

## 基于CT征象量化分析的肺结节恶性度分级

陈皓<sup>\*①②</sup> 段红柏<sup>①</sup> 郭紫园<sup>①</sup> 强永乾<sup>③</sup>

<sup>①</sup>(西安邮电大学计算机学院 西安 710121)

<sup>②</sup>(陕西省网络数据分析与智能处理重点实验室 西安 710121)

<sup>③</sup>(西安交通大学第一附属医院 西安 710061)

**摘要:** 为了提高肺结节恶性度分级的计算精度及可解释性, 该文提出一种基于CT征象量化分析的肺结节恶性度分级方法。首先, 融合影像组学特征和通过卷积神经网络提取的高阶特征构造分析CT征象所需的特征集; 接着, 在混合特征集的基础上利用进化搜索机制优化集成学习分类器, 实现对7种肺结节征象的识别和量化打分; 最后, 将7种CT征象的量化打分输入到一个利用差分进化算法优化产生的多分类器, 实现肺结节恶性度的分级计算。在实验研究中使用LIDC-IDRI数据集中的2000个肺结节样本进行进化集成学习器和恶性度分级器的训练和测试。实验结果显示对7种CT征象的识别准确率可达0.9642以上, 肺结节恶性度分级的准确率为0.8618, 精确率为0.8678, 召回率为0.8617, F1指标为0.8627。与多个典型算法的比较显示, 该文方法不但具有较高的准确率, 而且可对相关CT征象进行量化分析, 使得对恶性度的分级结果更具可解释性。

**关键词:** 肺恶性度分级; CT征象; 进化集成学习; 量化分析; 可解释性

中图分类号: TN911.73; TP391.41

文献标识码: A

文章编号: 1009-5896(2021)05-1405-09

DOI: 10.11999/JEIT200167

## Malignancy Grading of Lung Nodules Based on CT Signs Quantization Analysis

CHEN Hao<sup>①②</sup> DUAN Hongbai<sup>①</sup> GUO Ziyuan<sup>①</sup> QIANG Yongqian<sup>③</sup>

<sup>①</sup>(School of Computer, Xi'an University of Posts & Telecommunications, Xi'an 710121, China)

<sup>②</sup>(Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Post and Telecommunications, Xi'an 710121, China)

<sup>③</sup>(First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China)

**Abstract:** In order to improve the accuracy and interpretability of the grading of malignant nodules in the lung, a method is proposed to achieve grading automatically for lung nodules by using (Computed Tomography, CT) signs. Firstly, features sets are extracted of CT signs by combing the radiomics features with the higher-order features extracted by convolutional neural network. Then, the ensemble classifier is optimized by the evolutionary search mechanism based on the mixed feature sets, and it is used to realize quantitative scores for 7 CT signs. Finally, 7 quantitative scores are input to the optimized multi-classifier to achieve the grading of malignant nodules in the lung. In the experience, 2000 samples of lung nodules in LIDC-IDRI data set are used to train and test the proposed method. The results show that the recognition accuracy of the 7 CT signs can reach more than 0.9642, the grading accuracy reaches 0.8618, the precision reaches 0.8678, the recall reaches 0.8617, and the F1 index reaches 0.8627. With respect to typical algorithms, the proposed method not only has high accuracy, but also can quantitatively analyze the CT signs that make the grade result of malignancy more interpretive.

**Key words:** Lung malignancy grading; Computed Tomography (CT) signs; Evolutionary ensemble learning; Quantization analysis; Interpretability

收稿日期: 2020-03-13; 改回日期: 2020-09-25; 网络出版: 2020-10-16

\*通信作者: 陈皓 chenhao@xupt.edu.cn

基金项目: 国家自然科学基金(61876138, 61203311), 陕西省自然科学基金(2019JM-365), 陕西省教育厅自然科学专项(17JK0701), 陕西省网络数据分析与智能处理重点实验室开放课题基金(XUPT-KLND(201804)), 西安邮电大学创新基金(CXJJLI2018017)

Foundation Items: The National Natural Science Foundation of China (61876138, 61203311), The Natural Science Basic Research Program of Shaanxi Province (2019JM-365), The Scientific Research Program Funded by Shaanxi Provincial Education Department (17JK0701), The Science Foundation of the Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing (XUPT-KLND (201804)), The Innovation Funds of Xi'an University of Posts and Telecommunications (CXJJLI2018017)

## 1 引言

基于电子计算机断层扫描技术(Computed Tomography, CT)的影像学检查为肺内结节恶性度早期诊断和监测提供了基础<sup>[1]</sup>。正如多个医学协会的诊断指南中所述<sup>[2-4]</sup>, 肺结节影像的纹理和形态等征象与肺结节的恶性度分化有很大的关系, 而钙化、圆度和边缘清楚度等征象也有助于评估结节的恶性度。但由于人眼能够有效识别的灰阶信息仅占CT影像所提供信息的很小一部分, 且许多小结节是亚厘米结节, 同时考虑到医生水平的差异性, 故以人工分析为主的工作模式使得对肺结节的识别分析结果存在着偏差。利用计算机视觉技术辅助医师高效地获得肺内结节相关语义征象的量化结果, 并根据征象的量化结果对肺结节的恶性度进行分级, 不仅能提高医生的工作效率, 同时有助于提高诊断的准确率和可解释性。

传统的肺结节恶性度分级方法包括支持向量机(Support Vector Machine, SVM), K近邻等方法。如Rodrigues等人<sup>[5]</sup>通过结构共线矩阵提取肺结节的特征并使用多层感知机、SVM和K近邻3个分类器进行良恶性分类。Nobrega等人<sup>[6]</sup>使用朴素贝叶斯、多层感知器、SVM和K近邻进行良恶性的分类。传统的恶性度分级方法在结果可解释性上存在一定的优势, 方便进一步分析分类模型, 但在分类性能上并不理想。随着深度学习技术的发展, 卷积神经网络(Convolution Neural Network, CNN)在肺结节恶性度分类方面被广泛应用。如Zuo等人<sup>[7]</sup>提出一种多分辨率CNN从不同深度层中提取不同级别和分辨率的特征构建良恶性分类网络。Shen等人<sup>[8]</sup>通过CNN得到不同语义征象的表示形式, 然后组合起来用于推断肺结节良恶性分类结果。Wang等人<sup>[9]</sup>使用CNN模型中的3种模型对肺结节良恶性分类性能进行比较分析, 通过比较发现多通道CNN模型对结节的恶性肿瘤风险的分类能力最好。CNN方法可以做到端到端的学习, 但此类方法包含大量参数, 需要较大的内存和计算资源开销。同时深度学习算法属于“黑盒”, 无法进一步理解和分析模型是如何进行分类决策的。近年来一些集成学习方法, 如随机森林(Random Forest, RF)、极端梯度提升(eXtreme Gradient Boosting, XGBoost)和自适应增强(Adaptive Boosting, AdaBoost)等也得到了较多应用<sup>[10]</sup>。如Nishio等人<sup>[11]</sup>使用XGBoost进行肺结节的良恶性分类。吴艇帆等人<sup>[12]</sup>针对肺腺癌磨玻璃结节患者运用逻辑回归、人工神经网络、SVM和AdaBoost等4种经典算法, 进行模型性能的分析 and 比较。集成学习方法相比于传

统的方法在性能上得到了明显提高, 相对CNN方法在对结果进一步解释分析方面更具优势, 但此类方法的性能与训练方法及所采用的特征集结构有着直接关系。在特征提取方面。现有从CT中提取特征的方法大都以影像组学特征为主, 主要包括灰度统计特征、纹理特征、形状特征和三者的组合特征等。如Ferreira等人<sup>[13]</sup>通过共生矩阵提取纹理特征并从垂直线像素的灰度强度数组中提取边缘清晰度特征, 根据纹理和边缘清晰度的图像特征对肺结节进行良恶性分类。Wu等人<sup>[14]</sup>从形状、灰度和纹理方面提取特征。通过随机森林对肺结节良恶性进行分类。Shaukat等人<sup>[15]</sup>提取了图像纹理、2D和3D形状以及强度特征, 然后通过两层前馈神经网络实现良恶性分类。影像组学特征能够很好地对图像中的浅层信息进行量化, 并且具有较好的可解释性。但它不足以表达图像的全局信息, 同时过滤噪声能力较弱。此外提取影像组学特征时, 对包含伪影、空腔和骨头等结构信息需要进行特殊处理<sup>[16]</sup>。另一方面CNN作为一个性能优良的特征提取器, 能够很好地提取出图像全局性的高阶特征。常用的CNN特征包括深层卷积特征或浅层卷积特征。如Shi等人<sup>[17]</sup>采用VGG-16提取肺结节特征, 使用SVM分类器进行良恶性分类。Nobrega等人<sup>[6]</sup>通过使用VGG-16, VGG-9和MobileNet等网络提取肺结节的深层卷积特征, 然后使用朴素贝叶斯、多层感知器、SVM和K近邻进行良恶性的分类。CNN可以通过多层非线性机制获得高阶抽象特征。但是通过卷积得到的特征图分辨率较低, 且在卷积过程中容易丢失原图中很多底层的细节信息。上述两种特征提取方式都有其优势和局限性, 可结合两种特征提取方法来构造出更全面的特征以表达CT影像的信息。为了提高计算性能, 针对集成学习器的生成优化方法也开展了许多研究工作<sup>[18]</sup>。本文以随机森林为例进行分析, 如Elyan等人<sup>[19]</sup>将类分解的细粒度随机森林应用在医学诊断中。对原始数据集通过K-Means聚类可将 $n$ 个标签聚成 $K \times N$ 个标签, 增强数据标签的多样性以此来提高随机森林的性能。虽然在一定程度上提高了随机森林分类器的性能, 但只是通过对数据集的处理来生成更好的决策树, 对于决策树之间的结构组合并没有进行深入挖掘。如Miao等人<sup>[20]</sup>使用改进的随机森林对心力衰竭患者的死亡率进行预测建模。利用新颖的分裂规则和终止标准, 能识别更准确的预测因子以区分幸存者和非幸存者, 从而提高随机森林的性能。Paul等人<sup>[21]</sup>在构造每个节点的过程中, 不断更新特征的重要性功能列表。并且针对森林中的树木数量制定了

新的理论上限，最终收敛于一组很重要的特征。与传统的随机森林相比，该算法在检测乳腺组织病理的分类准确率上有显著提高。上述优化方法整体是一个正向的加法过程，缺乏逆向的反馈过程仅依靠提升单棵决策树的性能来改进集成学习器的整体准确率，并没有从整体上考虑不同决策树组合对性能的影响。

本文实现了对CT征象的量化分析和肺结节恶性度分级计算的方法。主要贡献包括：(1) 为了获得更丰富的影像信息，提出融合影像组学特征和通过CNN获得的高阶特征共同构造CT征象分析所需的特征集。(2) 为了充分利用语义标签，提出通过进化集成学习方法对7种CT征象进行量化分析。这为肺结节恶性度分级的计算提供了基础，同时也进一步提高了肺结节恶性度分级结果的可解释性。(3) 使用通过差分进化算法优化的多分类器来提高肺结节恶性度分级的准确率。这种方法具有结构简单，可解释性更佳等优点。本文第2节详细介绍本

文所提出的方法；第3节进行实验及讨论；第4节对本文工作进行了总结。

## 2 本文方法

本文使用7种语义征象开展研究工作，分别为钙化征象、球形度征象、边缘征象、分页征象、毛刺征象、纹理征象和精细度征象。整体思路为：首先，构造信息更为丰富的特征集；其次，针对7种征象，通过以RF为个体的群体进化搜索过程构造不同征象的量化评分器；最后，将7种征象的量化结果作为特征输入到恶性度分级模型计算分级结果。图1为此过程的流程图。

### 2.1 特征集

为了获得针对7种征象进行描述所需的更丰富的特征集，本文提出融合影像组学特征和CNN特征来组成特征向量。具体流程如图2所示。首先，使用自适应阈值对肺结节进行分割，再采用影像组学方法提取肺结节影像的低阶特征。其次，训练

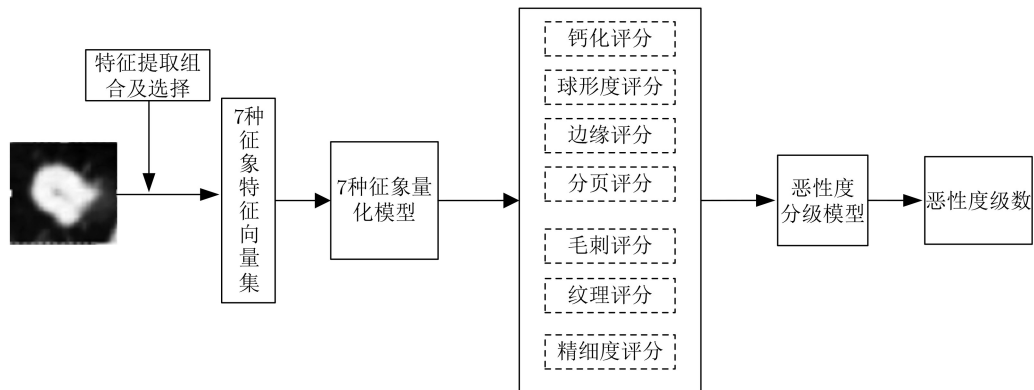


图1 肺结节恶性度分级流程

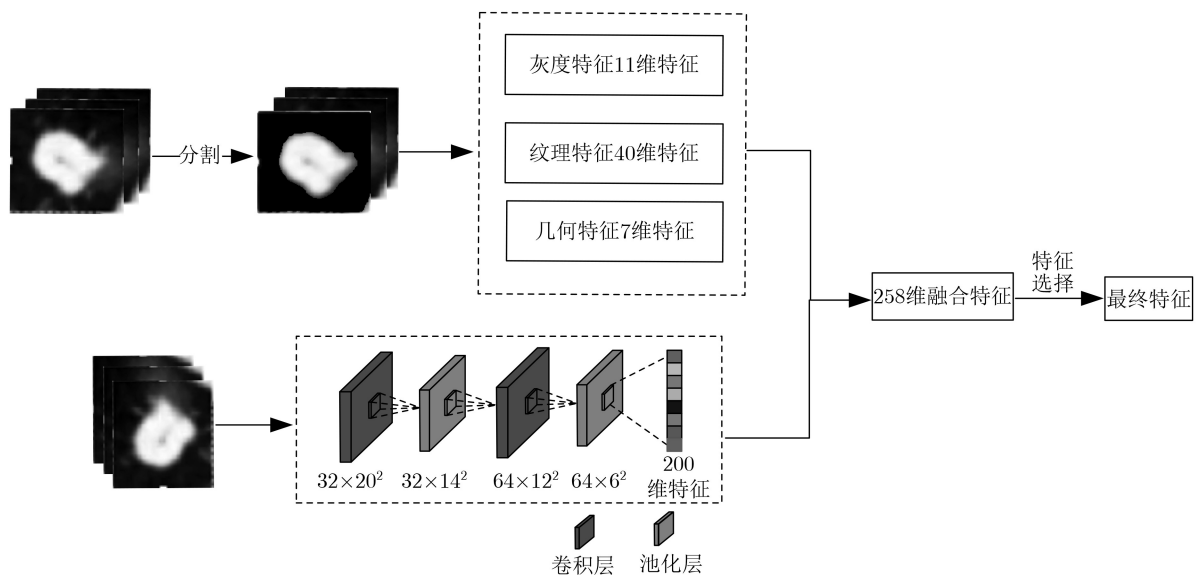


图2 多特征融合流程



7种征象对应的CNN模型，用来提取该征象的CNN高阶特征。最后，合并影像组学特征和该征象所对应的CNN高阶特征，并通过特征选择去除合并特征集中的冗余信息。此过程主要分为3个步骤，具体内容如下：

**步骤 1 影像组学特征提取。**主要从灰度、纹理和几何形状3个方面进行提取。对于灰度方面，根据直方图提取图像的最大值、最小值、均值、方差、偏度、峰度、直方图熵、能量、标准差、中值和均方根共11维特征。对于纹理方面，采用灰度共生矩阵在 $(0^\circ, 45^\circ, 90^\circ, 135^\circ)$ 上提取纹理能、纹理熵、对比度、相关性、差异性、纹理熵、同质性、协方差、纹理均值和逆差分矩共40维特征。对于几何形状方面，提取直径、椭圆的长短轴比例、离心率、周长、面积、紧凑度和矩形度共7维特征。

**步骤 2 CNN特征提取。**图2中给出了提取CNN特征所使用的卷积神经网络结构。主要有2个卷积层、2个池化层以及1个最终输出特征的全连接层。卷积核大小分别为 $32 \times 3 \times 3$ ， $64 \times 3 \times 3$ 。池化层大小分别为 $32 \times 2 \times 2$ ， $64 \times 2 \times 2$ 。全连接层大小为 $1 \times 200$ 。在训练模型时增加丢弃率为0.5的dropout层以及softmax层。并设置迭代次数为1000次，batch-size为44，使用Adam优化器，其中Adam的学习率取0.0001，损失函数为cross\_entropy。通过训练每种征象的分类准确率都可达到85%以上，最后从全连接层提取到200维的CNN特征。

**步骤 3 特征融合。**将58维的影像组学特征和每种征象的200维CNN特征进行合并，对合并后258维的特征通过特征重要性排名，并设置阈值进行特征选择，输出该征象的最终特征集。

## 2.2 基于进化集成学习的CT征象量化

基于前述特征集，可利用集成学习过程构造不同CT征象的分类器以实现量化计算。随机森林是一种典型的bagging算法，是随机子空间与bootstrap两种算法的结合。根据大数定理的证明随着决策树

的不断增长，最终随机森林的性能会趋于稳定。在决策树构造过程中采用两个随机过程，数据集有放回的随机采样以及构造每个决策树子节点特征向量的随机采样。随机过程的引入保证了整个随机森林中决策树的多样性，但会导致随机森林中存在冗余和性能低的决策树。而随机森林的最终结果是通过投票结果得到的，以此可以看出最终的结果与每棵性能树以及组合树之间的组合有着明显关系。但是bagging方法计算性能的提升是以系统复杂度的增大为代价的。这导致集成分类器的准确率与复杂度之间产生了一个很难调和的矛盾。能否在提高集成学习器准确率的同时优化其组成结构，从而降低模型的复杂度呢？实现思路为利用遗传规划算法(Genetic Programming, GP)的进化搜索过程来生成并优化集成学习器。流程图如图3所示。以下是进化集成学习器的具体步骤：

**步骤 1 初始化种群。**种群由NP个随机生成的个体构成，表示为 $\text{pop} = \{\text{RF}_1, \text{RF}_2, \dots, \text{RF}_{\text{NP}}\}$ ， $j = 1, 2, \dots, \text{NP}$ 。其中NP为种群规模， $\text{RF}_j$ 为tn个决策树的组合，tn为可变参数。为了增加种群的多样性，每次有放回地随机采样40%的数据集作为个体初始化的数据集，设置当前迭代数值 $\text{gen} = 0$ 。

**步骤 2 搜索运算。**种群进化主要依靠交叉和变异算子实现，按照交叉概率 $P_c$ 和变异概率 $P_m$ 选择搜索操作。搜索运算具体为：

(1) 变异运算。变异运算可定义为 $\text{RF}' = \odot \text{RF}$ ，其中RF为从种群 $\text{pop}^{\text{gen}}$ 中随机挑选的父代个体， $\text{RF}'$ 为变异产生的子代个体， $\odot$ 为变异运算符。由于个体的性能与每棵决策树的性能以及决策树之间组合的结构有很大关系，故采取2种变异策略对个体进行变异。

变异策略1为个体中树层面的变异，首先从个体RF中随机挑选两棵树，分别随机选取两棵树的交换点，再互换以交换点为根节点的子树，生成新的个体 $\text{RF}'$ 。

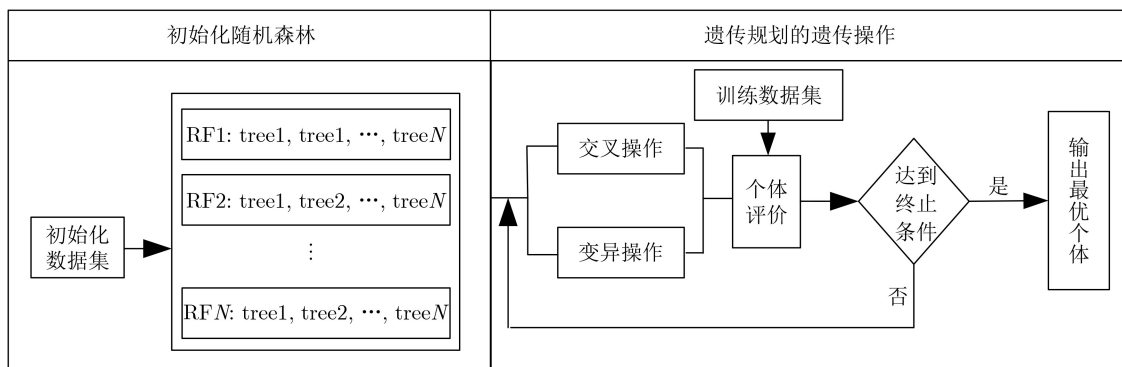


图3 基于进化搜索构造集成学习器的流程

变异策略2为个体结构层面的变异，从个体中增加树和删除树两个方面考虑。如果随机数小于0.5，则向RF中添加一棵随机生成的树，否则将从RF中删除一棵随机被选择的树，然后生成新的个体RF'。按照概率 $P_{ms}$ 将2种不同的变异策略进行有效的结合，不仅对个体的组合结构进行搜索，还对个体中的树进行搜索。

(2) 交叉运算。交叉运算可定义为 $\{RF_1', RF_2'\} = RF_1 \otimes RF_2$ ，其中 $RF_1$ 和 $RF_2$ 为从种群 $pop^{gen}$ 中随机的挑选两个父代个体， $RF_1'$ 和 $RF_2'$ 为交叉后产生的子代个体， $\otimes$ 为交叉运算符。由于个体中树的个数为可变大小，故交叉操作具体为：首先比较 $RF_1$ 和 $RF_2$ 中树的个数 $tn_1$ 和 $tn_2$ ，从树个数较多的 $RF_1$ 中无放回地随机挑选出 $tn_2$ 个树组成 $RF_3$ 。其次对 $RF_3$ 和 $RF_2$ 按概率 $P_{cs}$ 进行均匀交叉，生成新的个体 $RF_3'$ 和 $RF_2'$ ，最后将 $RF_3'$ 中的树放回 $RF_1$ 中得到 $RF_1'$ 。

步骤3 选择运算。对步骤2中产生的子代个体 $RF'$ 与父代个体 $RF$ 进行评估，选择性能较好的个体遗传到下一代种群 $pop^{gen+1}$ 中。采用准确率作为个体评估方式，计算公式为

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

其中，TP表示实际为正被预测为正的样本数量，FP表示实际为负被预测为正的样本数量，TN表示实际为正被预测为负的样本数量，FN表示实际为负被预测为负的样本数量。

步骤4 若下一代种群 $pop^{gen+1}$ 中个体数量小于种群规模NP，则转步骤2，否则转步骤5。

步骤5 置 $gen = gen + 1$ 。判断当前迭代次数 $gen$ 是否达到最大迭代数值NG，若满足则输出最优森林个体，否则转步骤2。

### 2.3 恶性度分级器

softmax回归模型是一种简单有效的多分类器，可作为恶性度分级计算的工具。为了保证恶性度分级计算的准确率，使用基于差分进化算法(Differential Evolution, DE)的强化学习过程对softmax模型进行优化学习。softmax多元模型可表示为

$$\begin{aligned} h_{\mathbf{W}}(X_i) &= \begin{bmatrix} P(y_i = 1|X_i, w_1) \\ P(y_i = 2|X_i, w_2) \\ \vdots \\ P(y_i = K|X_i, w_K) \end{bmatrix} \\ &= \frac{1}{\sum_{k=1}^K \exp(X_i w_k)} \begin{bmatrix} \exp(X_i w_1) \\ \exp(X_i w_2) \\ \vdots \\ \exp(X_i w_K) \end{bmatrix} \end{aligned} \quad (2)$$

其中， $y_i$ 为恶性度分级标签，有 $K$ 个不同的分类级

别。 $\mathbf{W} = [w_1 \ w_2 \ \dots \ w_K]$ 是 $[Q, K]$ 型矩阵， $Q$ 为特征集维度。 $X_i$ 为数据集中第 $i$ 个样本。通过softmax模型计算出样本 $X_i$ 对于每个类别的概率值，最大的概率值位置为该样本最终所属的恶性度类别。

对于softmax模型中的参数 $\mathbf{W}$ ，采用差分进化算法对其进行搜索优化。差分进化算法优化的目标函数是使肺结节恶性度分级的误差率最小化，其目标函数表示为

$$f_{DE} = \max \left\{ \sum_{i=1}^{tn} F_i \right\} \quad (3)$$

$$F_i = \begin{cases} 1, & h_w(X_i) \text{最大概率位置} = y_i \\ 0, & \text{其它} \end{cases} \quad (4)$$

其中， $F_i$ 为样本 $X_i$ 经过softmax函数分级的结果是否正确， $\sum_{i=1}^{tn} F_i$ 为 $n$ 个样本中分级结果正确的数量。另外，在优化过程中需将参数 $\mathbf{W}$ 转成 $Q \times K$ 维的行向量，而在恶性度分级时需将 $Q \times K$ 维的行向量转成 $[Q, K]$ 型矩阵，代入softmax函数中进行计算。

## 3 实验

LIDC-IDRI数据集由1018个病例组成，每个病例包含了4位医生对每个3 mm以上肺结节的坐标以及7种CT征象和恶性度的语义描述评分。4位医生的年资不同，导致对每个结节的标注信息存在差异性。本文将每个结节中4位医生标注的坐标求均值作为该结节的中心坐标，将4位医生对每个CT征象语义和恶性度评分求众数作为CT征象和恶性度的最终评分。最终提取大小为 $30 \times 30$ 像素的2000个肺结节影像样本作为本文的实验数据。由于钙化征象的评分的差异性较大，故本文将此征象的1~5分合成一类进行二分类实验。实验所采用的实验环境为：Inter® Core(i5) 处理器2.3 GHz CPU 8 GB RAM，实验所有的代码使用Python3.6实现，实验中搭建CNN模型时使用Keras2.2.2框架。对于进化集成学习器的参数设置如下：迭代次数NG为500，种群规模NP为200，交叉率 $P_c$ 为0.6，变异率 $P_m$ 为0.4，控制两种变异策略的概率 $P_{ms}$ 为0.5，均匀交叉概率 $P_{cs}$ 为0.5。对于DE优化softmax模型时，参数设置如下：迭代次数为500，交叉率CR为0.3，缩放因子 $F$ 为0.5。

### 3.1 评价指标

本文采用准确率(ACC)、精确率(Pre)、召回率(Rec)和F1-Score作为评价指标，通过五折交叉验证对CT征象的量化结果及恶性度分级进行评估。ACC的计算方法见式(1)，其他3个指标的计算公式为

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{F1-Score} = \frac{2 \cdot \text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}} \quad (7)$$

其中, Pre是指实际为正的样本占被预测为正的样本的比例。Rec是指实际为正的样本中被预测为正的样本占实际为正的样本的比例。F1-Score为Pre和Rec的调和平均数, 可用来检测模型的稳健性。

### 3.2 实验统计结果

#### 3.2.1 7种CT征象的特征集结构

表1为对7种CT征象在进行特征选择时选取的阈值及所对应的特征数量。采用嵌入式特征选择方法, 根据随机森林的特征重要性排名在不降低准确率的情况下设置阈值进行筛选。从表1中可以看出不同的语义征象选取的阈值不同, 对应的有效计算机特征数量也不同。说明不同语义征象在影像中的表现形式不同, 所对应的计算机特征存在差异性。

#### 3.2.2 计算精度

表2为7种CT征象量化和恶性度分级在4个指标上的实验统计结果。由表2可得到7种CT征象量化

表1 不同征象的有效特征

特征	阈值	特征总数量	影像组学特征数	CNN特征数
精细度	0.0050	69	28	41
球形度	0.0030	136	23	113
边缘	0.0035	100	28	72
分叶征	0.0009	203	34	169
毛刺征	0.0060	55	23	22
纹理征	0.0070	39	12	27
钙化征	0.0300	68	20	48

表2 恶性度分级和CT征象量化的精度

指标	精细度	球形度	边缘	分叶征	毛刺征	纹理征	钙化征	恶性度
ACC	0.9668	0.9764	0.9642	0.9693	0.9792	0.9728	0.9844	0.8618
Pre	0.9648	0.9772	0.9659	0.9522	0.9695	0.9733	0.9683	0.8678
Rec	0.9651	0.9785	0.9346	0.9497	0.9708	0.9736	0.9674	0.8619
F1	0.9649	0.9778	0.9499	0.9509	0.9701	0.9735	0.9678	0.8627

表3 恶性度分级模型的权重系数

恶性度等级	精细度	球形度	边缘	分叶征	毛刺征	纹理征	钙化征
1	-0.0703	<b>-0.8560</b>	0.3835	0.2935	0.2751	0.0848	-0.7668
2	-0.0452	0.3054	0.5043	-0.0062	0.0069	<b>-0.6435</b>	-0.4618
3	-0.4187	-0.3507	0.4151	0.3337	0.17592	-0.0367	<b>0.5758</b>
4	0.1984	-0.1948	<b>0.4221</b>	0.0369	0.3217	0.0113	0.2751
5	<b>0.8548</b>	-0.4759	0.2756	-0.2990	0.4641	-0.1176	0.5337

的准确率在0.9642~0.9844, 平均ACC为0.9733, 平均Pre为0.9673, 平均Rec为0.9628, 平均F1为0.9655。肺结节恶性度分级的ACC为0.8618, Pre为0.8678, Rec为0.8617, F1为0.8627。

#### 3.2.3 可解释性价值

表3为优化后的恶性度分级模型的参数值, 可以看出将每种CT征象的量化结果作为恶性度分级特征时, 每种征象的量化结果与最终恶性度分级结果之间的映射关系, 从而可以进一步理解和分析每种CT征象对恶性度分级结果的影响。从表3可以看出球形度评分对恶性度1级的影响较大, 纹理征评分对于恶性度2级的影响相对较大, 钙化征评分对恶性度3级时的影响相对较大, 边缘评分对于恶性度4级的影响相对较大, 精细度评分对恶性度5级的影响较大。

### 3.3 实验对比

#### 3.3.1 征象量化结果对比

为了验证本文提出的进化集成学习方法的有效性, 选取了RF, 极端随机树(Extremely randomized Trees, ET), AdaBoost, XGBoost和梯度提升树(Gradient Boosting Decision Tree, GBDT)共5种典型算法进行实验对比。对比算法采用sklearn算法包实现, 使用网格搜索法对于决策树个数参数进行优化, 其他的参数均为默认参数。表4给出了7种征象在不同集成学习器上的量化结果对比。图4给出了纹理征象的进化集成学习器在优化过程中最优个体的ACC及子树数量的变化曲线。

从表4中可以发现ET的性能在3个对比算法中表现最好, XGBoost相对RF在性能上有一定的提高。本文方法相对ET, ACC平均提高0.021, F1指标平均提高0.016。对比XGBoost, ACC平均提高0.028, F1指标平均提高0.020。对比RF, ACC平均

表 4 不同集成学习器的量化计算结果对比

对分类器	指标	精细度	球形度	边缘	分叶征	毛刺征	纹理征	钙化征
ET	ACC	0.9638	0.9526	0.9422	0.9603	0.9372	0.9572	0.9512
	Pre	0.9637	0.9558	0.9506	0.9616	0.9372	0.9574	0.9507
	Rec	0.9646	0.9532	0.8994	0.9590	0.9374	0.9572	0.9517
	F1	0.9638	0.9541	0.9197	0.9598	0.9365	0.9573	0.9511
	树个数	112	100	88	108	152	76	76
XGBoost	ACC	0.9621	0.9428	0.9452	0.9560	0.9275	0.9312	0.9498
	Pre	0.9619	0.9442	0.9504	0.9571	0.9276	0.9311	0.9489
	Rec	0.9625	0.9437	0.9266	0.9566	0.9289	0.9311	0.9506
	F1	0.9621	0.9438	0.9377	0.9566	0.9282	0.9311	0.9496
	树个数	188	180	192	188	176	110	80
RF	ACC	0.9585	0.9411	0.9422	0.9491	0.9490	0.9637	0.9471
	Pre	0.9583	0.9449	0.9399	0.9497	0.9492	0.9642	0.9466
	Rec	0.9593	0.9422	0.9166	0.9488	0.9502	0.9637	0.9477
	F1	0.9586	0.9432	0.9270	0.9491	0.9488	0.9639	0.9470
	树个数	128	172	188	136	128	128	72
本文方法	ACC	<b>0.9668</b>	<b>0.9764</b>	<b>0.9642</b>	<b>0.9693</b>	<b>0.9792</b>	<b>0.9728</b>	<b>0.9844</b>
	Pre	<b>0.9648</b>	<b>0.9772</b>	<b>0.9659</b>	<b>0.9522</b>	<b>0.9695</b>	<b>0.9733</b>	<b>0.9683</b>
	Rec	<b>0.9651</b>	<b>0.9785</b>	<b>0.9346</b>	<b>0.9697</b>	<b>0.9708</b>	<b>0.9736</b>	<b>0.9674</b>
	F1	<b>0.9649</b>	<b>0.9778</b>	<b>0.9499</b>	<b>0.9509</b>	<b>0.9701</b>	<b>0.9735</b>	<b>0.9678</b>
	树个数	<b>77</b>	<b>67</b>	<b>78</b>	<b>54</b>	<b>70</b>	<b>60</b>	<b>23</b>

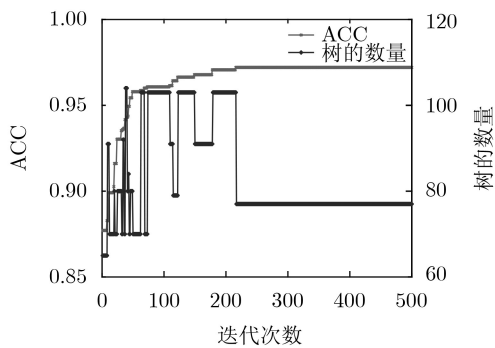


图 4 进化集成学习器的优化过程

提高0.023, F1指标平均提高0.014。在树个数方面可以看出本文方法产生的分类器所需的决策树个数最少,比RF平均降低了65%,比XGBoost平均降低72%,比ET平均降低39%。

从图4中可以明显看出最优个体的ACC为0.972,子树数量稳定为70。显然,相较于传统的集成学习方法,本文提出的进化集成学习方法不但可提高集成分类器的性能,还可在进化过程中不断优化自身的结构,以降低分类器的复杂度。

本文还与近几年文献中所提出的4种主要基于CNN的方法进行了对比。文献[11]通过CNN学习到不同CT征象的表示形式,然后组合起来用于推断肺结节良恶性分类结果。文献[22]提出一种多阶转

移学习法来评估肺结节的分级性能,通过构建每个CT语义的神经网络来提取特征。将每个特征作为输入用来探索哪些属性可以支持恶性肿瘤诊断。文献[23]通过提取影像组学特征并进行特征选择得到7个CT征象和恶性度的特征集,然后对其进行分类。文献[24]训练多个CT征象的CNN模型,再将征象的计算结果放入网络中进行恶性度分类。表5给出了上述文献方法与本文方法的征象量化及恶性度分级结果的ACC指标对比。

从表5中可以看出本文的恶性度分级准确率略低于文献[24],但明显好于其他3种算法,而对7种CT征象的评分结果则都好于文献[24]。这说明本文提出的基于进化集成学习的CT征象量化方法是有效的,而基于softmax模型的恶性度分级方法的准确率并不理想。但是,通过softmax模型可计算7种CT征象与5级恶性度之间的权重关系,在结果的可理解性方面相对于CNN方法具有明显优势。另一方面,文献[23]在特征提取方面只提取了影像组学特征,因此其精度不如本文方法。可见使用CNN特征加影像组学特征的特征集构造方法能够更全面地反映CT影像中的信息。

### 3.3.2 不同结构特征集的对比分析

为了说明影像组学特征和CNN特征融合的有



效性,以纹理征象为例来讨论特征集的聚类分析效果。因为纹理征象分值为1~5分,所以将其聚成5类。首先分别对影像组学特征和CNN特征以及融合特征3组不同结构的特征集进行聚类计算,对聚类效果采用互信息、均一性、同质性和完整性的调和平均值(v-measure)进行评价。这3个指标用来衡量聚类产生的类别与真实标签之间的关系,指标越大表明聚类效果越好。

从图5中可以看出融合后的特征相对2个单一结构特征的不同类簇之间的边界更清晰,同时对同一类簇内的内聚性效果更好。从表6也可以看出融合

后的特征在3个指标上的结果都更佳,这也说明了融合特征的有效性。

#### 4 结束语

本文提出了一种基于语义标签的恶性度分级方法。为了验证此方法,在LIDC-IDRI数据集上进行测试,同时与多种典型算法以及近几年相关文献中提出的CNN方法进行了对比。结果显示,本文方法具有较高的准确率,同时更有利于提高计算结果的可解释性。在讨论中分析了本文所采用的特征融合方法的有效性。

表5 相关文献的量化结果对比

文献	精细度	球形度	边缘	分叶征	毛刺征	纹理征	钙化征	恶性度
文献[11]	0.7190	0.5222	0.7250	/	/	0.8340	0.9080	0.8420
文献[22]	0.7163	0.7392	0.7332	0.7248	0.7422	0.7677	0.7390	0.8194
文献[23]	0.7431	0.7622	0.7013	0.8001	0.7828	0.9002	/	0.7556
文献[24]	0.8933	0.8933	0.8933	0.8933	0.8933	0.8933	0.8933	<b>0.8933</b>
本文方法	<b>0.9668</b>	<b>0.9764</b>	<b>0.9642</b>	<b>0.9693</b>	<b>0.9792</b>	<b>0.9728</b>	<b>0.9844</b>	0.8618

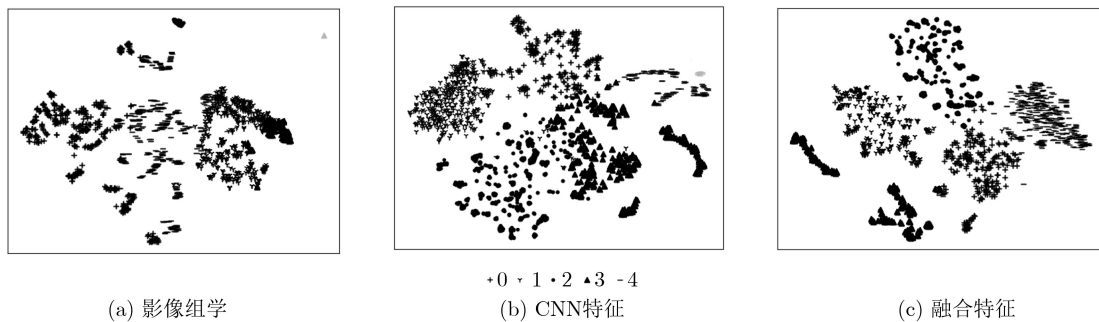


图5 不同特征集合的聚类结果可视化对比

表6 不同特征集合的聚类结果对比

特征集	均一性	v-measure	互信息
影像组学特征	0.32610	0.3233	0.3179
CNN特征	0.44342	0.4282	0.4118
融合特征	0.60850	0.5934	0.5771

#### 参考文献

- [1] MCWILLIAMS A, TAMMEMAGI M C, MAYO J R, *et al.* Probability of cancer in pulmonary nodules detected on first screening CT[J]. *New England Journal of Medicine*, 2013, 369(10): 910–919. doi: [10.1056/NEJMoa1214726](https://doi.org/10.1056/NEJMoa1214726).
- [2] NAIDICH D P, BANKIER A A, MACMAHON H, *et al.* Recommendations for the management of subsolid pulmonary nodules detected at CT: A statement from the Fleischner society[J]. *Radiology*, 2013, 266(1): 304–317. doi: [10.1148/radiol.12120628](https://doi.org/10.1148/radiol.12120628).
- [3] GOULD M K, DONINGTON J, LYNCH W R, *et al.* Evaluation of individuals with pulmonary nodules: When is

it lung cancer?: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines[J]. *Chest*, 2013, 143(5S): e93S–e120S. doi: [10.1378/chest.12-2351](https://doi.org/10.1378/chest.12-2351).

- [4] TRAVIS W D, BRAMBILLA E, NOGUCHI M, *et al.* International association for the study of lung cancer/American thoracic society/European respiratory society: International multidisciplinary classification of lung adenocarcinoma[J]. *Proceedings of the American Thoracic Society*, 2011, 8(5): 381–385. doi: [10.1513/pats.201107-042ST](https://doi.org/10.1513/pats.201107-042ST).
- [5] RODRIGUES M B, DA NÓBREGA R V M, ALVES S S A, *et al.* Health of things algorithms for malignancy level classification of lung nodules[J]. *IEEE Access*, 2018, 6: 18592–18601. doi: [10.1109/ACCESS.2018.2817614](https://doi.org/10.1109/ACCESS.2018.2817614).
- [6] DA NÓBREGA R V M, PEIXOTO S A, DA SILVA S S P, *et al.* Lung nodule classification via deep transfer learning in CT lung images[C]. The IEEE 31st International Symposium on Computer-Based Medical Systems, Karlstad,



- Sweden, 2018: 244–249.
- [7] ZUO Wangxia, ZHOU Fuqiang, LI Zuoxin, *et al.* Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection[J]. *IEEE Access*, 2019, 7: 32510–32521. doi: [10.1109/ACCESS.2019.2903587](https://doi.org/10.1109/ACCESS.2019.2903587).
- [8] SHEN Shiwen, HAN S X, ABERLE D R, *et al.* An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification[J]. *Expert Systems with Applications*, 2019, 128: 84–95. doi: [10.1016/j.eswa.2019.01.048](https://doi.org/10.1016/j.eswa.2019.01.048).
- [9] WANG Huafeng, ZHAO Tingting, LI Lihong, *et al.* A hybrid CNN feature model for pulmonary nodule malignancy risk differentiation[J]. *Journal of X-Ray Science and Technology*, 2018, 26(2): 171–187. doi: [10.3233/XST-17302](https://doi.org/10.3233/XST-17302).
- [10] 褚征, 于炯. 基于随机森林的流处理检查点性能预测[J]. *电子与信息学报*, 2020, 42(6): 1452–1459. doi: [10.11999/JEIT190552](https://doi.org/10.11999/JEIT190552).  
CHU Zheng and YU Jiong. Performance prediction based on random forest for the stream processing checkpoint[J]. *Journal of Electronics & Information Technology*, 2020, 42(6): 1452–1459. doi: [10.11999/JEIT190552](https://doi.org/10.11999/JEIT190552).
- [11] NISHIO M, NISHIZAWA M, SUGIYAMA O, *et al.* Computer-aided diagnosis of lung nodule using gradient tree boosting and bayesian optimization[J]. *PLoS One*, 2018, 13(4): e0195875. doi: [10.1371/journal.pone.0195875](https://doi.org/10.1371/journal.pone.0195875).
- [12] 吴艇帆, 张仁寿. 机器学习在肺内恶性磨玻璃密度结节的应用研究[J]. *广州大学学报: 自然科学版*, 2018, 17(3): 33–39.  
WU Tingfan and ZHANG Renshou. Research on the application of machine learning in the malignant grinding glass density nodules of lung[J]. *Journal of Guangzhou University: Natural Science Edition*, 2018, 17(3): 33–39.
- [13] FERREIRA JR J R, OLIVEIRA M C, and DE AZEVEDO-MARQUES P M. Characterization of pulmonary nodules based on features of margin sharpness and texture[J]. *Journal of Digital Imaging*, 2018, 31(4): 451–463. doi: [10.1007/s10278-017-0029-8](https://doi.org/10.1007/s10278-017-0029-8).
- [14] WU Wenhao, HU Huihui, GONG Jing, *et al.* Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis[J]. *Physics in Medicine & Biology*, 2019, 64(3): 035017. doi: [10.1088/1361-6560/aafab0](https://doi.org/10.1088/1361-6560/aafab0).
- [15] SHAUKAT F, RAJA G, ASHRAF R, *et al.* Artificial neural network based classification of lung nodules in CT images using intensity, shape and texture features[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2019, 10(10): 4135–4159. doi: [10.1007/s12652-019-01173-w](https://doi.org/10.1007/s12652-019-01173-w).
- [16] 李双双, 侯震, 刘娟, 等. 影像学分析与建模工具综述[J]. *中国医学物理学杂志*, 2018, 35(9): 1043–1049. doi: [10.3969/j.issn.1005-202X.2018.09.010](https://doi.org/10.3969/j.issn.1005-202X.2018.09.010).  
LI Shuangshuang, HOU Zhen, LIU Juan, *et al.* Review of radiomic analysis and modeling tools[J]. *Chinese Journal of Medical Physics*, 2018, 35(9): 1043–1049. doi: [10.3969/j.issn.1005-202X.2018.09.010](https://doi.org/10.3969/j.issn.1005-202X.2018.09.010).
- [17] SHI Zhenghao, HAO Huan, ZHAO Minghua, *et al.* A deep CNN based transfer learning method for false positive reduction[J]. *Multimedia Tools and Applications*, 2019, 78(1): 1017–1033. doi: [10.1007/s11042-018-6082-6](https://doi.org/10.1007/s11042-018-6082-6).
- [18] 刘家辰, 苗启广, 曹莹, 等. 基于混合多样性生成与修剪的集成单类分类算法[J]. *电子与信息学报*, 2015, 37(2): 386–393. doi: [10.11999/JEIT140161](https://doi.org/10.11999/JEIT140161).  
LIU Jiachen, MIAO Qiguang, CAO Ying, *et al.* Ensemble one-class classifiers based on hybrid diversity generation and pruning[J]. *Journal of Electronics & Information Technology*, 2015, 37(2): 386–393. doi: [10.11999/JEIT140161](https://doi.org/10.11999/JEIT140161).
- [19] ELYAN E and GABER M M. A fine-grained random forests using class decomposition: An application to medical diagnosis[J]. *Neural Computing and Applications*, 2016, 27(8): 2279–2288. doi: [10.1007/s00521-015-2064-z](https://doi.org/10.1007/s00521-015-2064-z).
- [20] MIAO Fen, CAI Yunpeng, ZHANG Yuxiao, *et al.* Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest[J]. *IEEE Access*, 2018, 6: 7244–7253. doi: [10.1109/ACCESS.2018.2789898](https://doi.org/10.1109/ACCESS.2018.2789898).
- [21] PAUL A, MUKHERJEE D P, DAS P, *et al.* Improved random forest for classification[J]. *IEEE Transactions on Image Processing*, 2018, 27(8): 4012–4024. doi: [10.1109/tip.2018.2834830](https://doi.org/10.1109/tip.2018.2834830).
- [22] WANG Qingfeng, CHENG Jiezh, LIU Zhiqin, *et al.* Multi-order transfer learning for pathologic diagnosis of pulmonary nodule malignancy[C]. 2018 IEEE International Conference on Bioinformatics and Biomedicine, Madrid, Spain, 2018: 2813–2815.
- [23] YANG Jing, LI Na, FANG Shuai, *et al.* Semantic features prediction for pulmonary nodule diagnosis based on online streaming feature selection[J]. *IEEE Access*, 2019, 7: 61121–61135. doi: [10.1109/ACCESS.2019.2903682](https://doi.org/10.1109/ACCESS.2019.2903682).
- [24] WU Botong, ZHOU Zhen, WANG Jianwei, *et al.* Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction[C]. The IEEE 15th International Symposium on Biomedical Imaging, Washington, USA, 2018: 1109–1113.
- 陈皓: 男, 1978年生, 博士, 副教授, 硕士生导师, 主要研究方向为医疗大数据。  
段红柏: 男, 1993年生, 硕士生, 研究方向为数据挖掘和模式识别。  
郭紫园: 女, 1996年生, 硕士生, 研究方向为数据挖掘和进化计算。  
强永乾: 男, 1965年生, 博士, 副教授, 硕士生导师, 研究方向为医学影像学。