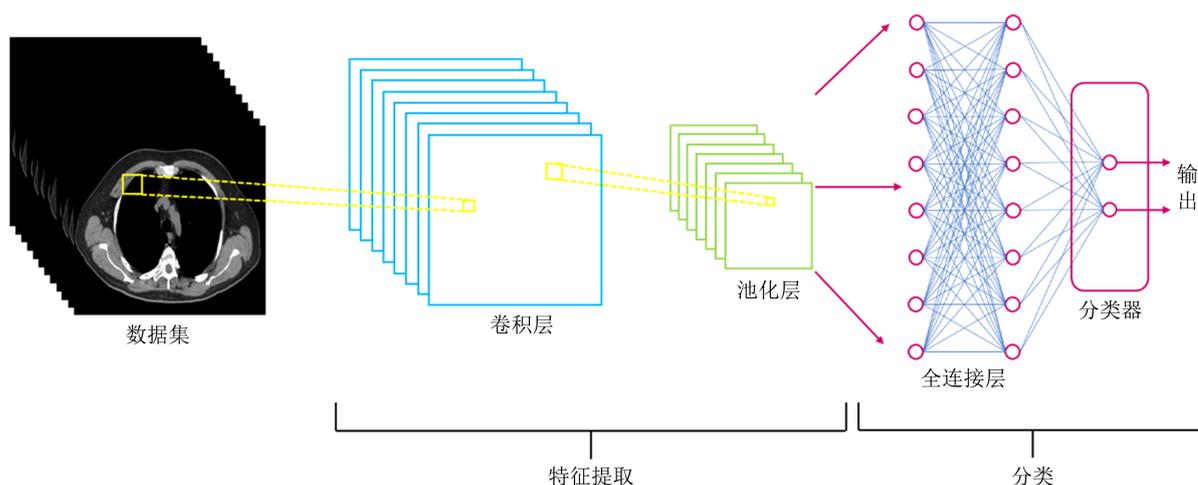


Figure 4. Structure diagram of 3D-CNN

图 4. 3D-CNN 结构图

4.3. 数据集的选择

目前利用公开数据库进行肺结节分类的研究,主要利用的是 LIDC-IDRI 数据库,但是良恶性评估的标签来自于这个数据库提供的两个数据:1) 利用 4 位胸腔放射科医生对肺结节良恶性的 5 个主观评估作为肺结节的良恶性等级评价,这个主观评价由可扩展标记语言(extensible markup language, XML)文件中提取,如图 5 所示;2) 利用数据集提供的 157 个病人的诊断结果作为肺结节的良恶性等级评价,这个评价是来自“tcia-diagnosis-data.csv”,如图 6 所示。国内外公认的诊断肺癌的“金标准”是细胞、组织学诊断方法,这种方法的诊断正确性优于其他任何诊断方法[35],所以将利用“金标准”诊断方法的数据用于研究将更具有临床意义。

刘一璟[36]等利用 LIDC-IDRI 数据集,将 4 位胸腔放射科医生对肺结节良恶性的 5 个主观评估作为肺结节的良恶性等级评价,这 5 个主观评估的分数为从 1 到 5 的整数,分数越高则肺结节为恶性越明显,而后分别为对每一个肺结节的医生良恶性评估进行平均,高于 3 的划为恶性,低于 3 的划为良性,等于 3 的划为不确定。最终得到 2637 个肺结节样本,其中 1361 个良性样本,640 个恶性样本,636 个不确定样本,不确定样本不用于实验。最后在提出的 DenseNet-centercrop-BC 网络中 AUC 分值最高可达到 0.9313。

Devinder [37]等利用 LIDC-IDRI 数据集提供的 157 个病人的诊断结果作为肺结节的良恶性等级评价,这 157 个病人的诊断结果来自于活检、外科手术、随访两年这三种手段,三个等级评价分别为良性、原发性恶性和转移性恶性,相对于放射科医生对肺结节的主观判断,诊断结果作为肺结节的良恶性等级评价更应该是金标准。而后,通过将提取的肺结节放射组学特征作为训练目标,在自动编码器上训练及测试,最终获得了 75.01%的准确性。

```
<?xml version="1.0" encoding="UTF-8"?>
- <LidcReadMessage xmlns="http://www.nih.gov" uid="1.3.6.1.4.1.14519.5.2.1.6279.6001.1308168927827.0">
  - <ResponseHeader>
    <Version>1.7</Version>
    <MessageId>1151420</MessageId>
    <DateRequest>2006-06-05</DateRequest>
    <TimeRequest>17:05:26</TimeRequest>
    <RequestingSite>removed</RequestingSite>
    <ServicingSite>removed</ServicingSite>
    <TaskDescription>Second unblinded read</TaskDescription>
    <CtImageFile>removed</CtImageFile>
    <SeriesInstanceUid>1.3.6.1.4.1.14519.5.2.1.6279.6001.303494235102183795724852353824</SeriesInstanceUid>
    <StudyInstanceUID>1.3.6.1.4.1.14519.5.2.1.6279.6001.339170810277323131167631068432</StudyInstanceUID>
    <DateService>2006-06-05</DateService>
    <TimeService>17:05:26</TimeService>
    <ResponseDescription>1 - Reading complete</ResponseDescription>
    <ResponseComments>Merged, reader anonymized, unblinded responses</ResponseComments>
  </ResponseHeader>
  - <readingSession>
    <annotationVersion>3.12</annotationVersion>
    <servicingRadiologistID>anonymous</servicingRadiologistID>
    + <unblindedReadNodule>
      - <unblindedReadNodule>
        <noduleID>3</noduleID>
        - <characteristics>
          <subtlety>5</subtlety>
          <internalStructure>1</internalStructure>
          <calcification>6</calcification>
          <sphericity>3</sphericity>
          <margin>4</margin>
          <lobulation>1</lobulation>
          <spiculation>1</spiculation>
          <texture>5</texture>
          <malignancy>3</malignancy>
        </characteristics>
        - <roi>
          <imageZposition>1550.5</imageZposition>
          <imageSOP_UID>1.3.6.1.4.1.14519.5.2.1.6279.6001.498089111228175009979138545584</imageSOP_UID>
```

Figure 5. Part of the XML file

图 5. 部分 XML 文件

	A	B	C	D	E	F
	TCIA Patient ID	Diagnosis at the Patient Level 0=Unknown 1=benign or non-malignant	Diagnosis Method 0 = unknown 1 = review of radiological images to show	Primary tumor site for metastatic disease	Nodule 1 Diagnosis at the Nodule Level 0=Unknown 1=benign or non-malignant	Nodule 1 Diagnosis Method at the Nodule Level 0 = unknown 1 = review of
1						
2	LIDC-IDRI-0068	3	4	Head & Neck Cancer	3	4
3	LIDC-IDRI-0071	3	1	Head & Neck	1	1
4	LIDC-IDRI-0072	2	4	Lung Cancer	1	4
5	LIDC-IDRI-0088	3	0	Uterine Cancer	0	0
6	LIDC-IDRI-0090	2	3	NSCLC	2	3
7	LIDC-IDRI-0091	3	4	urothelial carcinoma	3	4
8	LIDC-IDRI-0100	3	1	Testis	1	1
9	LIDC-IDRI-0118	3	0	Prostate	0	0

Figure 6. Part of the “tcia-diagnostic-data.csv” file

图 6. 部分 “tcia-diagnosis-data.csv” 文件

4.4. 肺结节多分类

一般对于肺结节的分类研究，研究者们选择将肺结节分为两类，即良性与恶性，很少有研究者们选择对肺结节恶性再进行分类的研究，目前多分类研究有对肺结节的恶性程度进行等级分类，但是对于恶性结节的种类进行分类，例如原发性恶性和转移性恶性，在临床上是个难题，这两个诊断相当重要，极大影响了患者肺癌分期、后续治疗手段选择，甚至决定了患者能否治愈或预后。

Kang [38]等利用 LIDC-IDRI 数据集提供的 157 个病人的诊断结果作为肺结节的良恶性等级评价，采用三维多视图卷积神经网络(MV-CNN)来探讨肺结节的分类，对肺结节进行二分类(良性和恶性)和三元分类(良性、原发恶性和转移恶性)，最后的二分类错误率为 4.59%，三元分类错误率为 7.70%。

5. 讨论与展望

在上述研究中, 通过研究者设计特征提取算法实现的肺结节分类任务取得了不错的结果, 相对于深度学习方法提取特征更有针对性, 但是目前利用人工提取的特征还存在一些局限性: 1) 在放射组学中, 非标准化获取的数据集会增加特征描述噪声, 影响预测的准确性; 2) 提取放射组学特征时, 需要对图像进行分割, 分割算法的不同, 会造成后续提取特征的差异性; 3) 利用 CT 图像的底层特征, 如灰度特征、纹理特征、几何特征等, 这些特征一般是人工设计的, 存在着主观差异性; 4) 需要考虑到临床中系统水平的变化, 扫描仪的参数设置和操作员的精确度都需要考虑到。

深度学习模型相对于基于人工特征的模型受主观因素影响更少, 获取特征更便捷。上述研究中的 SADE 是多个 AE 堆叠而成的, 可以获得更抽象的特征, 由于加了噪声, 训练后的模型对噪声信息不敏感, 泛化能力更好, 鲁棒性能更强, 但是相对于 CNN 模型, SADE 的输入对维度有一定的要求, 需要将肺结节图像通过维度变换降为一维数据, 才能进一步提取肺结节特征或者进行分类。Sun [39]等设计并实现了三种基于多通道 ROI 的深度结构算法: 卷积神经网络(CNN)、深度信念网络(DBN)和堆叠去噪自编码器(SDAE), 数据来自于 LIDC-IDRI 数据库 1018 例病例, 对于每个样本的感兴趣区域(region of interest, ROI), 生成三幅图像, 分别是包含结节及其周围区域的原始 ROI 图像、仅包含结节的 ROI 和梯度图像, 最后将三幅图像组合生成一个 RGB 图像作为每个样本最终的 ROI, 通过比较三个模型的性能, CNN 的曲线下最高面积(AUC)为 0.899 ± 0.018 , 明显高于传统 CADx 的 $AUC = 0.848 \pm 0.026$ 。DBN 的结果也略高于 CADx, 而 SDAE 略低于 CADx。由此说明, 基于 CNN 的架构模型相对于 DBN 和 SADE 具有不可替代的优势。由于肺部 CT 可以展示肺部的空间形态信息, 且肺结节是一个立体的组织, 所以获得空间信息更加的重要, 2D-CNN 与 3D-CNN 相比较, 不能够充分地学习肺结节的空间信息, 目前 3D-CNN 具有提取的特征多, 训练时产生的参数多等弊端, 未来对于 3D-CNN 的研究应该从优化参数等方向着手改善算法, 这样才可以在减少参数量的同时不至于丢失需要的立体结构信息。Transformer 目前在计算机视觉领域掀起一阵浪潮, 由于其在自然语言处理领域强大的表示能力, 许多研究者将其扩展到计算机视觉任务中, 将 Transformer 与 CNN 进行比较, 目前在图像分类方面, Generative Pretraining from Pixel (iGPT) [40]和 Vision Transformer (ViT) [41]是使用 Transformer 进行图像分类的两个模型, 在 Chen [40]等人的研究中发现 Transformer 更适用于大型数据集, 若能够将 Transformer 结构的网络在大规模图像数据集上进行预训练, 然后迁移至目标数据集中, 效果可以与当前最好的 CNN 算法的网络相媲美。Dosovitskiy [41]等人在研究中发现 Transformer 相比 CNN 缺少一些归纳偏差, 例如平移等变和局部性, 但在数据集足够的情况下, 泛化能力同样也是可以达到很强。若要利用 Transformer 进行图像的分类, 那么在图像预处理阶段, 需要对图像进行维度的转换, 相对于 CNN 网络可以直接将图像作为输入相比较, Transformer 更复杂, 但是 Transformer 相较于 CNN 其更关注全局信息, 能建模更加长距离的依赖关系, 而 CNN 较关注局部信息, 全局信息的捕捉能力弱。Transformer 由于缺少 CNN 的归纳偏差问题的同时, 同样也避免了 CNN 中存在的归纳偏差等问题。在肺结节分类的任务中, 未来可以尝试去使用 Transformer 算法, 先在大规模数据集中进行预训练, 再迁移至肺部 CT 图像数据集, 如果想要直接在目标数据集中使用 Transformer 算法, 就要求大型的肺部 CT 图像数据集的出现, 所以若想要在目前存在的肺部 CT 图像数据集上利用 Transformer 算法进行肺结节分类研究, 还需要进行一定的预处理。

在进行肺结节分类研究时, 数据的选择也是至关重要的, 选择具有“金标准”的数据比医生主观评估的数据更具与临床意义。对于区分原发性肺癌和转移性肺癌, 是一个临床难题, 但是这两个诊断又相当重要, 极大影响了患者肺癌分期、后续治疗手段选择, 也影响了患者预后处理选择, 因此肺结节的多种分类具有必要性。

综上所述我们认为在未来利用具有金标准的肺结节 CT 图像数据, 运用 3D-CNN 提取的特征和人工提取的肺结节特征相结合的方法, 对肺结节实现多分类将具有更好的发展前景。

致 谢

我们承认国家癌症研究所和国家卫生研究院基金会在创建免费公开的 LIDC-IDRI 中扮演的关键角色, 使得科研者们对于肺结节的研究更便捷化。我们特别感谢 7 个学术中心和 8 个医学成像公司用于协作和创建数据库。最后, 我们感谢编辑和匿名审稿人对于本文章的帮助。

基金项目

国家自然科学基金项目(61931013, 81701644, 11832003); 重点研发项目(2017YFC0111104)。

参考文献

- [1] Siegel, R.L., Miller, K.D. and Jemal, A. (2020) Cancer Statistics, 2020. *CA: A Cancer Journal for Clinicians*, **70**, 7-30. <https://doi.org/10.3322/caac.21590>
- [2] World Health Organization (2020) World Health Statistics 2020. World Health Organization, Geneva.
- [3] Armato III, S.G., Roberts, R.Y., Meyer, C.R., *et al.* (2007) The Lung Image Database Consortium (LIDC): Ensuring the Integrity of Expert-Defined “Truth”. *Academic Radiology*, **14**, 1455-1463. <https://doi.org/10.1016/j.acra.2007.08.006>
- [4] 刘颖, 赖敏. 早期肺癌影像学诊断研究进展[J]. 影像研究与医学应用, 2019, 3(1): 10-11.
- [5] 贾群玲. CT 与 MRI 诊断孤立性肺结节良恶性的准确性分析[J]. 中国 CT 与 MRI 杂志, 2016, 14(10): 42-45.
- [6] Patel, V.K., Naik, S.K., Naidich, D.P., *et al.* (2013) A Practical Algorithmic Approach to the Diagnosis and Management of Solitary Pulmonary Nodules. *Chest*, **143**, 840-846. <https://doi.org/10.1378/chest.12-1487>
- [7] National Lung Screening Trial Research Team (2011) The National Lung Screening Trial: Overview and Study Design. *Radiology*, **258**, 243-253. <https://doi.org/10.1148/radiol.10091808>
- [8] Naidich, D.P., Marshall, C.H., Gribbin, C., *et al.* (1990) Low-Dose CT of the Lungs: Preliminary Observations. *Radiology*, **175**, 729-731. <https://doi.org/10.1148/radiology.175.3.2343122>
- [9] Armato III, S.G., McLennan, G., Bidaut, L., *et al.* (2011) The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans. *Medical Physics*, **38**, 915-931. <https://doi.org/10.1118/1.3528204>
- [10] Setio, A.A.A., Traverso, A., De Bel, T., *et al.* (2017) Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: The LUNA16 Challenge. *Medical Image Analysis*, **42**, 1-13. <https://doi.org/10.1016/j.media.2017.06.015>
- [11] 王婧璇, 林岚, 赵思远, 等. 基于深度学习的肺结节计算机断层扫描影像检测与分类的研究进展[J]. 生物医学工程学杂志, 2019, 36(4): 670-676.
- [12] Way, T.W., Sahiner, B., Chan, H.-P., *et al.* (2009) Computer-Aided Diagnosis of Pulmonary Nodules on CT Scans: Improvement of Classification Performance with Nodule Surface Features. *Medical Physics*, **36**, 3086-3098. <https://doi.org/10.1118/1.3140589>
- [13] Anirudh, R., Thiagarajan, J.J., Bremer, T. and Kim, H. (2016) Lung Nodule Detection Using 3D Convolutional Neural Networks Trained on Weakly Labeled Data. *Medical Imaging 2016: Computer-Aided Diagnosis. International Society for Optics and Photonics*, **9785**, 978532. <https://doi.org/10.1117/12.2214876>
- [14] Liu, Y., Balagurunathan, Y., Atwater, T., *et al.* (2017) Radiological Image Traits Predictive of Cancer Status in Pulmonary Nodules. *Clinical Cancer Research*, **23**, 1442-1449. <https://doi.org/10.1158/1078-0432.CCR-15-3102>
- [15] Hawkins, S., Wang, H., Liu, Y., *et al.* (2016) Predicting Malignant Nodules from Screening CT Scans. *Journal of Thoracic Oncology*, **11**, 2120-2128. <https://doi.org/10.1016/j.jtho.2016.07.002>
- [16] Dilger, S.K., Uthoff, J., Judisch, A., *et al.* (2015) Improved Pulmonary Nodule Classification Utilizing Quantitative Lung Parenchyma Features. *Journal of Medical Imaging*, **2**, 041004. <https://doi.org/10.1117/1.JMI.2.4.041004>
- [17] Aerts, H.J.W.L., Velazquez, E.R., Leijenaar, R.T.H., *et al.* (2014) Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nature Communications*, **5**, Article No. 4006. <https://doi.org/10.1038/ncomms5006>

- [18] Wang, J., Liu, X., Dong, D., *et al.* (2016) Prediction of Malignant and Benign of Lung Tumor Using a Quantitative Radiomic Method. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, 16-20 August 2016, 1272-1275. <https://doi.org/10.1109/EMBC.2016.7590938>
- [19] Chen, S., Qin, J., Ji, X., *et al.* (2017) Automatic Scoring of Multiple Semantic Attributes with Multi-Task Feature Leverage: A Study on Pulmonary Nodules in CT Images. *IEEE Transactions on Medical Imaging*, **36**, 802-814. <https://doi.org/10.1109/TMI.2016.2629462>
- [20] 杨佳玲, 赵涓涓, 强彦, 等. 基于深度信念网络的肺结节良恶性分类[J]. 科学技术与工程, 2016, 16(32): 69-74.
- [21] LeCun, Y., Boser, B., Denker, J.S., *et al.* (1989) Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, **1**, 541-551. <https://doi.org/10.1162/neco.1989.1.4.541>
- [22] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [23] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
- [24] Donahue, J., Anne Hendricks, L., Guadarrama, S., *et al.* (2015) Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 2625-2634. <https://doi.org/10.1109/CVPR.2015.7298878>
- [25] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [26] Szegedy, C., Liu, W., Jia, Y., *et al.* (2015) Going Deeper with Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [27] Hua, K.L., Hsu, C.H., Hidayati, S.C., *et al.* (2015) Computer-Aided Classification of Lung Nodules on Computed Tomography Images via Deep Learning Technique. *OncoTargets and Therapy*, **8**, 2015-2022. <https://doi.org/10.2147/OTT.S80733>
- [28] Mastouri, R., Khelifa, N., Neji, H. and Hantous-Zannad, S. (2020) A Bilinear Convolutional Neural Network for Lung Nodules Classification on CT Images. *International Journal of Computer Assisted Radiology and Surgery*, **16**, 91-101. <https://doi.org/10.1007/s11548-020-02283-z>
- [29] 吴世洋, 任劲松, 张冉, 等. 基于卷积神经网络的肺结节良恶性分类[J]. 中国医学工程, 2020, 28(1): 1-3.
- [30] Shen, W., Zhou, M., Yang, F., *et al.* (2015) Multi-Scale Convolutional Neural Networks for Lung Nodule Classification. In: *International Conference on Information Processing in Medical Imaging*, Springer, Cham, 588-599. https://doi.org/10.1007/978-3-319-19992-4_46
- [31] Liu, S., Xie, Y., Jirapatnakul, A. and Reeves, A.P. (2017) Pulmonary Nodule Classification in Lung Cancer Screening with Three-Dimensional Convolutional Neural Networks. *Journal of Medical Imaging*, **4**, 041308. <https://doi.org/10.1117/1.JMI.4.4.041308>
- [32] Ciompi, F., Chung, K., Van Riel, S.J., *et al.* (2017) Towards Automatic Pulmonary Nodule Management in Lung Cancer Screening with Deep Learning. *Scientific Reports*, **7**, Article No. 46479. <https://doi.org/10.1038/srep46878>
- [33] Zheng, J., Yang, D., Zhu, Y., *et al.* (2020) Pulmonary Nodule Risk Classification in Adenocarcinoma from CT Images Using Deep CNN with Scale Transfer Module. *IET Image Processing*, **14**, 1481-1489. <https://doi.org/10.1049/iet-ipr.2019.0248>
- [34] Shen, W., Zhou, M., Yang, F., *et al.* (2017) Multi-Crop Convolutional Neural Networks for Lung Nodule Malignancy Suspiciousness Classification. *Pattern Recognition*, **61**, 663-673. <https://doi.org/10.1016/j.patcog.2016.05.029>
- [35] 王强修, 李钧, 朱良明. 肺癌诊断与治疗[M]. 北京: 人民军医出版社, 2013.
- [36] 刘一璟, 张旭斌, 张建伟, 等. DenseNet-centercrop: 一个用于肺结节分类的卷积网络[J]. 浙江大学学报(理学版), 2020, 47(1): 20-26.
- [37] Kumar, D., Wong, A. and Clausi, D.A. (2015) Lung Nodule Classification Using Deep Features in CT Images. 2015 12th Conference on Computer and Robot Vision, Halifax, 3-5 June 2015, 133-138. <https://doi.org/10.1109/CRV.2015.25>
- [38] Kang, G., Liu, K., Hou, B. and Zhang, N. (2017) 3D Multi-View Convolutional Neural Networks for Lung Nodule Classification. *PLoS ONE*, **12**, e0188290. <https://doi.org/10.1371/journal.pone.0188290>
- [39] Sun, W., Zheng, B. and Qian, W. (2017) Automatic Feature Learning Using Multichannel ROI Based on Deep Structured Algorithms for Computerized Lung Cancer Diagnosis. *Computers in Biology and Medicine*, **89**, 530-539. <https://doi.org/10.1016/j.compbiomed.2017.04.006>
- [40] Chen, M., Radford, A., Child, R., *et al.* (2020) Generative Pretraining from Pixels. *Proceedings of the 37th Interna-*

tional Conference on Machine Learning, **119**, 1691-1703.

- [41] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2020) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929.

附录

Table A1. A list of literature on classification of pulmonary nodules based on traditional artificial features
附表 1. 基于传统人工特征的肺结节分类文献列表

参考文献	数据来源	数据预处理	分类器	分析结果
Liu <i>et al.</i> [14]	私人数据集	提取 24 个放射学特征 (包含基于尺寸的特征)	线性分类器	肺结节恶性肿瘤分类准确率达 81%
Samuel <i>et al.</i> [15]	国家肺部筛查试验 (ACRIN 6684) 公开数据	提取了 219 个 3D 图像特征, 从中选取 23 个稳定特征	随机森林分类器	预测 1 年和 2 年后会癌变的结节, 准确率分别为 80% 和 79%
Samantha <i>et al.</i> [16]	14 例来自 NLST 爱荷华州队列; 另外的 36 例病例来自 COPDGene 研究	对 CT 图像进行结节分割和实质分割, 然后对分割后的图像提取结节定量特征、肺实质特征以及全局特征	人工神经网络	肺结节良恶性分类准确率 96%
Hugo <i>et al.</i> [17]	Lung1, Lung3, H&N1, Lung2 和 H&N2	提取肿瘤的形状、肿瘤的形状、肿瘤的纹理以及小波特征四个方面的特征	-	放射组学特征对肺癌的预后和验证都有很强的作用
Wang <i>et al.</i> [18]	LIDC-IDRI	提取 15 个具有预测价值的放射学特征	支持向量机	肺肿瘤的良恶性分类达到的最好的准确率为 76.1%

Table A2. A list of literature on classification of pulmonary nodules based on deep learning
附表 2. 基于深度学习的肺结节分类文献列表

参考文献	数据来源	数据预处理	算法	分析结果
Chen <i>et al.</i> [19]	LIDC-IDRI	提取 Haar 特征和 Hog 特征, 利用 SDAE 和 CNN 对每个样本提取深度特征, 将四种特征进行异构	SADE、CNN	研究提出的语义属性评分方案可以提供更丰富的结核定量评估
杨佳玲等[20]	LIDC-IDRI	提取 87 个不同的特征, 它们分别来自于形状、灰度、纹理和空间位置这四个方面	DBN	肺结节良恶性分类准确率高达 95.3%
Hua <i>et al.</i> [27]	LIDC-IDRI	选择直径大于 3 mm 的肺结节进行实验	2D-CNN	肺结节良恶性分类灵敏度为 73.3%, 特异度为 78.8%
Rekka <i>et al.</i> [28]	LUNA 16	利用两个 CNN 网络分别提取特征, 再通过外矩阵乘积对每个图像、相同位置的特征进行合并	2D-CNN	肺结节与非结节的分类准确率可达 91.99%
吴世洋等[29]	LIDC-IDRI	肺部 CT 图像进行分割获得肺结节图像, 利用 CNN 进行特征提取, 将每张肺结节图像转化为一个 192 维的特征图	2D-CNN	肺结节良恶性分类准确率为 84.4%
Shen <i>et al.</i> [30]	LIDC-IDRI	从不同尺度的肺结节图像中提取结节特征来捕获良恶性结节之间的异质性, 然后将提取的同一个结节的三种不同尺度的抽象特征进行融合	Multi-scale-CNN	肺结节良恶性分类准确率达 86.8%
Liu <i>et al.</i> [31]	肺癌筛查研究(国家肺癌筛查试验(NLST)) 40 和早期肺癌行动计划(ELCAP)39	以结节为中心进行裁剪, 对裁剪后的三维图像区域进行重新采样, 得到各向同性的固定大小的三维图像	3D-CNN	肺结节良恶性分类的 AUC 值为 0.732
Ciampi <i>et al.</i> [32]	意大利 MILD 筛查实验和丹麦 DLCST 筛查实验	直接处理原始数据	多流多尺度卷积神经网络	通过分析任意数量二维界面图来学习肺结节的三维表示, 能够提取到多尺度的特征, 从而更加有利于肺结节的分类
Zheng <i>et al.</i> [33]	复旦大学中山医院(ZSDB)提供的 CT 数据库	对肺结节图像进行分割	包含尺度转移模块和多特征融合运算的深度卷积神经网络 STM-Net	肺腺癌肺结节的精度和 AUC 分别为 95.455% 和 0.987
Shen <i>et al.</i> [34]	LIDC-IDRI	采用多裁剪卷积神经网络对卷积特征图中裁剪不同区域	Multi-crop-CNN	肺良恶性分类准确率达 87.14%