

基于联合训练的分类器的乳腺癌图像分类

张晋凯, 高翔, 王鹏, 白艳萍, 梅银珍

(中北大学 数学学院, 太原 030051)

摘要: 利用机器学习的乳腺癌组织病理图像诊断节省了大量的人力物力, 因此提高乳腺癌组织病理图像识别准确率有很好的现实意义; 针对单一分类器和集成学习分类器模型观测域有限容易陷入局部最优的问题, 提出一种基于联合训练的分类器模型; 通过单一分类器相互影响扩大观测感知域来寻找损失最小的估计点, 根据估计点来迭代优化超参数进而联合训练出拟合性能最好的分类器, 这样既汲取不同分类器模型的可取之处来增强泛化能力, 又加大了模型观测域在可以更快的得到全局最优的同时提升了识别准确率; 实验表明, 提出的联合训练的分类器能够提升乳腺癌组织病理学图像的分类性能, 在不同放大倍数 $40\times$ 、 $100\times$ 、 $200\times$ 、 $400\times$ 下图像良恶性分类准确率分别为 99.67%、98.08%、99.01%、96.34%。

关键词: 乳腺癌图像; 分类器模型; 联合训练; 深度学习; 机器学习

Breast Cancer Image Classification Based on Joint Training Classifier

ZHANG Jinkai, GAO Xiang, WANG Peng, BAI Yanping, MEI Yinzen

(School of Mathematics, North University of China, Taiyuan 030051, China)

Abstract: The histopathological image diagnosis of breast cancer using machine learning saves a lot of manpower and material resources, so it is of good practical significance to improve the histopathological image recognition accuracy of breast cancer. Aiming at the problem that the observation domain of single classifier and ensemble learning classifier model is limited and easy to fall into local optimality, a classifier model based on joint training is proposed. Through the interaction of single classifier, the observation perception domain is expanded to find the estimated point with the least loss, and the hyperparameters are iteratively optimized according to the estimated point and then jointly trained the classifier with the best fitting performance, which not only absorbs the desirability of different classifier models to enhance the generalization ability, but also increases the model observation domain to obtain the global optimal faster while improving the recognition accuracy. The experimental results show that the proposed jointly training classifier can improve the classification performance of the breast cancer histopathological images, and the accuracy of benign and malignant classification at different magnifications of $40\times$, $100\times$, $200\times$ and $400\times$ is 99.67%, 98.08%, 99.01% and 96.34%, respectively.

Keywords: breast cancer images; classifier models; joint training; deep learning; machine learning

0 引言

根据世界卫生组织国际癌症研究机构 (IARC) 的 latest 统计数据显示: 2021 年乳腺癌成为全球发病率最高的癌症, 其发病率和死亡率均为女性癌症首位, 并且整体发病率呈上升和年轻化趋势^[1]。因此, 乳腺肿瘤的早期检测和诊断治疗显得尤为重要。目前组织病理学分析是乳腺癌诊断的主要方法。病理学家根据活检获得的病理图像来判断乳腺癌良恶性^[2]。但是由于病理图像本身的复杂度会导致病理学家有细微的诊断分歧进而减缓诊断效率, 并且长时间工作也会影响诊断的结果甚至是误判, 从而需要大量人力物力来分析病理学图像。近年来, 利用机器学习诊断乳腺癌的计算机自动识别技术愈发成熟, 它通过分析大量的乳腺癌病理图像数据来提高诊断的准确率与效率。卷积神经网络

提取组织病理学图像特征后, 利用机器学习模型来对乳腺癌图像进行分析, 进行乳腺癌良恶性分类。

Spanhol 等^[3] 利用卷积神经网络进行特征提取与之前使用手工制作的纹理描述符进行比较, 并使用最近邻算法、支持向量机、决策树、随机森林等不同的分类器对乳腺癌病理图像特征分类, 实现了识别率的提高。Hou 等^[4] 采用数据增强和迁移学习来改进的深度卷积神经网络模型对乳腺癌病理图像进行自动分类, 既加快了训练速度不需要重新训练初始化权重又避免训练图像样本量少而出现深度学习模型过拟合。Hameed 等^[5] 提出一种集成深度学习方法来对癌类和非癌类的乳腺癌组织病理学图像进行分类, 其采用经过预训练微调的 VGG16 和 VGG19 网络模型的集合预测概率的平均值来提升单模型分类的性能指标。于凌涛等^[6] 采用基于 Inceptionv3 架构的卷积神经网络和图像分块

收稿日期: 2023-02-22; 修回日期: 2023-02-28。

基金项目: 国家自然科学基金项目 (61774137); 山西省基础研究计划资助项目 (202103021224195, 202103021224212, 202103021223189, 20210302123019); 山西省回国留学人员科研项目 (2020-104, 2021-108)。

作者简介: 张晋凯 (1996-), 男, 山西忻州人, 硕士, 主要从事图像处理与识别、机器学习方向的研究。

通讯作者: 梅银珍 (1977-), 女, 山西原平人, 博士, 副教授, 主要从事应用数学方向的研究。

引用格式: 张晋凯, 高翔, 王鹏, 等. 基于联合训练的分类器的乳腺癌图像分类[J]. 计算机测量与控制, 2023, 31(5): 228-234.

化思想对乳腺癌病理图像数据集良恶性分类。Sharma 等^[7]对比分析了基于手工制作的颜色、纹理等特征与基于预训练网络作为特征提取器分别来训练传统分类器支持向量机的乳腺癌组织病理学图像分类性能, 表明使用预训练网络作为特征提取器表现出更好的性能。Yurttakal 等^[8]采用梯度提升和深度学习的堆叠集成模型诊断乳腺癌肿瘤, 结果表明集成模型优于一些单一的机器学习方法。由于数据集训练样本数量少以及癌细胞的粘连性和颜色分布不均匀等问题, 现阶段研究大多是乳腺癌良恶性二元分类, 对良恶性子类分类研究较少, 为了进一步满足临床应用的需求, Han 等^[9]提出了一种结构化深度卷积神经网络模型, 对类内和类间分层特征空间自动学习低级到高级的语义特征和判别性分层特征, 用于对相似度高的不同类别精准化多分类, 实验表明模型多分类精度较高且具有一定稳定性。利用卷积神经网络可以提取出不同类别病理图像的局部细节特征和全局结构特征, 但对提取的特征进行分析训练分类器进而准确的进行图像分类也是重中之重。

现有的关于乳腺癌病理图像分析的模型大多基于单一的模型和集成学习^[10]模型, 单一分类器模型分类精确率不高, 集成分类器模型也存在着模型复杂、参数量多导致的计算量大和训练时间过长的问题。本文构建了联合训练的分类器的乳腺癌病理图像分析模型, 通过贝叶斯优化对几种分类器共同训练寻找最优超参数以提升分类器模型的准确率, 来辅助医生进行诊断。在提高医生的工作效率的同时减少因经验不足而导致的误诊或漏诊病例的现象。

1 方法与原理

1.1 卷积神经网络

卷积神经网络 (CNN, convolutional neural network) 的研究始于 20 世纪 80 至 90 年代, 到 21 世纪后, 随着深度学习理论的提出、数值计算设备的改进以及各种标准的数据集的出现, 卷积神经网络得到了快速发展, 一些卷积神经网络的经典模型层出不穷, 比如 VGG^[11]、ResNet、DenseNet^[12]等。

在乳腺癌组织病理学图像中, 不管是轮廓、边缘、纹理等局部细节更丰富的低层特征, 还是通过深层卷积网络得到的感受野更大的、全局信息更多的高层特征都包含很

多的图像特征信息。在传统的卷积神经网络中, 如果有 L 层, 那么有 L 个连接, 基于 DenseNet 的卷积神经网络将每一层所有的特征图进行通道拼接, 然后作为输入传递给之后所有层, 通过建立前面所有层与后面每一层的密集连接来实现特征重用, 如果一个 DenseNet 网络有 L 层, 那么其有 $L(L+1)/2$ 个连接。这种连接方式不仅加强了特征的传递, 缓解了深层网络梯度消失的问题, 而且减少了参数量避免过拟合。这些特点让 DenseNet 在参数和计算成本更少的情形下比其他卷积神经网络实现更优的性能。因此本文采用 DenseNet201 预训练网络来提取乳腺癌图像卷积特征。如图 1 所示, DenseNet201 模型包含 4 个稠密块, 每个稠密块由多个 1×1 卷积层, 3×3 卷积层构成的块 (block) 组成。各个稠密块中 block 的数量分别为 6, 12, 48, 32。稠密块越多相对应通道数也增加, 模型也更加复杂。所以稠密块后面是 1×1 卷积层和步幅为 2 的 2×2 平均池化层组成的过渡块, 卷积层用来降低通道数, 池化层进一步降低模型复杂度, 最后一个稠密块后面没有过渡块。另外, 通过迁移学习将在 ImageNet 数据集预训练好的模型参数在新的全连接层和输出层进行训练和微调, 加快训练速度。

1.2 分类器

分类器^[13]是一种机器学习技术, 它可以用来将输入数据分类到不同的类别中。它的基本原理是, 通过学习训练数据, 构建一个模型, 用于将新的输入数据分类到正确的类别中。分类器可以用于许多不同的应用, 包括文本分类、图像分类、语音识别等。分类器首先从训练数据中提取特征, 然后, 使用这些特征构建一个模型, 用于将新的输入数据分类到正确的类别中。分类器可以使用不同的算法, 比如支持向量机^[14]、K 最邻近法^[15]、决策树、朴素贝叶斯^[16]等。这些算法都有自己的优点和缺点, 因此, 在使用分类器时, 需要根据实际情况选择合适的算法。分类器主要目标是尽可能地逼近真正的分类模型, 有效地提取输入数据的特征, 并且减少参数的数量来降低模型的复杂度, 提高模型的泛化能力和可解释性。

1.2.1 支持向量机

支持向量机 (SVM, support vector machine) 是一个机器学习的过程, 一般来讲, 它是一种二分类模型, 其基

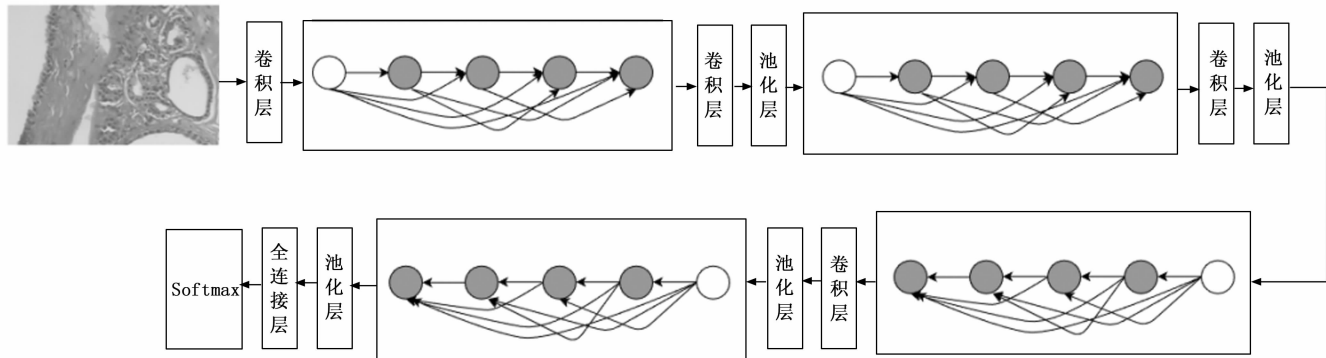


图 1 包含 4 个稠密块的 DenseNet201 结构图

本模型定义为使特征空间间隔最大的线性分类器，它的学习策略是通过在高维空间中寻找一个最大间隔分类超平面，将不同类别的数据样本点分开而使不同类别的点之间的间隔最大，将目标函数优化问题最终转化为一个凸二次规划问题的求解。由于现实问题中样本并不总是线性可分的，因此引入核函数的思想，将样本从原始空间映射到一个更高维的特征空间，找到合适的超平面，并通过序列最小优化算法求解。SVM 模型分类性能的好坏很大程度上取决于模型参数的选择，SVM 中核函数参数 g 和惩罚因子 C 是决定 SVM 性能的重要参数。

1.2.2 其他分类器

K 最近法 (KNN, k-nearest neighbor) 是一种基于距离的分类方法，待求数据样本的类别是根据特征空间中最近邻的 k 个数据样本的类别来决定的。它的基本原理是，首先按照新样本数据特征与训练样本特征的距离进行排序，然后根据与其最近邻的 k 个样本类别所属最多来判定新样本的类别。如果这 k 个样本大多数属于某一个类别，则该样本也属于这个类别以及具有这个类别上样本的特性，否则再按照某种判别规则确定新样本类别。

决策树 (DT, decision tree) 通过构建一种树形结构来描述一系列的决策过程，根据决策过程将样本进行分类。它的每个内节点表示一个属性特征，每个叶节点代表一个类别输出，我们希望根据最少的内节点可以将所有样本尽可能划分到同一类别中即叶子节点上，也就是说最重要一点是选择最优的划分特征。决策树模型训练复杂度低但不稳定，训练数据轻微变化可能导致节点选择不同，通过一定剪枝策略来避免数据的过拟合。

朴素贝叶斯 (NB, naive bayesian) 是一种基于贝叶斯定理与特征条件独立假设的分类方法，它用后验概率来判断样本类别，朴素贝叶斯模型假设特征之间相互独立，首先根据已知分类的训练样本集计算在各类别下各个特征属性的条件概率估计，然后根据贝叶斯定理和各个类别、特征的概率来判断样本最大概率的类别。朴素贝叶斯模型可以处理多类别问题而且泛化能力也较强，在特征独立假设的前提下与其他分类方法相比具有最小的误差率。

集成学习通过将多个相同分类器或多个分类性能不同的简单分类器组合在一起对应同一样本数据的判断结果来达到学习的目的。它不要求每个学习器性能最好，但多个学习器博采众长来选择分类最优结果，其预测精度和稳定性相比较于单一分类器大部分都提到了提升，但本质上没有改变单个分类器分类准确率，是通过一定的组合策略来提高分类器准确率。随机森林就是一种以决策树作为基预测器的集成学习方法。

1.2.3 贝叶斯优化

贝叶斯优化^[17] (BO, bayesian optimisation) 是一种利用先验知识逼近未知目标函数的后验分布从而调节超参数的超参数优化算法。其通过高斯过程建立目标函数的概率模型，并用它来选择最优的超参数来评估真实的目标函数。

我们的目标并不是使用尽可能多的数据点完全推断未知的目标函数，而是希望能选取获得最优性能的超参数，即求得最大化目标函数的超参数，因此超参数选择就可以看作是一种最优化问题。根据目标函数的初始候选解集合找到下一个最佳观测点，该点具有高均值或高方差来避免陷入局部最优，并将该点加入集合中，重复这一步骤，直至迭代终止，从而最快找到全局最优解。网格搜索通过穷举模型所有超参数组合来寻找最优超参数，当模型超参数多时优化速度慢且针对非凸问题容易陷入局部最优。随机搜索通过随机探索超参数空间的值，但它求解过程中没有利用之前已搜索点的信息，因此贝叶斯优化比网格搜索和随机搜索更为有效。

单个分类器都可进行超参数优化来寻找模型性能最佳的超参数，进而减少模型损失提高分类精度。本文使用贝叶斯优化的支持向量机 (BO-SVM) 来进行对比。采用径向基核函数训练 SVM 分类器，并通过使用贝叶斯超参数优化 SVM 中核函数参数 g 和惩罚因子 C ，找到最小化五折交叉验证损失的超参数。

径向基核函数：

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (1)$$

加入惩罚因子 C 后，支持向量机优化的目标函数和约束条件：

$$\begin{aligned} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ s. t. \quad 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (2)$$

1.2.4 联合训练的分类器

现有的关于乳腺癌病理图像分析的模型大多基于单一分类器模型和集成分类器模型。单一分类器优化是对自身超参数寻找最优结果，当单个分类器无法很好地学习样本时，模型的泛化能力会受到影响，分类器在测试集上的表现也会变差。为了提高模型的准确率，可以使用集成分类器模型，它通过为每种分类器找到最佳超参数，并通过集成学习策略构建集成分类器来增强泛化能力。总体来说，集成分类器模型是在结合的多个单分类器模型优化结果中寻找最优结果，其泛化能力是远大于单个分类器的泛化能力的，但它并不改变单分类器分类准确率。而且随着基分类器个数的增加模型复杂度变高，训练时间也会变长。

本文针对这些问题提出一种联合训练的分类器，不是根据超参数优化迭代结果直接优化单一分类器，而是先联合多个分类器 SVM、KNN、决策树、随机森林和朴素贝叶斯对超参数观测感知域进行扩大以便有损失更小的估计点，再根据估计点来迭代优化超参数进而联合训练出拟合性能较好的分类器。这既汲取不同分类器模型的可取之处来增强泛化能力又加大了模型观测域，在避免陷入局部最优的同时提升分类准确率。图 2 为寻找观测最小值与估计最小值的优化过程，根据最小观测值找到损失最小的估计点方向，下一轮迭代继续朝这个方向寻找最小估计点，如果使

损失降低, 根据找到的估计点更新拟合函数, 如果到某一个值优化梯度成负方向则到达局部最优, 通过增加随机数扰动跳出局部最优找到下一个梯度变正的点, 直到迭代结束。那么损失最小的观测点和估计点也就找到了。表 1 为高斯拟合目标函数迭代步骤。

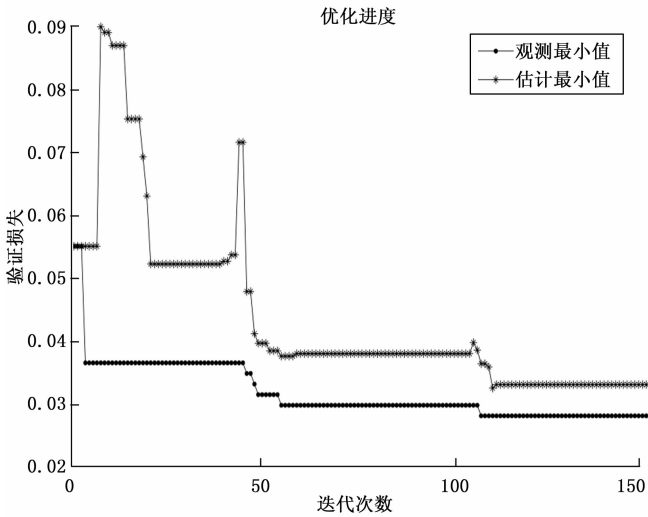


图 2 寻找观测值与估计值的优化过程

表 1 高斯过程拟合目标函数步骤

基本步骤:	
1.	根据已知数据集选择概率代理模型并通过高斯过程回归对目标函数进行建模, 计算每一点处函数的均值与方差
2.	根据均值和方差构造采集函数, 并通过贝叶斯优化过程找到最小估计点
3.	对得到的最小估计点采样得到采样点并更新分类器的超参数以优化模型
4.	循环 2 并不断迭代求得拟合目标函数最优解

表 2 为某一组数据联合训练的分类器迭代优化过程。由表可得出, 迭代次数为 13, KNN 模型为最佳观测分类器, 其超参数 Num Neighbors 为 7; 迭代次数为 61, KNN 模型为最佳估计分类器即返回的模型为 KNN 模型, 其超参数 Num Neighbors 为 16。

2 实验结果与分析

2.1 数据集

2016 年, Spanhol 等人^[18]发布了 BreakHis 乳腺癌组织病理图像公开数据集。该数据集包含了来自 82 位患者的 7 909 幅乳腺癌良恶性组织病理图像, 其中良性 2 480 例, 恶性 5 429 例。本文根据不同放大倍数 40×、100×、200×、400×分为 4 组数据, 针对每组数据, 随机将图像的 70% 作为训练集, 图像的 30% 作为测试集, 具体分布情况如表 3 所示。

良性和恶性乳腺肿瘤又各包括 4 种子类, 良性子类包括腺病 (A)、纤维腺瘤 (F)、叶状肿瘤 (PT)、管状腺瘤 (TA); 恶性子类包括导管癌 (DC)、小叶癌 (LC)、粘液

癌 (MC)、乳头状癌 (PC)。

表 2 联合训练的分类器迭代优化过程

迭代次数	评估结果	验证损失	分类器	超参数	参数值
1	Best	0.04	KNN	Num Neighbors	6
2	Accept	0.08	Tree	Min Leaf Size	1
3	Accept	0.531 2	NB	Distribution Names	Kernel
				width	0.000 194 6
4	Accept	0.088	Tree	Min Leaf Size	2
5	Best	0.033 6	ensemble	Method	Ada Boost M2
				Num Learning Cycles	209
				Min Leaf Size	3
				Max Num Splits	114
6	Accept	0.12	Tree	Min Leaf Size	20
7	Accept	0.531 2	SVM	Coding	Onewsone
				Box Constraint	4.020 1
				Kernel Scale	0.294 54
8	Accept	0.848	SVM	Coding	Onewsall
				Box Constraint	0.006 115 3
				Kernel Scale	0.002 354 6
9	Accept	0.228 8	Tree	Min Leaf Size	50
10	Accept	0.193 6	NB	Distribution Names	Kernel
				width	9.654 9
11	Accept	0.054 4	ensemble	Method	Bag
				Num Learning Cycles	250
				Min Leaf Size	6
				Max Num Splits	18
12	Accept	0.065 6	ensemble	Method	Bag
				Num Learning Cycles	203
				Min Leaf Size	20
				Max Num Splits	69
13	Best	0.032	KNN	Num Neighbors	7
.....
61	Accept	0.033 6	KNN	Num Neighbors	16
62	Accept	0.268 8	KNN	Num Neighbors	110
63	Accept	0.24	SVM	Coding	Onewsall
				Box Constraint	0.470 85
				Kernel Scale	5.361
.....
136	Accept	0.531 2	KNN	Num Neighbors	312

表 3 按放大倍数的良恶性分类图片数目

放大倍数	良性	恶性	总计
40×	625	1 370	1 995
100×	644	1 437	2 081
200×	623	1 390	2 013
400×	588	1 232	1 820
总计	2 480	5 429	7 909
病人人数	24	58	82

2.2 评价指标

本文主要的评价指标是图像级准确率, 令 N_{rec_i} 为分类

正确的图片数量， N 为测试集样本的数量，则图像级准确率定义为：

$$Image_accuracy = \frac{N_{rec-1}}{N} \quad (3)$$

为了更直观的反映真实的良恶性样本与预测的良恶性样本的差异，进一步评估本文方法的乳腺癌图像识别性能，本文采用混淆矩阵来对分类结果进行分析，以及敏感性，特异性，精确率和 $F1_score$ 四种评价指标。敏感性表示在所有恶性样本中恶性样本被诊断正确的概率，特异性表示在所有良性样本中良性样本被诊断正确的概率，精确率表示被诊断为恶性的样本中诊断正确的概率， $F1_score$ 是敏感性和精确率的调和平均数，来平衡两者使其同时达到最高。

2.3 结果比较

根据单一分类器与本文所提出的联合训练的分类器的比较，实验结果如表 4 所示。可以看出无论是对单一分类器 SVM，还是对超参数优化后的 BO-SVM，联合训练的分类器对不同放大倍数的图像分类准确率都得到了提升，在 $40\times$ 、 $100\times$ 、 $200\times$ 、 $400\times$ 下得到的准确率分别为 99.67%、98.08%、99.01%、96.34%。结果表明本文提出的联合训练的分类器一定程度上能有效提高乳腺癌病理图像识别准确率。

表 4 图像层面的准确率

	40×	100×	200×	400×
SVM	96.99%	94.55%	96.85%	94.87%
BO-SVM	99.33%	97.44%	99.01%	95.05%
联合训练的分类器	99.67%	98.08%	99.01%	96.34%

图 3 为不同放大倍数下的混淆矩阵。横坐标表示样本的预测标签，纵坐标表示样本的真实标签。从图 3 可以看出，良性样本被误判为恶性样本的数量并不多，反而 400 倍数恶性样本被误判为良性样本的数量比较多，说明样本不均衡对该模型二分类的影响并不大。但是由于 400 倍数图像放大倍数过大，图像中包含更多精确的病灶位置局部信息，缺少了全局信息，导致识别准确率较低。

表 5 不同放大倍数下的良恶性识别结果

放大倍数	准确率 /%	敏感性 /%	特异性 /%	精确率 /%	$F1_score$ /%
40×	99.67	99.51	98.93	98.94	99.22
100×	98.08	97.91	96.37	96.43	97.16
200×	99.01	99.28	98.40	98.41	98.84
400×	96.34	93.24	98.86	98.80	95.94

乳腺癌组织病理图像良恶性分类是为了更好的帮助病理学家来对患者进行检测。将良性误分类为恶性可能会花费一定时间来重新检测，而将恶性误分类为良性则可能导致患者错失最佳治疗时间。则敏感性越高，可以越快让病人得到及时治疗，表 5 中 4 种放大倍数的敏感性分别为 99.51%、97.91%、99.28%、93.24%，很大程度上避免了乳腺癌恶性被漏诊的情况。 $F1_score$ 分别为 99.22%、97.16%、98.84%、95.94%，也说明了恶性样本在不漏诊的同时也有较高的精确率。

表 6 为与文献中几种方法的比较，文献 [19-21] 以 Protocol 原则划分数据集，文献 [21-22] 以图像为单位划分数据集，所用数据集和数据集划分比例均与本文一致，

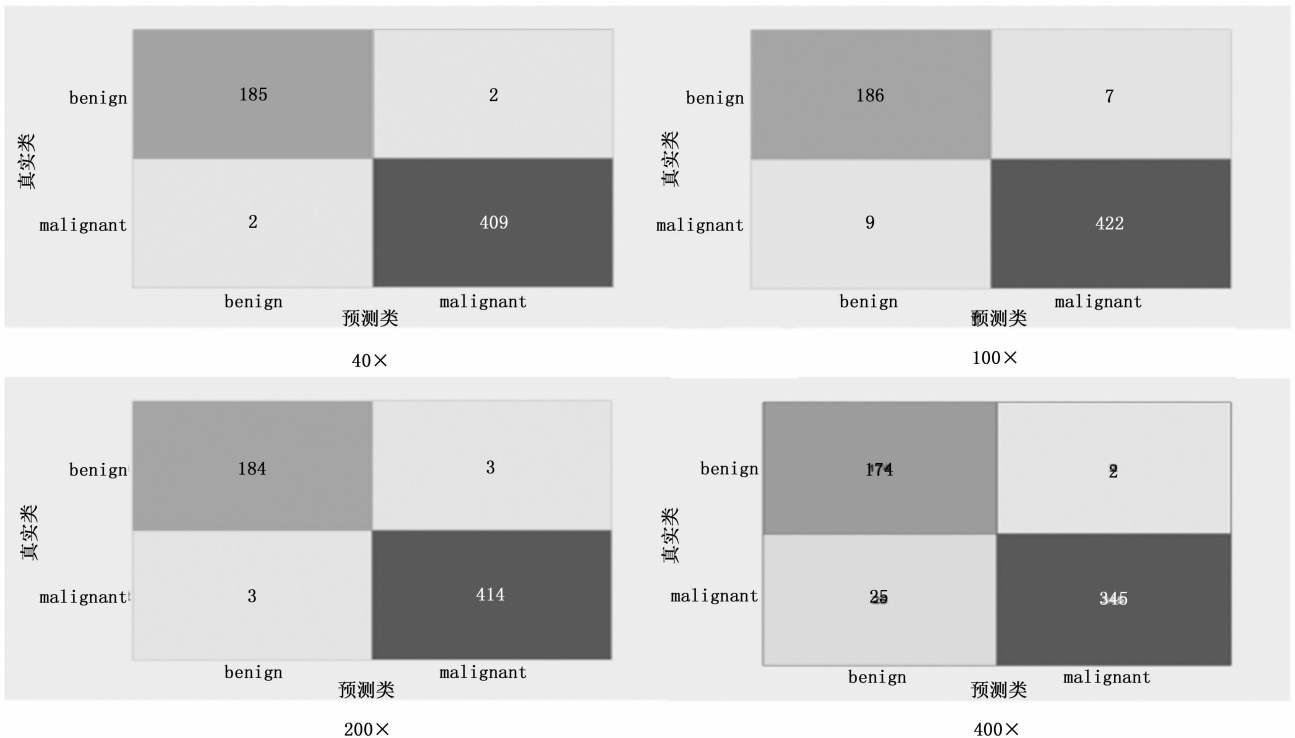


图 3 二分类不同放大倍数的混淆矩阵

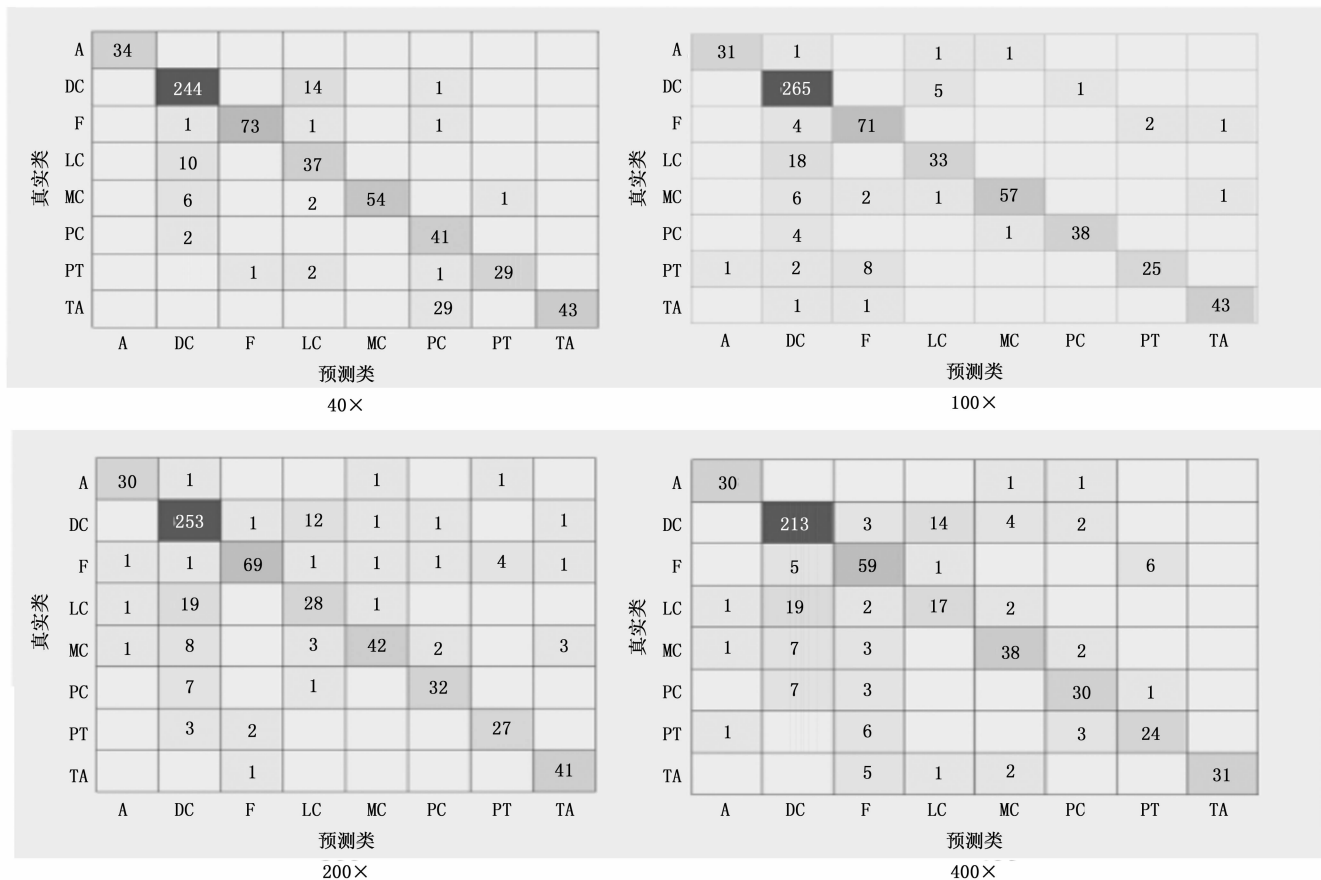


图 4 八分类不同放大倍数的混淆矩阵

其中训练集 70%，测试集 30%。对比图像级准确率可以看出，在 40×和 200×的放大倍数上，本文方法的识别准确率均高于文献中方法的识别准确率。在 100×和 400×放大倍数上，文献 [21] 的识别准确率略高于本文方法，而其他文献的识别准确率均比本文的较低。可以看出相较于文献中提出的方法，本文方法在乳腺癌组织病理学图像识别问题中是有效的，但依然有提升的空间。

表 6 与其他方法的图像级准确率对比

文献	40×	100×	200×	400×
文献[19]	87.71%	88.92%	90.12%	85.16%
文献[20]	90.40%	90.80%	93.45%	88.45%
文献[21]	99.00%	98.50%	98.50%	98.00%
文献[22]	92.87%	94.63%	94.28%	93.78%
文献[23]	98.33%	97.12%	97.85%	96.15%
本文	99.67%	98.08%	99.01%	96.34%

为了更好的评估实验结果，我们又对不同的子类准确率进行了比较，图 4 是其对应的混淆矩阵。A 表示腺病，F 表示纤维腺瘤，PT 表示叶状肿瘤，TA 表示管状腺瘤，DC 表示导管癌，LC 表示小叶癌，MC 表示粘液癌，PC 表示乳头状癌。可以看出，由于导管癌的样本数量远远超过其他子类，样本严重不均衡导致识别结果偏向于样本多的类，

子类图像的准确率还有待提升。

3 结束语

本文针对单一分类器超参数观测域小提出了一种联合训练的分类器。通过联合几种分类器扩大其观测域来迭代优化超参数进而联合训练出拟合性能较好的分类器。在 BreakHis 数据集的实验结果表明提出的联合训练的分类器可以有效提高乳腺癌病理图像识别准确率。对不平衡样本数据的问题未来可以通过数据扩充与增强方法进一步提高不同放大倍数的图像及子类的识别准确率。

参考文献:

- [1] DAR R A, RASOOL M, ASSAD A, et al. Breast cancer detection using deep learning: Datasets, methods, and challenges ahead [J]. Computers in biology and medicine, 2022, 149: 106073 - 106073.
- [2] 满 芮, 杨 萍, 季程雨, 等. 乳腺癌组织病理学图像分类方法研究综述 [J]. 计算机科学, 2020, 47 (S2): 145 - 150.
- [3] SPANHOL F A, OLIVEIRA L S, PETITJEAN C, et al. Breast cancer histopathological image classification using convolutional neural networks [C] // IEEE, 2016: 2560 - 2567.
- [4] HOU Y. Breast cancer pathological image classification based on deep learning [J]. Journal of X-ray Science and Technology,

