

## 基于跨域小样本分类的甲状腺超声图像识别模型

**摘要：**背景--通过对我国甲状腺癌现状分析，对于甲状腺癌发病率，中国平均水平略高于世界水平；对于医疗资源建设和利用水平较落后的地区，早期甲状腺癌的发现率要低于较发达地区。超声检查是一种高分辨率、高经济性、高安全性的无创检查方法，对于甲状腺内结节性病灶以及弥漫性病变具有较高的检查灵敏度。甲状腺超声图像是超声检查的结果，它可以反映甲状腺内结节性病灶的形态、大小、数目、内部结构等特征，帮助医生对病灶进行定性诊断，辅助判断其良恶性的可能性，为患者临床早期诊断和治疗提供帮助。因此，甲状腺超声图像对甲状腺内结节性病灶检查具有重要的意义。本文旨在通过分析临床甲状腺超声图像数据集，训练一种基于跨域小样本分类的集成学习模型来提高甲状腺癌诊断准确率和非典型恶性病灶检出率，建立网站来降低医疗资源落后地区诊疗的成本，提高居民甲状腺肿瘤的体检检出率。方法--我们使用 946 张甲状腺超声图像作为数据集，根据基于医学图像强细节性、强精准性等特征进行图像预处理，在前期融合了 **ResNet-18** 和 **EfficientNet-b5** 两个子算法模型组成的**集成学习模型**的基础上，我们优化了模型，采用**投票机制**对数据集进行训练和预测，其中 850 张提供良恶性标签的图像作为训练集、验证集，另外 96 张未知标签的图像作为测试集，选用五折交叉验证法对数据集进行划分、验证，将每张图像大小调整为  $224 \times 224$  像素，利用 **CLAHE 算法**增强局部对比度，减少噪声，后将图片进行随机水平翻转等数据增强处理和归一化处理。选择 **Adam** 算法作为优化器来优化模型对数据集的拟合，利用 **MSE 损失函数**作为计算 loss 值算法，最终以 F1-Score 作为评价指标来评价模型的性能。测试结果--我们设计和筛选了包含三个最优模型的基于硬投票的集成学习模型，模型准确率分别为 88.82%，90.00%，88.82%，F1-score 分别为 0.8902，0.8994，0.8876。

**关键词：**ResNet 模型；EfficientNet 模型；集成学习；CLAHE 算法；Adam 优化器；MSE 损失函数

# 一、研究背景

## 1.1 现状分析

首先，我们分析了 2021 年中国各地区基层医疗卫生机构（包括社区卫生服务中心（站）、街道卫生院、乡镇卫生院、村卫生室、门诊部、诊所）的数量，结果如下图所示：

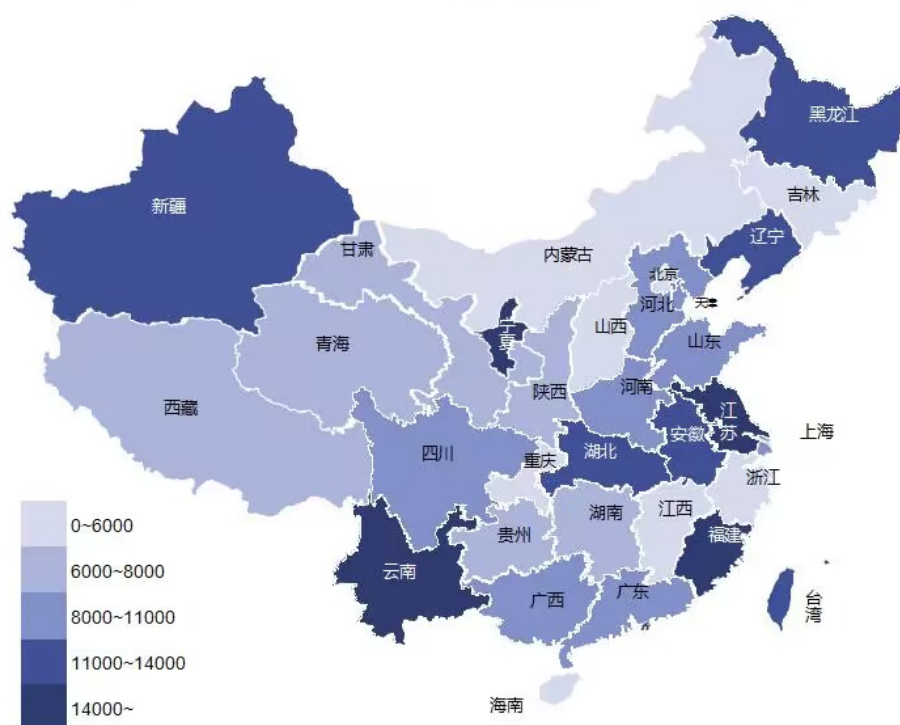


图 1. 中国 2021 年各地区基层医疗卫生机构数量分布热力图<sup>[1]</sup>

在医疗领域，甲状腺疾病是常见的问题。其中包括甲状腺良性疾病和甲状腺恶性疾病，甲状腺癌是内分泌系统最常见的恶性疾病。我们分析了近 20 年中国与世界甲状腺癌发病率，并将二者进行对比，结果如下图所示：

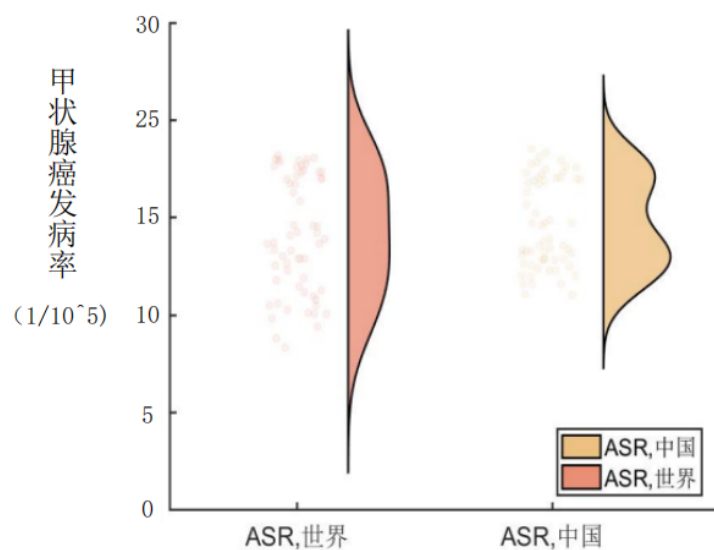


图 2. 近 20 年中国和世界甲状腺癌发病率对比云雨图<sup>[2]</sup>

针对以上结果分析，我们得出以下结论：近 20 年以来，中国甲状腺癌发病率平均水平略高于世界水平，中国甲状腺癌高发发病率的数集群水平高于世界平均水平。

由于甲状腺早期恶性肿瘤通常对仪器辅助诊断的依赖性很高，而部分基层地区设备设施的落后和医生经验不足可能会导致早期肿瘤检出率大幅度降低，因此我们产生了融合 ResNet 和 EfficientNet 的集成学习模型帮助判断甲状腺肿瘤超声图像良恶性的想法，同时建立了网站，能够供用户上传超声图像并获得模型判断结果，帮助提高医疗资源相对较差的地区早期甲状腺癌诊断率。

## 1.2 针对基层医疗的背景调查

### 1.2.1 中国基层医疗资源建设与利用水平评价体系

为了调查中国各个地区基层医疗资源的相对水平，我们建立了一套 2017-2021 年中国基层医疗资源建设与利用水平评价体系，具体的建模指标如下表所示：

表 1. 2017-2021 中国基层医疗资源建设与利用水平评价体系<sup>[1]</sup>

中国基层医疗资源建设与利用水平	准则层	指标层		正负向
		一级指标	二级指标	
	资源投入层	设施	各地区基层医疗卫生机构数量	正向指标
			各地区村卫生室数量	正向指标
			各地区乡镇卫生院数量	正向指标
			各地区社区卫生服务中心(站)数量	正向指标
			各地区基层医疗卫生机构床位数	正向指标
		人员分配	各地区基层医疗卫生机构医疗人员密度	正向指标
			各地区村卫生室医疗人员密度	正向指标
			各地区乡镇卫生院医师日均负担	负向指标
			各地区社区卫生服务站医师日均负担	负向指标
	资源利用层	诊疗情况	各地区基层医疗卫生机构平均诊疗人次	正向指标
			各地区乡镇卫生院病床使用率	正向指标
			各地区社区卫生服务站病床周转率	负向指标
		医疗消费	各地区县属及以下人均卫生总费用	正向指标

1.2.2 熵权法

熵权法利用各个指标所含信息熵的大小来赋权，某一指标的信息熵越小说明该指标的变异程度越大，提供的信息就更多，理应赋予更大的权重，该方法通过衡量指标变异程度的大小确定权重的大小赋予各个指标较为客观的权重。熵权法的原理及公式如下所示：

（1）数据标准化处理。通过将各个指标数据标准化来消除各个指标因为量纲不同可能产生的偏差，数据标准化方法分为两种分别对应于正向指标与负向指标，其中正向指标同基层医疗资源建设与利用水平呈正相关，负向指标同基层医疗资源建设与利用水平呈负相关，例如：各地区乡镇卫生院医师日均负担。

对于正相关指标采用的标准化处理方法为：

$$x_{ij}^{+} = \frac{x_{ij} - \min_{1 \leq i \leq m} \{x_{ij}\}}{\max_{1 \leq i \leq m} \{x_{ij}\} - \min_{1 \leq i \leq m} \{x_{ij}\}} \tag{1,1}$$

对于负相关指标采用的标准化处理方法为：

$$x_{ij}^{+} = \frac{x_{ij} - \min_{1 \leq i \leq m} \{x_{ij}\}}{\max_{1 \leq i \leq m} \{x_{ij}\} - \min_{1 \leq i \leq m} \{x_{ij}\}} \tag{1,2}$$

（2）计算各个指标数据的的信息熵。第 i 个指标的信息熵为：

$$e_i = -k \sum_{j=1}^m \frac{x_{ij}}{\sum_{j=1}^m x_{ij}} \ln \frac{x_{ij}}{\sum_{j=1}^m x_{ij}}, k = \frac{1}{\ln m}, e_i \geq 0 \tag{1,3}$$

(3) 利用信息熵的大小求出熵权大小。第  $i$  个指标的熵权为:

$$u_i = \frac{1-e_i}{n-\sum_{i=1}^n e_i} \quad (1,4)$$

然后,在上述评价体系的建立基础上,我们采用熵权法确定指标的信息熵和权重。2017-2021 中国基层医疗资源建设与利用水平评价体系中各个指标的信息熵和权重如下表所示:

表 2. 2017-2021 中国基层医疗资源建设与利用水平评价体系中各个指标的信息熵和权重<sup>[1]</sup>

一级指标总权重	一级指标	二级指标	二级指标权重	信息熵
0.3576	设施	各地区基层医疗卫生机构数量	0.1234	2.0360
		各地区村卫生室数量	0.0455	2.1392
		各地区乡镇卫生院数量	0.0087	1.9435
		各地区社区卫生服务中心(站)数量	0.1027	1.9779
		各地区基层医疗卫生机构床位数	0.0773	1.9877
0.2488	人员分配	各地区基层医疗卫生机构医疗人员密度	0.1031	2.0775
		各地区村卫生室医疗人员密度	0.0043	1.9938
		各地区乡镇卫生院医师日均负担	0.1	1.5690
		各地区社区卫生服务站医师日均负担	0.0414	2.0750
0.2499	诊疗情况	各地区基层医疗卫生机构平均诊疗人次	0.1099	1.8314
		各地区乡镇卫生院病床使用率	0.0998	2.1570
		各地区社区卫生服务站病床周转率	0.0402	1.9652
0.1437	医疗消费	各地区县属及以下人均卫生总费用	0.1437	2.1050

最后,我们利用线性加权取得最终基层医疗资源建设与利用水平结果评分,具体的公式如下:

$$S = \sum_{j=1}^n (a_j x'_{ij}) \quad (1,5)$$

### 1.2.3 模糊聚类

根据上述最基于模糊关系的模糊聚类分析是一种针对所研究的对象按某种标准进行模糊分类的数学方法,它是进行信息综合评价的基础之一。模糊聚类的原理和公式如下所示:

(1) 数据标准化。建立原始数据矩阵,利用  $n$  个省份 ( $x_1, x_2, \dots, x_n$ ) 的  $m$  个指标数据建立原始矩阵如下:

对原始数据矩阵进行平移-极差变换,公式如下:

$$x'' = \frac{x'_{ik} - \min_{1 \leq i \leq n} \{x'_{jk}\}}{\max_{1 \leq i \leq n} \{x'_{jk}\} - \min_{1 \leq i \leq n} \{x'_{jk}\}} \quad (1,6)$$

(2) 建立模糊相似矩阵。我国各个省份的医疗支持方式及各建设方案的效果由于地理、政策、历史等原因差异较大因此本文采用指数相似系数法来构建模糊相似矩阵。指数相似系数法公式如下：

$$r_{jj} = \frac{1}{m} \sum_{k=1}^m e^{\left\{ -\frac{3}{4} \times \frac{(x_{ij} - x_{jk})^2}{S_k^2} \right\}} \quad (1,7)$$

$$\text{其中, } S_k = \frac{1}{n} \sum_{k=1}^m (x_{ij} - \bar{x}_k)^2, x_k = \frac{1}{n} \sum_{k=1}^m x_{ik}, (k = 1, 2, \dots, m) \quad (1,8)$$

(3) 根据模糊相似矩阵采用直接聚类法进行聚类。根据上述评分结果，我们采用了基于模糊关系的模糊聚类分析法，对中国 31 省份进行了分类，聚类结果如下表所示：

表 3. 中国 31 省份医疗资源建设和利用水平聚类分析结果<sup>[1]</sup>

类别	省份
第一类	河北 山西 内蒙古 吉林 福建 广西 海南 贵州 西藏 甘肃 青海 宁夏 新疆
第二类	黑龙江 江西 河南 重庆 云南
第三类	天津 辽宁 安徽 山东 湖北 湖南 四川 陕西
第四类	北京 上海 江苏 浙江 广东

第一类省份的平均评价得分为 0.5201，第二类省份的平均评价得分为 0.5044，第三类省份的平均评价得分为 0.7208，第四类省份的平均评价得分为 0.8466。

由于一二类省份与三四类省份间差异较大，我们将三四类省份归为甲类地区（医疗资源建设与利用水平较高的地区），将一二类省份归为乙类地区（医疗资源建设与利用水平较低的地区）。

我们将甲、乙两个地区甲状腺癌体检检出人次占总发病人次的比率进行对比，输出结果如下图所示：

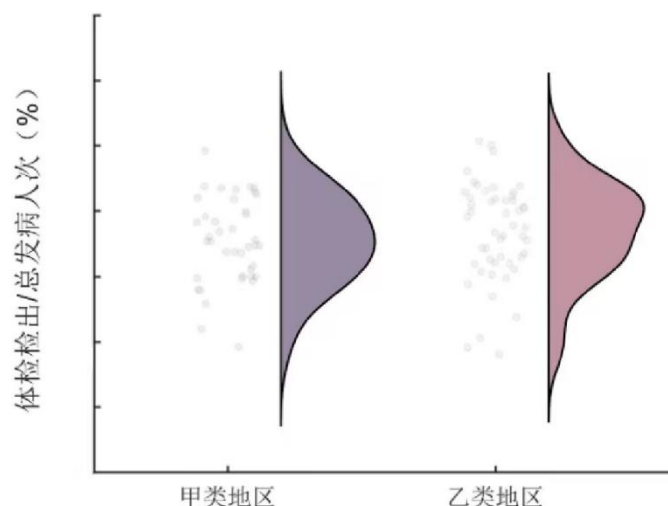


图 3. 甲乙两类地区甲状腺癌体检检出人次占总发病人次的比率对比云雨图

针对以上结果分析，我们得出以下结论：由于大部分甲状腺癌是在体检中发现的，而医疗资源建设和利用水平较落后的地区，人们的体检意识大多薄弱，早期甲状腺癌的发现率要低于较发达地区。

基于以上结论，我们认为，建立一个易于使用的甲状腺医学图像分类模型在线影像识别系统和联结基层医生与更富有经验医生的医学影像分析联盟，为提高甲状腺癌检出率和提高患者预后及幸福感提供新的思路。

## 二、问题定义与分析

### 2.1 问题定义

超声检查是一种常见的用于甲状腺疾病咨询的无创检查方法，能够通过声波图像显示甲状腺的大小、形状、位置、质地、边缘、内部结构等特征。超声检查在病理特征上区别良性和恶性肿瘤具有更为显著的能力，为甲状腺癌临床早期诊断和治疗提供极大的帮助。

根据上述现状分析和甲状腺超声图像的特点，我们总结出目前甲状腺癌的超声检查存在以下问题：

(1) 在近 20 年中，中国甲状腺癌发病率较世界甲状腺发病率平均水平增长更为显著。

(2) 对于医疗水平、医疗资源相对落后的地区，居民健康管理意识较低，体检人数较少，并且居民拥有较少的途径来获取医疗资源，可能承担不了较高的医疗卫生花费。

(3) 甲状腺癌发病隐匿，早期常无明显症状，病人多以甲状腺结节为首发症状，甲状腺结节有良性和恶性之分，如果未及时准确判断甲状腺结节的良恶性，会造成甲状腺病人的误诊和延迟诊断。

(4) 由于涉及到医学专业知识和临床经验，超声医师对于甲状腺超声图像的诊断具有主观性，并且超声图像存在分辨率和对比度较低、固有斑点噪声过大等问题。

## 2.2 问题分析

随着科技的不断进步，人工智能算法逐渐引起人们的重视，计算机辅助的肿瘤定位将降低误诊率，为判断甲状腺病变的性质判别提供了新的方法。为了解决甲状腺超声图像的误诊和分辨率和对比度较低的问题，我们基于甲状腺超声图像，训练一个人工智能算法，提供一个可靠的、自动的、快速的解决方案（包括分类和判断），可以在临床应用中帮助医生准确诊断甲状腺病变的良恶性。

为了提高模型性能，避免部分外部因素对模型的干扰，训练之前需要对超声图像进行预处理，并进行有效的特征提取，我们利用神经网络的深度学习算法，设计一种集成学习模型，结合图像分析技术和医疗知识对图像进行训练。我们的数据集由 850 张提供良恶性标注的甲状腺超声图像（410 张良性病变图像和 440 张恶性病变图像，训练集）和 96 张甲状腺超声图像（测试集）组成，我们期望模型可以对甲状腺超声影像正确的分类，并在未知的数据上进行准确的预测，帮助医生在更短的时间内诊断患者的甲状腺良性或恶性疾病。本次比赛中，我们将提交一个基于深度学习的集成学习模型。

在此基础上，为了降低医疗资源落后地区诊疗的成本，提高居民甲状腺肿瘤的体检检出率，我们将在建立的网站上开通邮件功能，打通该地区医生与更有经验的医生的沟通渠道，使得甲状腺癌可以及早发现和进行治疗。



## 三、数据分析及预处理

### 3.1 数据评价指标

F1-Score 又称为平衡 F 分数 (Balanced F Score)，是统计学中用来衡量二分类模型精确度的一种指标，本次竞赛中 F1-Score 作为最终测评的方法。它是精确率和召回率的调和平均数，利用 F1-Score 可以区别算法和模型的优劣。它的取值范围是 0 到 1，1 代表本次测试模型的输出最好，0 代表本次测试模型的输出结果最差。F1-Score 的计算公式为：

$$F1 - Score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 * (precision * recall)}{(precision + recall)} \quad (3.1)$$

2.1 式中，precision 表示模型的精确率，recall 表示模型的召回率。利用数据得出二分类结果的混淆矩阵，如下表所示：

表 4. 二分类结果的混淆矩阵 (Confusion matrix)		
实际情况	预测结果	
	正例	负例
正例	真正例	假负例
负例	假正例	真负例

真正例 (True Positive, TP)：表示被模型预测为正例的正样本。

真负例 (True Negative, TN)：表示被模型预测为负例的负样本。

假正例 (False Positive, FP)：表示被模型预测为正例的负样本。

假负例 (False Negative, FN)：表示被模型预测为负例的正样本。

(1) 通过第一步的统计值计算精确度、召回率和准确率。上述数值计算公式如下：

$$precision = \frac{TP}{TP + FP} \quad (3.2)$$

$$recall = \frac{TP}{TP + FN} \quad (3.3)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4)$$

精确率 (precision) 表示被分类器判定为正例中正样本的比重，召回率 (recall) 表示被预测为正例的占实际总的正例的比重，准确率 (accuracy) 表示分类器对整个样本预测正确的比重。

(2) 通过第二步的计算结果计算得出 F1-Score，计算公式如下：

$$F1 - Score = \frac{2*(precision*recall)}{(precision+recall)} \quad (3.5)$$

## 3.2 模型参数

Epoch, Batch, Iterations 均为神经网络超参数，三种参数的定义如下表所示：

表 5. 三种神经网络超参数的定义

名词	定义
Epoch	使用训练集的全部数据对模型进行一次完整的训练，被称为“一代训练”
Batch	使用训练集中的一小部分样本对模型权重进行一次反向传播的参数更新，这一小部分样本被称为“一批数据”
Iteration	使用一个 Batch 对模型进行一次参数更新的过程，被称为“一次训练”

Epoch 表示训练中遍历训练集中所有样本的次数，在本次梯度下降的模型训练过程中，神经网络逐渐从不拟合状态到优化状态，达到最优状态后若迭代次数持续增加，则会进入过拟合状态。因此 Epoch 的取值并非越大越好，一般取值在 50-200 之间，数据多样性增强，相对应的 Epoch 值也会增加。

我们通过把全部训练数据集分为若干个批（Batch），按照批来更新参数。每个 Batch 中样本的个数叫做样本大小（Batchsize）。Batchsize 最大的是样本总数 N，即全数据集（Full batch learning）形式；最小是 1，即每次训练一个样本，这就被称为在线学习（Online Learning）。当我们处理较大的数据集时，一次向网络喂入全部的数据不一定会得到很好的训练效果，因此一般设置在 50-400 左右比较适宜。

一个 Iteration 等于使用 Batchsize 个样本训练一次，也称为迭代。每一次迭代都是一次权重更新，每一次权重更新都需要 Batchsize 个数据进行 Forward 运算得到损失函数。

Epoch, Batch, Iterations 的换算公式如下：

$$One\ Epoch = Numbers\ of\ Iterations = N = \frac{\text{训练样本的数量}}{Batchsize} \quad (3.6)$$

### 3.3 数据预处理

针对甲状腺超声图像数据集的预处理，我们组选择五折交叉验证法。五折交叉验证法是一种常用的机器学习模型评估和参数选择的方法。它的基本思想是将数据集平均分成五份，每次用其中一份作为测试集，其他四份作为训练集，训练模型并计算测试误差。重复这个过程五次，然后取五次测试误差的平均值作为最终的评估指标。五折交叉验证法的优点是可以充分利用数据，避免了单次划分训练集和测试集可能带来的偏差和方差。它也可以用于比较不同模型或参数的性能，选择最优的模型或参数。

甲状腺超声图像数据集预处理的过程如下：

(1) 我们在 `src/functions.py` 中使用 `five_fold_cross_validation()` 函数实现五折交叉验证法。我们将赛方提供的所有图像以 1 为起点进行编号，前 410 项为良性病例，后 440 项为恶性病例。将其按照 1-850 进行编号。然后将编号进行随机打乱后分为五组，每次选用其中一组进行实验。

(2) 使用 `augmentations.Resize(224, 224)` 将图片缩放到 224x224 像素的大小。

(3) 使用 `augmentations.CLAHE(clip_limit=4.0, tile_grid_size=(4, 4), p=0.9)` 对图片进行对比度受限的自适应直方图均衡化，其中 `clip_limit` 表示对比度的限制程度，`tile_grid_size` 表示分割图片的网格大小，`p` 表示应用该操作的概率。

(4) 使用 `augmentations.Normalize(mean=(0.485, 0.456, 0.406), std=(0.229, 0.224, 0.225), max_pixel_value=255.0, p=1.0)` 对图片进行归一化，其中 `mean` 和 `std` 分别表示三个通道（红、绿、蓝）的均值和标准差，`max_pixel_value` 表示像素值的最大值，`p` 表示应用该操作的概率。

(5) 使用 `ToTensorV2()` 将图片转换为 PyTorch 张量，并将通道顺序从 RGB 变为 BGR。

(6) 在训练集中，还会使用 `augmentations.HorizontalFlip(p=0.5)` 随机地对图片进行水平翻转，其中 `p` 表示应用该操作的概率。

## 四、模型设计与选择

### 4.1 CLAHE 算法

#### 4.1.1 CLAHE 算法的采用

我们研究的是甲状腺肿瘤超声图像，对于肿瘤显像的边缘特征、内部回声等细节要求较高。简单的直方图均衡算法（Histogram Equalization, HE）能够有效提高图像全局对比度，通过这种方法，亮度可以更好地在直方图上分布，这样就可以用于增强局部的对比度而不影响整体的对比度。但直方图均衡算法在计算量和噪声处理方面存在不足，可能会增加背景噪声的对比度并且降低有用信号的对比度。

比较而言，我们采用自适应直方图均衡化（Adaptive Histogram Equalization, AHE）的方法，能够调整图像亮度分布和增强程度，对边缘等细节特征进行有效处理，但是由于图像被等分为几块分别处理，这会导致算法的时间复杂度提高，图像处理的效率也会有所降低。

在 HE 和 AHE 算法基础上，我们最终选择**限制对比度自适应直方图均衡算法（Contrast Limited Adaptive Histogram Equalization, CLAHE）**，它继承了 AHE 的全部优势，同时避免图像失真、减弱噪声放大产生的负面影响。

#### 4.1.2 CLAHE 算法的原理

CLAHE 算法的公式及原理如下：

（1）分块。输入一张图像，对其进行均匀分割，得到大小相等的子区域，并假设各子区域内包含的所有像素值和为 S。

（2）计算直方图。以  $U_{ab}(k)$  表示各个子区域的直方图，其中 k 表示局部子区域的灰度级数，取值范围为  $[0, N-1]$ ，N 表示该子区域内的最大灰度级数。

（3）计算受限值。

$$\beta = \frac{S}{N} \left\{ 1 + \frac{\alpha}{100} (l_{max} - 1) \right\} \quad (4.1)$$

其中， $\alpha$  表示截断系数，取值范围是  $[0, 100]$ ； $l_{max}$  为最大斜率，取值范围为  $[1, 4]$ ， $l_{max}$  的取值决定了对比度增强的幅度。

(4) 像素点重分配。每一块子域都有对应的受限值 $\beta$ ，利用 $\beta$ 截取该子域的直方图  $U_{ab}(k)$ ，并得到截取总数为：

$$E_a = U_{ab}(k) - \beta \quad (4.2)$$

截取的像素总数需要平均分配到直方图的各灰度级中，计算平均分配的像素数为：

$$b_v = E_a / N \quad (4.3)$$

计算分配剩余像素数步长为：

$$L = L_a / E_a \quad (4.4)$$

其中， $L_a$ 为灰度范围长度。循环检测直方图中各灰度级数的大小，当其值小于该区域内受限值 $\beta$ ，为其分配平均像素数 $b_v$ 。重复进行分配操作，直到截取的像素总数被分配完成。裁剪分配的过程如下图所示：

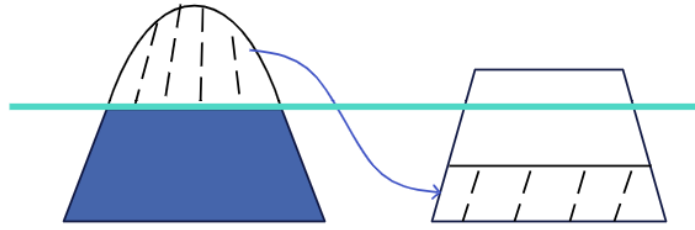


图 4. CLAHE 算法裁剪分配过程

(5) 直方图均衡。将对比度受限后每个子块的灰度直方图进行均衡化处理。

(6) 双线性插值进行灰度值重构。对经过均衡处理后的直方图进行灰度双线性插值。

假设函数 $f(x)$ 有 4 个已知点的值分别为  $P_{11}(x_1, y_1)$ 、 $P_{12}(x_1, y_2)$ 、 $P_{21}(x_2, y_1)$ 、 $P_{22}(x_2, y_2)$ ，则在函数 $f(x)$ 上任意点 $p = (x, y)$ 可以通过线性插值近似求出。

在  $x$  方向上进行线性插值，得到：

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(P_{11}) + \frac{x - x_1}{x_2 - x_1} f(P_{12}) \quad (4.5)$$

$$R_1 = (x, y_1) \quad (4.6)$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(P_{21}) + \frac{x - x_1}{x_2 - x_1} f(P_{22}) \quad (4.7)$$

$$R_2 = (x, y_2) \quad (4.8)$$

在  $y$  方向上进行线性插值，得到：

$$f(p) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \quad (4.9)$$

$$R_1 = (x, y_1) \tag{4.10}$$

$$R_2 = (x, y_2) \tag{4.11}$$

其中， $f(P_{11})$ 、 $f(P_{12})$ 、 $f(P_{21})$ 和 $f(P_{22})$ 表示经过直方图均衡化后，4个相邻区域中的直方图的灰度值。利用上述的方法，计算输出图像各点的灰度值，实现像素点灰度值的重构。

上述 CLAHE 算法的流程如下图所示：

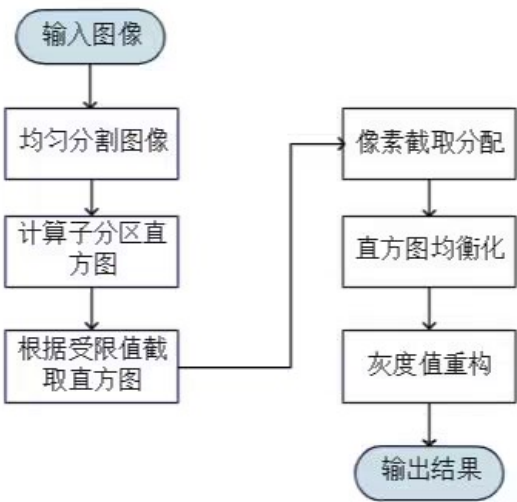


图 5. CLAHE 算法的流程图

### 4.1.3 CLAHE 算法的选择

我们从甲状腺超声图像数据集中选择第 83 号良性肿瘤超声图像和第 600 号恶性肿瘤作为测试的案例，分别采用 HE 算法和 CLAHE 算法运行，运行后的对比结果如下：

第 83 号良性肿瘤超声图像：

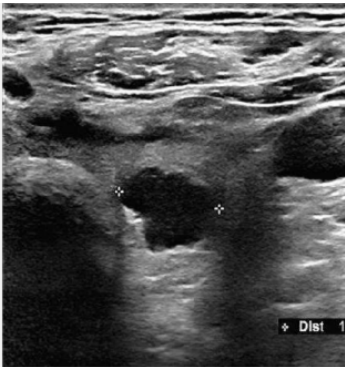


图 6-1 原始图像

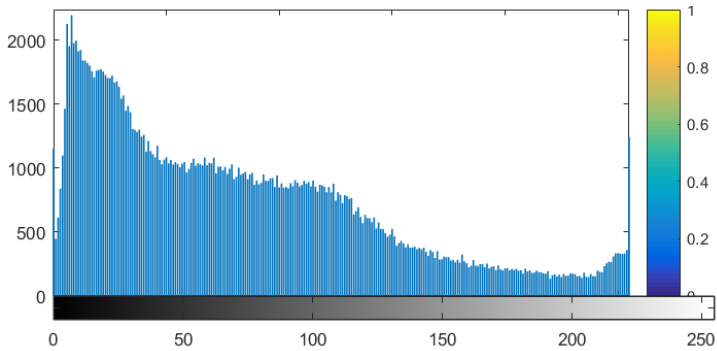


图 6-2 原始灰度直方图



图 7-1. 传统 HE 算法处理

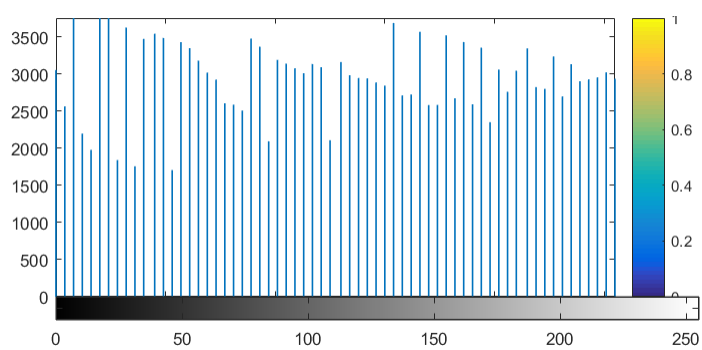


图 7-2. 传统 HE 算法均衡化处理后直方图

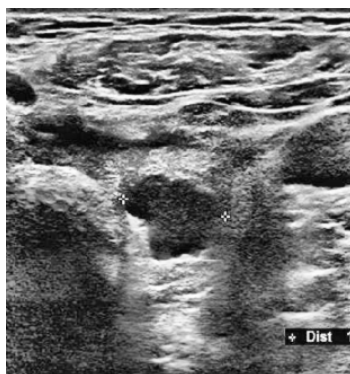


图 8-1. CLAHE 处理

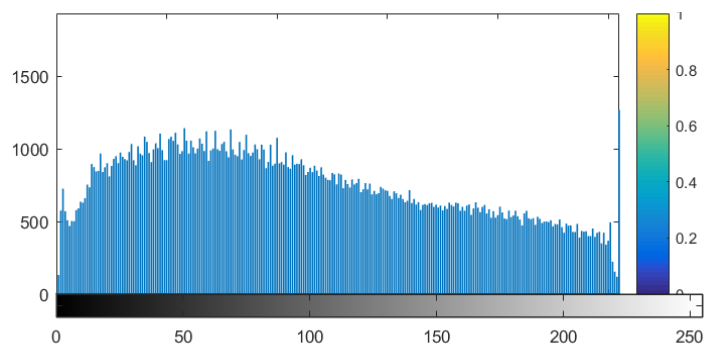


图 8-2. CLAHE 算法均衡化处理后直方图

第 600 号恶性肿瘤:



图 9-1.原始图像

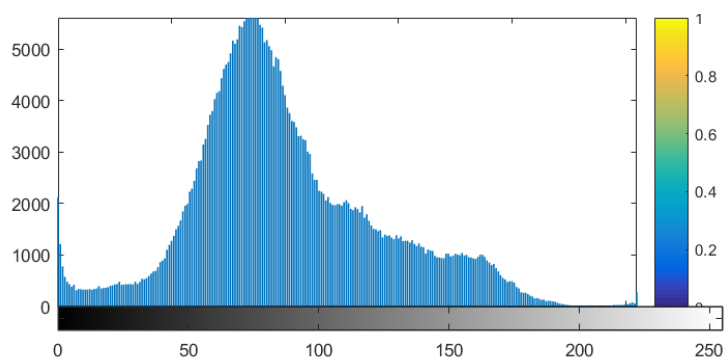


图 9-2.原始灰度直方图



图 10-1. 传统 HE 算法处理

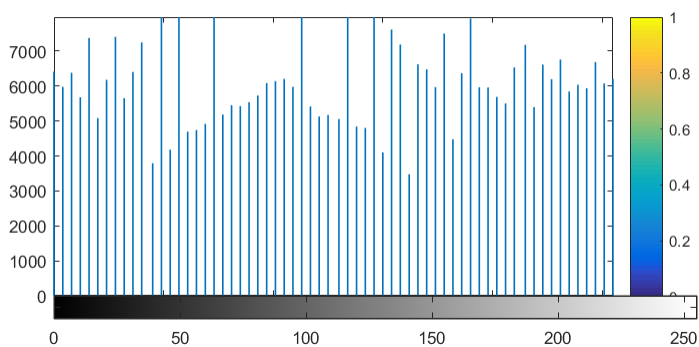


图 10-2. 传统 HE 算法均衡化处理后直方图



图 11-1. CLAHE 处理

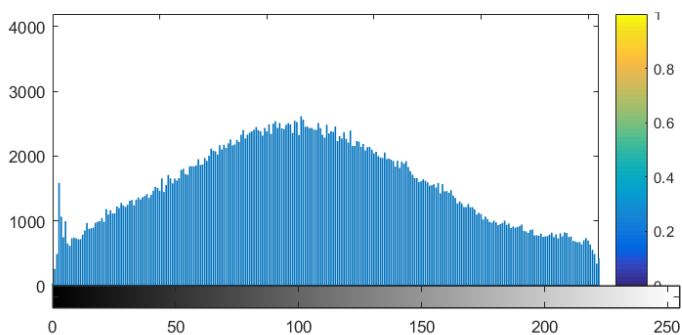


图 11-2. CLAHE 算法均衡化处理后直方图

通过对比 HE 算法和 CLAHE 算法处理后的图像及灰度直方图,我们组总结出 HE 算法和 CLAHE 算法的优劣如下图所示:



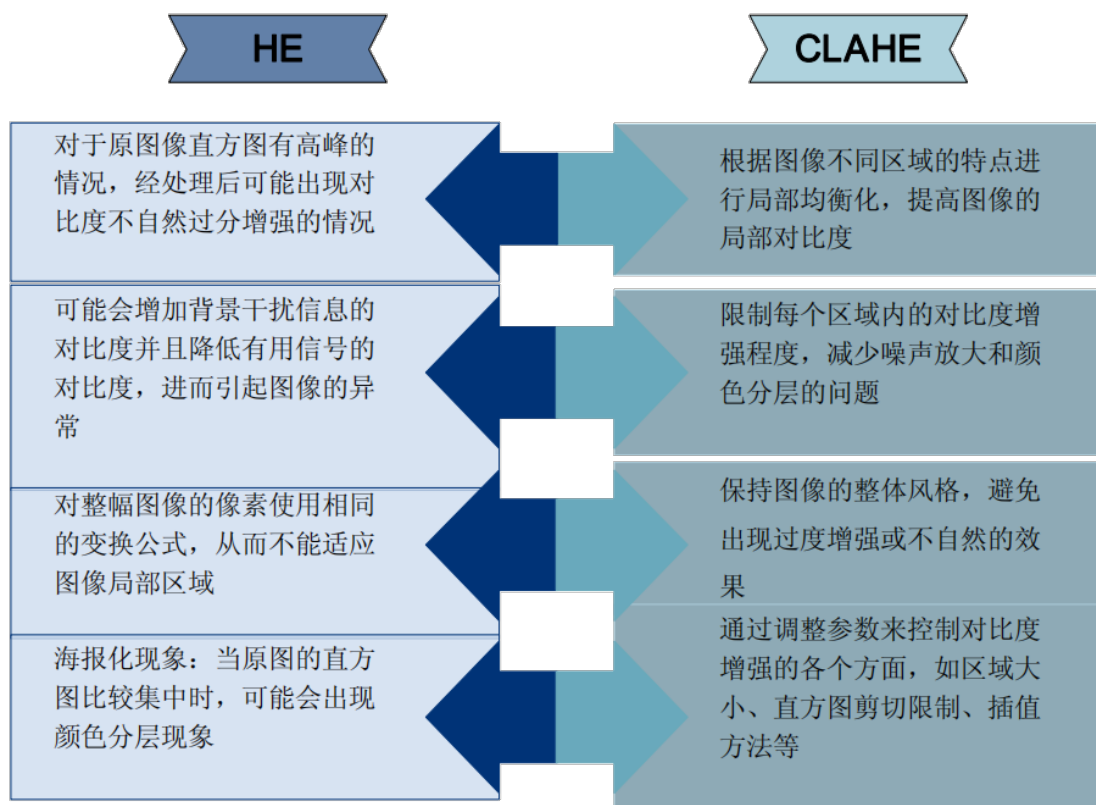


图 12. HE 算法与 CLAHE 算法优势对比图

## 4.2 ResNet 模型

### 4.2.1 ResNet 模型的采用

我们知道加深层对提升性能很重要，加深层可以使模型在训练集和测试集上的性能更好。这是由于模型的复杂度的提高，会导致模型的表达能力更强，因此对于潜在的映射关系会拟合的更好。但是，在深度神经网络中，如果过度加深层的话，很多情况下学习将不能顺利进行，导致最终性能不佳。为了解决深度网络中这种“退化”的问题，我们选择残差神经网络（Residual Network，ResNet）作为本次甲状腺超声图像自动分类的模型。

为了解决深度神经网络中“退化”的问题，在 ResNet 中，导入了“快捷结构”。通过引入这种快捷结构，即使加深层，也能持续高效地学习（当然，层的加深也是有限度的）。快捷结构只是原封不动地传递输入数据，所以反向传播时会将来自上游的梯度原封不动地传向下游。因此，基于快捷结构，之前因为加深层而导致的梯度变小的梯度消失问题就有望得到缓解。

## 4.2.2 模型的架构设计

我们将 ResNet 划分为 5 个构建层（Building Layer），其中一个构建层可以包含 1 个或多个网络层以及构建块。ResNet 模型的架构图如下图所示：

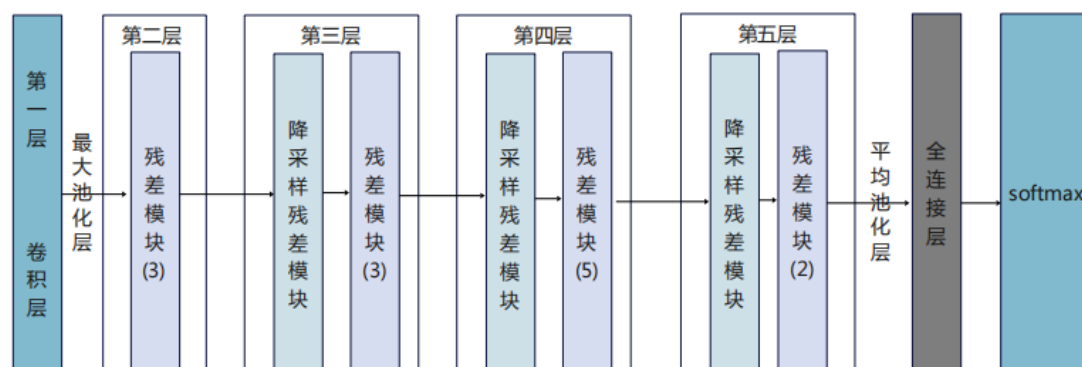


图 13. Resnet 模型架构图

- (1) 第一个构建层，是由 1 个卷积层和最大池化层构成。
- (2) 第二个构建层，是由 3 个残差模块构成。
- (3) 第三、第四、第五构建层，都是由降采样残差模块开始，紧接着 3 个、5 个、2 个残差模块。

## 4.3 EfficientNet 模型

EfficientNet 是 Google 研究小组 2019 年在论文《EfficientNet: Rethinking Model Sealing for Convolutional Natural Networks》中提出的模型，该模型基于网络深度（depth）、宽度（width）和输入分辨率三个维度的缩放来寻找最优模型。

- 增加网络的深度能够处理更加丰富、复杂的特征，但是网络过深会面临梯度消失、训练困难的问题。
- 增加网络的宽度能够获得更高细粒度的特征并且也更容易训练，但是宽度较大而深度较浅的网络往往很难学习到更深层次的特征。
- 增加输入网络的图像分辨率能够使得网络学习与表达的内容更加丰富，有利于提高准确度；但是对于非常高的输入分辨率，准确率的增益存在递减效应，且大分辨率图像会使计算量增加。

EfficientNet 能够调整深度、宽度、输入分辨率到最佳状态，集成上述几个方面中的优势，完成一个优秀的网络设计。

此外，提高神经网络的性能不仅仅在于堆叠层数，更重要的是网络要可以训练和收敛，参数量要小并且方便训练。EfficientNet 满足上述的特点，可以利用更少的参数量进行最好的识别效率，提高网络的性能。

## 4.4 集成学习模型

### 4.4.1 集成学习模型的设计

集成学习（Ensemble Learning）是一种有效的机器学习技术，可以提高分类精度并增强模型鲁棒性。集成学习的核心思想是，整合多个不同的模型，通过加权平均或投票等方式获得最终预测结果。集成多个模型可以有效降低过拟合的风险，提高模型的泛化能力和稳定性。

EfficientNet<sup>[3]</sup>是最新的卷积神经网络模型之一，最初用于图像分类任务。ResNet<sup>[4]</sup>则是一个经典的卷积神经网络模型，其通过残差连接（Residual Connection）减小了梯度消失问题。这两个模型都以其高效的特性在图像分类领域取得了成果，这两个模型都以其高效的特性在图像分类领域取得了成果。

在本次竞赛图像分类应用中，我们最终设计了两个卷积神经网络模型组成的集成学习模型，即选用 ResNet-18 和 EfficientNet-B5，使得我们的模型拥有集成学习的能力（同一个预测任务用多个内部模型预测，取最好结果的模型输出预测结果），拥有更高的性能。

子模型融合的方式使得模型更加具有灵活性和适应性，能够适应各种不同的数据和任务。此外，由于我们采用了多个模型的输出，所以我们可以根据模型的表现和效果更好地对模型进行优化和调节。

### 4.4.2 集成学习模型的选择

ResNet 和 EfficientNet 这两个模型都是经典卷积神经网络模型，它们在迭代的各个阶段中都有不同的表现。

**ResNet-18:** 轻量化模型，有效避免过拟合问题，缺点是处理复杂图像能力较差；在刚开始的几次迭代中，误差率下降得比较快。接着，随着迭代次数的增加，误差率下降的速度会变慢，说明模型的收敛速度逐渐变慢，需要更多的迭代来到达最优解。

**EfficientNet-B5**：在图片相对复杂、含信息熵较多的情况下，采用 EfficientNet-b5 进行预测，由于 EfficientNet-B5 的权重规模更大、隐藏层更多，可以提取更多图片信息，缺点是容易出现过拟合问题。EfficientNet-B5 模型在开始迭代时，误差率可能存在放大的情况，也就是误差率会在前几个迭代中有小幅上升的趋势，然后逐渐下降。在后续的训练中，EfficientNet-B5 的误差率下降速度将逐渐加快，直到误差率收敛到一个相对稳定的水平。

采用两种子模型集成的模型可提升模型鲁棒性、泛化能力。ResNet-18 和 EfficientNet-B5 子模型随迭代次数增加时误差率的热力图如下图所示：

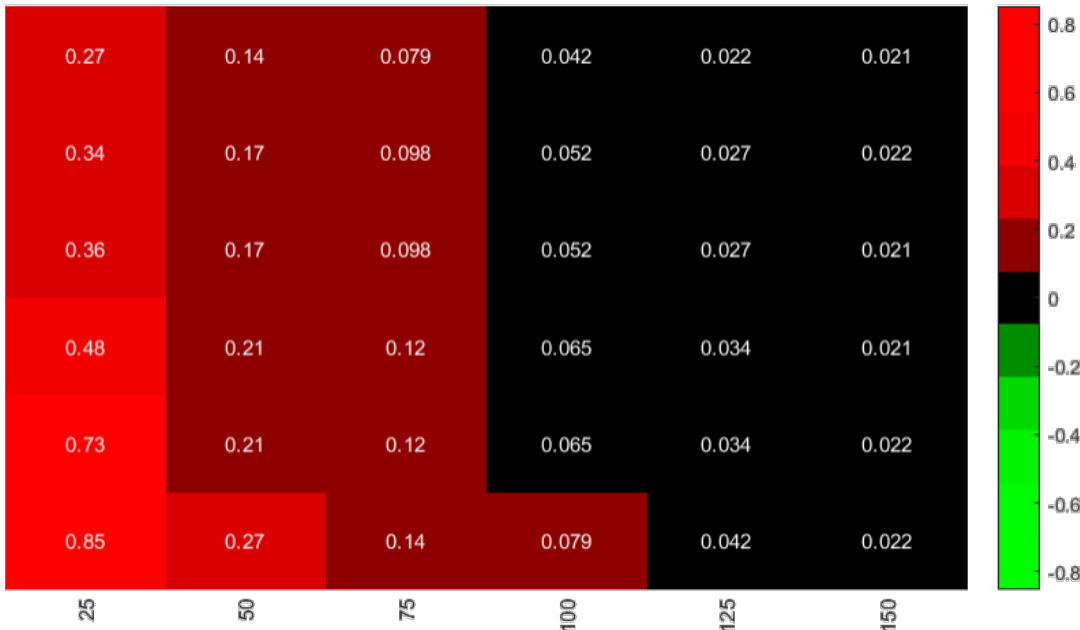


图 14. ResNet-18 模型随迭代次数增加误差率的变化热力图

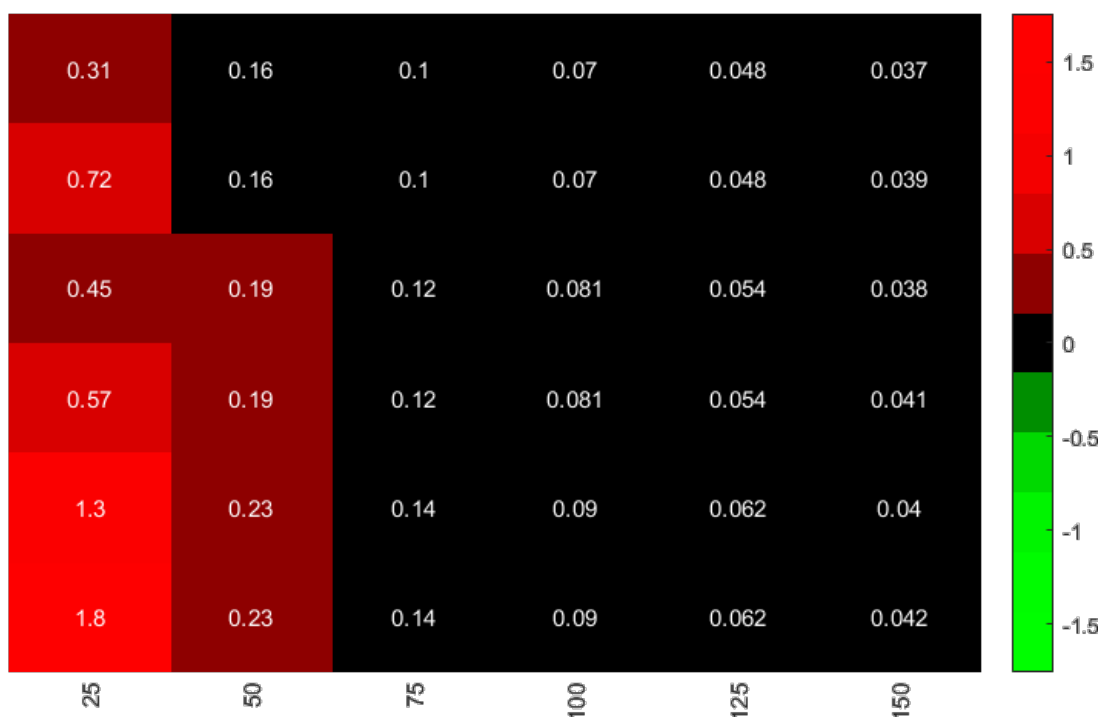


图 15. EfficientNet-b5 模型随迭代次数增加误差率的变化热力图

综上所述，我们的集成学习模型具有明显的优点，适用于多种场景，并可作为一个有效的工具来提高在图像分类领域中的性能和表现。我们组选择集成学习模型具有以下优势：

（1）提高了模型的准确性：由于将不同的模型集成在一起，避免了单个模型可能出现的偏差或错误，提高了模型的准确性。

（2）增强了模型的泛化能力：集成多个模型并通过投票或加权平均方式获得结果，可以降低过拟合的风险，增强模型的泛化能力。

（3）提高了模型的鲁棒性：集成不同的模型，可以使得模型更加稳健和可靠，并避免了单个模型可能出现的数据特异性或不完整性。

（4）灵活、高效、可扩展：我们的集成学习模型灵活、高效，并且可扩展，可以适用于各种不同的图像分类任务和不同规模的数据集。

#### 4.4.3 集成学习模型的优化方案探索

为了进一步提高模型准确度，提升模型应用推广的可行性，我们针对目前模型的可行优化方案进行了一系列探索。我们考虑到提高模型准确性、减少过拟合是我们最主要的需求，计划引进更多的模型结构，建立一种投票机制。

投票机制是集成学习中常用的一种方法，可以将多个弱学习器的预测结果进行投票或加权投票，最终决定集成模型的预测结果。具体的投票机制模型选择我们将在第五章详细展开。



图 16. 基于投票机制的集成学习模型相较于普通集成学习模型的优势

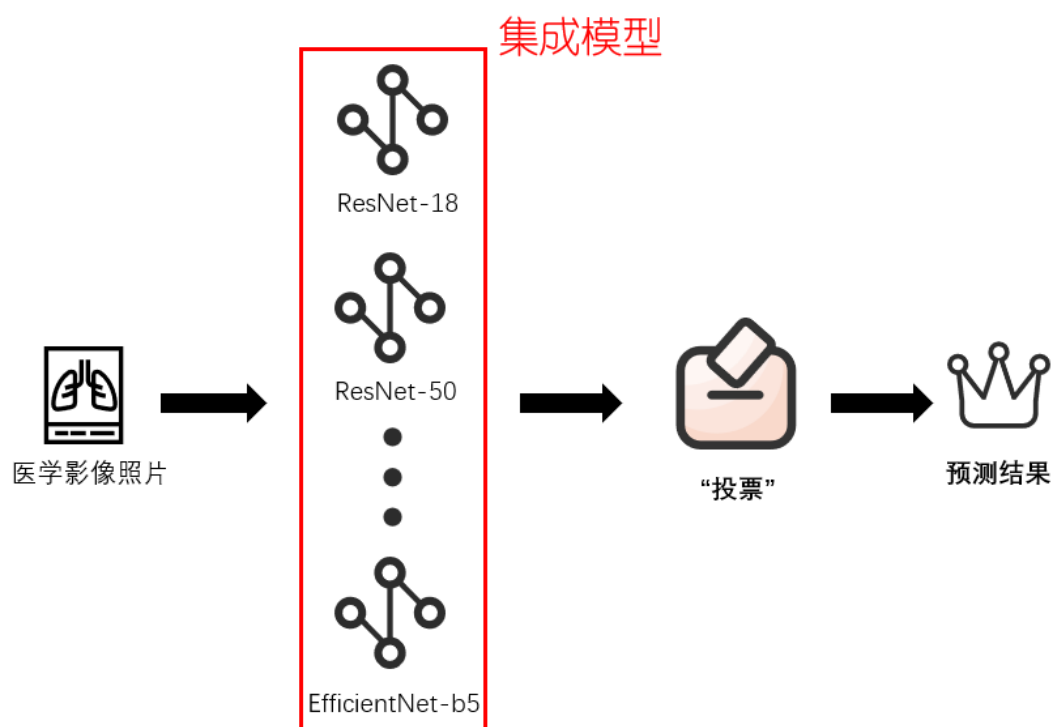


图 17.投票机制示意图

#### 4.5 Adam 优化器<sup>[5]</sup>

### 4.5.1 Adam 优化器的采用

在深度学习中，我们的目标是通过不断改变网络参数，使得参数能够对输入做各种非线性变化的拟合输出。我们通常将更新参数的算法称为优化器，它是通过训练优化参数，来最小化（最大化）损失参数。因此，为了使我们的模型逼近或达到最优值，我们需要各种优化器来更新和计算影响模型训练和输出的网络参数。

随机梯度下降（Stochastic Gradient Descent, SGD）是一种经典的优化器，通过梯度下降的方法，不断调整模型的参数，使模型的损失函数最小化。SGD 运算简单，通过少量的计算量就可以得到很好的结果，但是它的收敛速度会受到学习率的影响，容易陷入局部最小值。

针对 SGD 的缺点，我们最终选择自适应矩估计（Adaptive Moment Estimation, Adam）作为本次模型的优化器。Adam 优化器是一种对随机梯度下降法（Stochastic Gradient Descent, SGD）的扩展，可以对梯度的一阶矩估计（梯度的均值）和二阶矩估计（梯度的未中心化的方差）进行综合考虑，最终计算出更新步长。它的计算效率高，收敛速度快，更适合深度学习模型的优化。

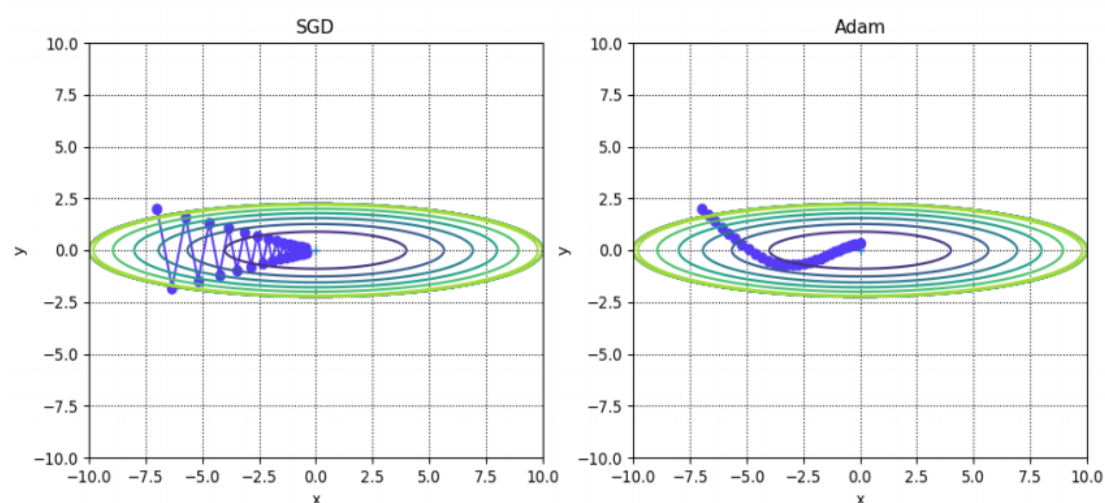


图 18. 两种优化器的最优化梯度更新路径

### 4.5.2 Adam 优化器的原理

Adam 优化器以梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率，引入动量和自适应学习率使算法加快模型学习、提高网络识别精度并在收敛时减

小震荡。以梯度下降第  $t$  次迭代为例，Adam 优化器公式及原理如下：

(1) 更新时间步长。

$$t = t + 1 \tag{4.12}$$

(2) 计算梯度。

$$g_t = \nabla_{\theta} f_t(\theta_{t-1}) \tag{4.13}$$

式中， $g_t$  为历史梯度值， $f(\theta)$  是参数  $\theta$  的随机目标函数值（损失函数值）， $\theta_{t-1}$  为  $t - 1$  时刻的参数值。

(3) 计算历史梯度的一阶指数平滑值，用于得到具有动量的梯度值。

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{4.13}$$

式中， $m_t$  为一阶动量项， $\beta_1$  为超参数，默认值为 0.9。

(4) 计算历史梯度的平方的一阶指数平滑值，用于得到每个权重参数的学习率权重参数。

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot (g_t)^2 \tag{4.14}$$

式中  $v_t$  为二阶动量项， $\beta_2$  为超参数，默认值为 0.999。

(5) 计算参数更新值，参数更新值正比于历史梯度的一阶指数平滑值，反比于历史梯度的平方的一阶指数平滑值。

$$\theta_t = \theta_{t-1} - \frac{\alpha \cdot m_t}{\sqrt{v_t} + \epsilon} \tag{4.15}$$

式中， $\alpha$  为步长或学习率，初始值为 0.001， $\epsilon$  的作用是避免除数变为 0，本次取值为  $10^{-8}$ 。

4.5.3 Adam 优化器的选择

利用 Matlab2016 分别计算出运用 SGD 和 Adam 优化器处理后各个变量参数的值，如下表所示：

表 6. SGD 和 Adam 优化器处理后变量参数对比		
优化器	SGD	Adam
耗时/s	18.252	13.014
loss	0.3011	0.0878
acc	0.9247	0.9666
val_acc	0.9218	0.9553



我们组分别以 MAE 和 MSE 作为损失函数，以 Epoch 值作为参数，分别运用 SGD 和 Adam 优化器进行处理，测试的结果如下图所示：

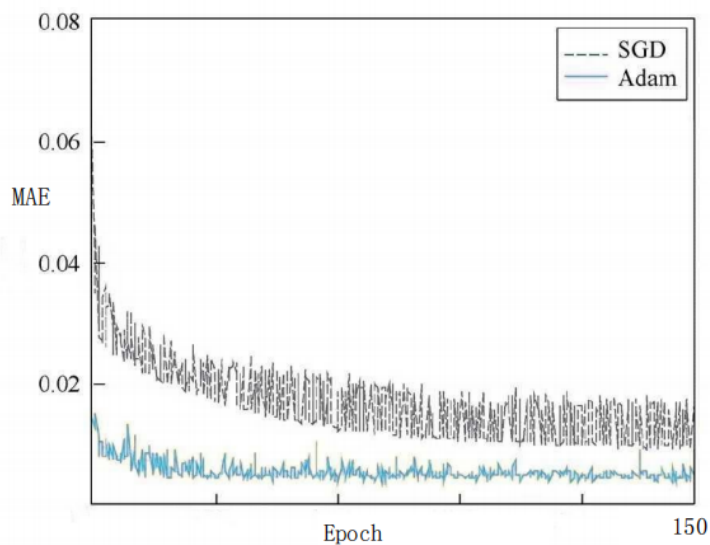


图 19. 以 MAE 为损失函数，运用 SGD 和 Adam 优化器测试结果

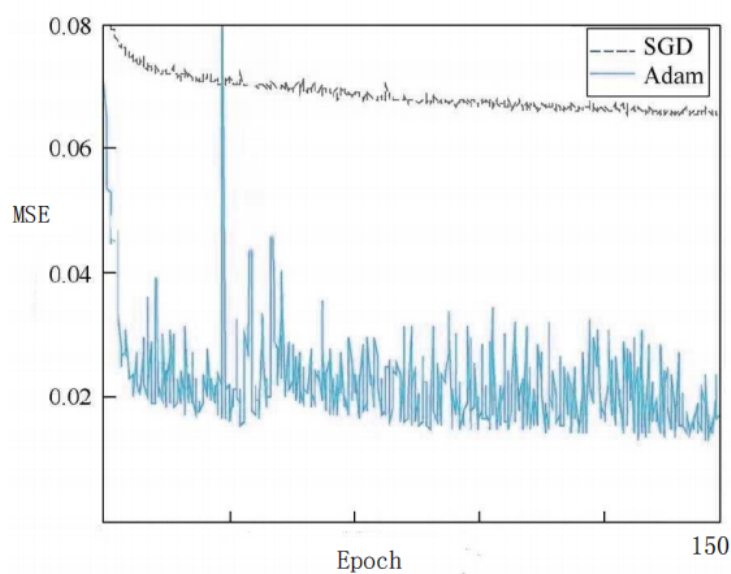


图 20. 以 MSE 为损失函数，运用 SGD 和 Adam 优化器测试结果

通过对比上述 SGD 和 Adam 优化器处理后的测试结果，相比于 SGD 优化器，Adam 优化器具有以下优势：

- (1) 测试结果准确率高、耗时短、计算高效，方便实现。
- (2) 更新步长和梯度大小无关，只和 $\alpha$ 、 $\beta_1$ 、 $\beta_2$ 有关系。并且由它们决定步长的理论上限。

(3) 对目标函数没有平稳要求，即损失函数可以随着时间变化。并且随着 Epoch 值的增加，Adam 优化器处理后的 MAE 和 MSE 损失值均低于 SGD 处理后的值，说明经 Adam 优化器处理后的模型拟合的很好。

(4) 能较好处理稀疏梯度（梯度在很多步长处都是 0 的情况）。

#### 4.6 MSE 损失函数的选择

我们知道，损失函数（Loss Function）是用来估量模型的预测值与目标值的偏离程度。损失函数是模型对数据拟合程度的反映，拟合越差，损失函数的值就应该越大，同时我们还期望，损失函数在比较大时，它对应的梯度也要比较大，这样的话更新变量就可以更新的更快一点。损失函数中有两种常见的回归损失函数，一种是均方误差（Mean Square Error, MSE），另一种是平均绝对误差（Mean Absolute Error, MAE）。

平均绝对误差（Mean Absolute Error, MAE）是目标值与预测值之差绝对值和的均值，表示了预测值的平均误差幅度，不需要考虑误差的方向，其公式如下所示：

$$MAE = \frac{\sum_{i=1}^n |y_i - f(x)|}{n} \quad (4.16)$$

均方误差（Mean Square Error, MSE）是预测值 $f(x)$ 与目标值 $y$ 之间差值平方和的均值，其公式如下所示：

$$MSE = \frac{\sum_{i=1}^n (y_i - f(x))^2}{n} \quad (4.17)$$

例如图 13 和 14 所示，MAE 作为损失函数比较稳定，对于离群点不太敏感。因为 MAE 计算的是误差 $(y - f(x))$ 的绝对值，对于任意的输入值，都有着稳定的梯度，不会导致梯度爆炸问题。但是 MAE 在 $(y - f(x) = 0)$ 点处不可导，因此导数不连续，求解效率低。另外，对于小的损失值，其梯度也是大的，在深度学习中不利于收敛和模型的学习。

所以在这种情况下，我们组最终选择均方误差损失函数（MSE），因为 MSE 的函数曲线光滑、连续并且处处可导，便于我们使用 Adam 算法，而且梯度值是动态变化的，随着误差的减小，梯度也在减小，因此有利于快速收敛，更适合模型的学习。

## 五、测试报告

### 5.1 集合模型的筛选和性能分析

我们首先将大赛提供的数据集（不包括需要预测的 96 张图像）中 850 张已知病变性质的甲状腺超声图像进行**五折分割**，**0.8** 的比例（680 张）作为**训练数据**，**0.2** 的比例（170 张）作为**测试数据**。

在本次试验中，我们使用类似于**投票机制**的**集成模型**对预测结果进行综合研判，具体实现方法为：用**奇数**个不同的神经网络模型进行图像分类，每个模型都有独立的参数和结构。当我们给定相同的图像输入时，所有的模型都会给出一个分类结果，我们选择最多数的结果作为最终的分类。例如，我们使用 ResNet-18、ResNet-50 和 EfficientNet-B0 三个模型，每个模型都对同一张图像做出阳性或阴性的判断，如果有两个模型判断结果为阳性，一个模型判断结果为阴性，那么我们就认为这张图像是阳性的。

我们在 ResNet 与 EfficientNet 中选择了一些可能较为适合本次图像分类问题的结构，通过对训练结果进行分析，我们列出前七名具有较优 F1-Score 和准确性的模型，计划在其中做一些筛选，得到最优化集成解决方案。具体处理方案和性能如下：

表 7. 较优模型训练方法及性能

模型名称	Backbone 结构	层数方案	数据增强	Epoch	F1-score	Accuracy
model1	ResNet-18	冻结前 4 层	处理 1	37	0.8659	87.06%
model2	ResNet-18	全开放	处理 1	32	0.8902	88.82%
model3	ResNet-50	冻结前 4 层	处理 1	39	0.8795	88.24%
model4	EfficientNet-b0	全开放	处理 1+处理 2	13	0.8994	90.00%
model5	EfficientNet-b0	全开放	处理 1	21	0.8876	88.82%
model6	EfficientNet-b0	全开放	处理 1	11	0.8851	88.82%
model7	EfficientNet-b5	全开放	处理 1	17	0.8834	88.82%

其中，**处理 1\***指使用 *alumentations.Resize* 函数将图像的宽度和高度都缩放到 224 像素。然后使用 *alumentations.CLAHE* 函数对图像进行对比度受限的

自适应直方图均衡化以增强图像的对比度和细节，此函数的参数 *clip\_limit* 表示对比度的最大值，被设置为 4.0；*tile\_grid\_size* 表示分割图像的网格大小，被设置为 (4, 4)，*p* 表示应用这个函数的概率，被设置为 0.9。接着使用 *albumentations.Normalize* 函数对图像进行归一化，以将图像的像素值转换为标准正态分布，便于后续的模式训练，此函数的参数 *mean* 表示归一化的均值，被设置为 (0.485, 0.456, 0.406)；参数 *std* 表示归一化的标准差，被设置为 (0.229, 0.224, 0.225)；*max\_pixel\_value* 表示图像的最大像素值，被设置为 225.0；*p* 表示应用这个函数的概率，被设置为 1.0。最后使用 *ToTensorV2* 函数将图像转换为 PyTorch 张量。

**处理 2** 指图像被转换为 PyTorch 张量前对图像进行水平翻转，*p* 表示应用这个函数的概率，被设置为 0.5。

以上模型训练时，**学习率**均被设定为 0.0001，**学习率衰减**均被定义为 25，**Batchsize** 均被设定为 5。

从表 7 中，我们首先排除了准确性明显较低的 model1 和 model3，保留了 F1-Score 和准确性明显优越的 model4，并根据前文模型设计的研究决定保留基于 Resnet-18 的 model2。根据上述考虑和奇数个模型以实现投票机制的需求，最终我们计划在如下方案中选择其一：

表 8. 待评判的两个集成学习解决方案

方案一	保留 5 种模型结构，分别为 model2、model4、model5、model6、model7
方案二	保留 3 种模型结构，分别为 model2、model4、model5

我们针对方案一和方案二分别进行了训练，并结合双线性插值的方法分别绘制了两个方案的准确率概率密度散点图，辅助我们做出最终的选择。

双线性插值是一种基于图像领域的线性插值方法。它考虑了目标像素周围的 4 个邻近像素，并根据距离和权重进行加权平均，可以更平滑地预测和处理图像的变化。此外，双线性插值方法能够识别和保留边缘信息，相对保持边缘的锐利度，减少插值过程中可能引入的模糊或失真，对于医学图像中的结构定位和形态分析至关重要。

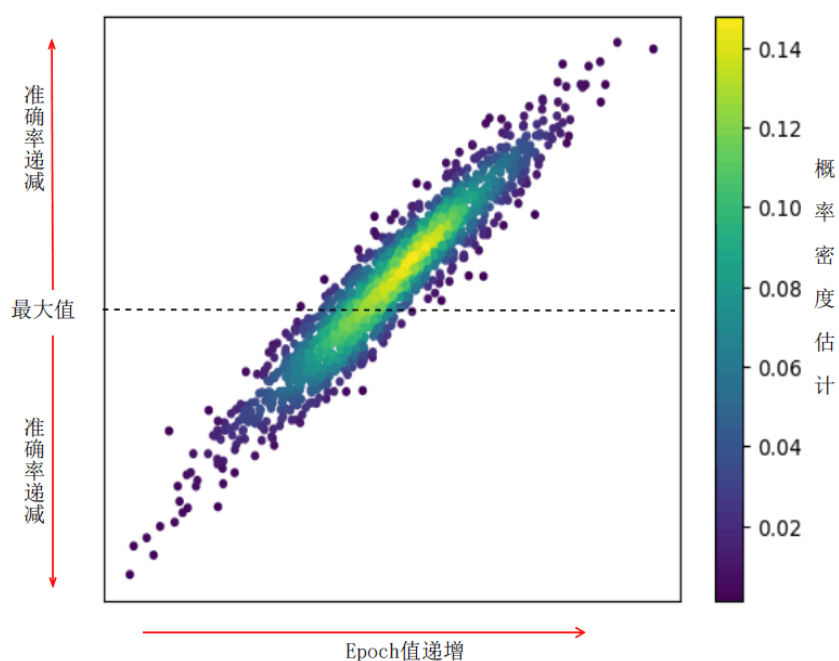


图 21. 方案一的准确率概率密度散点分布图

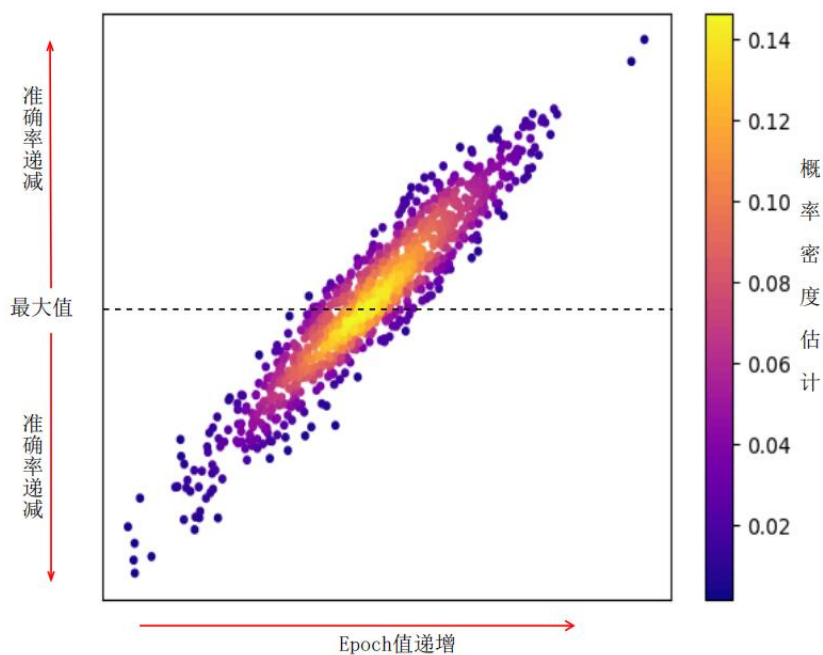


图 22. 方案二的准确率概率密度散点分布图

从图中我们可以分析出，方案一中概率密度最高的区域集中在准确率最大值上方的训练记录中，即有大量次数的训练结果准确率低于最理想的范围，且密度散点形成的彩色条带两端的散点分布较多，整体数据点的集中性较差；方案二中

概率密度分布中心区域即为最高准确率所在区域，数据点集中性强，彩色条带两端的散在数据点明显较方案一减少。因此，我们决定选择方案二的模型进行实际应用，即表 7 中 **model2、model4、model5** 组成的**基于投票机制的集成学习模型**。

5.2 对赛方测试图像分类结果

表 9. 赛方测试图像 ID-类别预测表

良性	恶性
1,3,4,9,10,13,14,17,19,20,21,23,26,27,30, 31,32,33,38,41,42,44,47,50,51,57,58,59,6 2,63,66,68,72,75,79,80,81,82	2,5,6,7,8,11,12,15,16,18,22,24,25,28,29,34,35,36,37, 39,40,43,45,46,48,49,52,53,54,55,56,60,61,64,65,67, 69,70,71,73,74,76,77,78,83,84,85,86

六、代码简介

代码文件说明：

- （一）研究背景——中国基层医疗资源建设与利用水平评价体系的建立
- InformationEntropy\_weight.m**: 利用熵权法对建立的模型中各个指标的信息熵和权重进行计算。
- LinearWeightedScoring.m**: 用线性加权法实现了对我国 31 个省份地区在我们建立的评价体系中的评分计算。
- fuzzyClustering.m**: 实现了基于模糊关系的模糊聚类分析，利用平移-极差变换进行数据预处理，建立基于指数相似系数的模糊矩阵，并采用 F 统计量法进行聚类分析，对全国各省份在基层医疗建设与利用水平上的评分进行分类。
- primaryCareDistributionHeatMap.m**: 本文图 1. 中国 2021 年各地区基层医疗卫生机构数量分布热力图的绘制代码。
- rainCloud.r**: 本文图 2. 近 20 年中国和世界甲状腺癌发病率对比云雨图以及图 3. 甲乙两类地区甲状腺癌体检检出人次占总发病人次的比率对比云雨图的绘制代码。
- （二）CLAHE 增强算法的选择
- clahe.m**: 实现了 CLAHE 算法，能够对图像进行增强化处理。

**greyImage.m:** 在算法实现的基础上，运用算法对图像增强并输出，并绘制出使用算法前后图片的灰度分布直方图。

### （三）优化器的选择

**SGD.m 和 Adam.m:** 通过 5 个步骤实现 SGD 优化器、Adam 优化器及它们的优化过程：

1. 定义真实参数
2. 随机生成数据
3. 实现 SGD 优化器
4. 模型参数更新
5. 返回模型参数

**AdamSGDCompare.m:** 实现了 SGD 优化器和 Adam 优化器的性能对比，主要针对两个优化器在应用时的耗时、误差率、准确率及稳定准确率四个指标。

**gradientUpdatePathFig.py:** 实现了 SGD、Momentum、AdaGrad、Adam 四种优化器的最优化梯度更新路径图的绘制，用于模型建立阶段优化器选择时辅助决策的重要依据之一，考虑到本题图像处理的特征需求，为了使优化器最终选择决策的理由更加直观鲜明，我们在报告中只保留了 SGD、Adam 两种优化器的可视化路径图像对比。

### （四）集成学习模型的设计

**seaFig.m:** 以 Epoch 取值的若干小区段为自变量，对比 EfficientNet-B5 和 ResNet-18 两种算法模型在训练过程中随 Epoch 值增加的不同表现（以每 5 个单位 Epoch 值为取值分隔，取误差率数值作为衡量指标），输出两个模型各自的热力图。

### （五）模型应用

**functions.py:** 实现五折交叉验证法分类、自动检测最佳训练设备、清空文件夹等函数。

**prediction.py:** 实现 Predict 类用于加载模型, 实现 predict 方法用于对图片进行预处理并得出分析结果; 实现 Voting 类获得投票模型结果; 并将结果保存。

**split\_dataset.py:** 调用 functions.py 中的函数, 实现对数据集进行五折交叉验证的划分。

**Train\_EfficientNet\_B0(F1).py:** 实现给定超参数和训练集对 EfficientNet-B0 模型进行训练, 并根据 F1 记录最优值。

**Train\_ResNet18(acc).py:** 实现给定超参数和训练集对 ResNet18 模型进行训练, 并根据 acc 记录最优值。

**Train\_ResNet18(F1).py:** 实现给定超参数和训练集对 ResNet18 模型进行训练, 并根据 F1 记录最优值。

**evaluate\_model.py:** 实现使用指定模型对测试集进行评估, 计算模型 Precision、Recall、F1-Score、Accuracy。

**app.py:** 使用 Flask 框架搭建 Web 服务器, 实现超声图像在线评估和诊断。为 TMICM 在线影像识别系统后端。

#### (六) 模型测试报告

**bilinearInterpolation.m:** 缩放图像后实现双线性插值方法对本文所及超声图像进行处理, 通过嵌套的循环遍历缩放后的图像的每个像素位置, 计算对应原图像中的四个像素点的坐标, 并获取其颜色值。计算插值结果, 并将其赋值给输出图像的相应位置。

**ProbabilityDensityScatter.m:** 本文图 21-22. 方案一和方案二的准确率概率密度散点分布图的绘制代码。

## 七、方案实施

为了降低医疗资源落后地区诊疗的成本, 提高居民甲状腺肿瘤的体检检出率, 减免医生部署模型的技术壁垒, 减少资源浪费, 我们开发了 TMICM (Thyroid Medical Image Classification Model, 甲状腺医学图像分类模型) 在线影像识别系



统。依托 TMICM 在线图像分类系统，医生可以轻易地上传用户的超声影像并实时获得模型分析结果。医生也可以通过邮件系统与更多医生进行交流来获得更全面的結果，同时对模型结果进行优化和改进。我们计划建立包含医学生、基层医生、综合医院医生的医学影像分析联盟，从在线图像分类系统的邮件功能出发，设法打通基层医生与更有经验的医生的沟通渠道，使得甲状腺癌可以及早发现和治疗，提升患者的预后生存质量和幸福感。



图 23. TMICM 在线影像识别系统

## 八、成员分工

- 组长：** 模型训练；训练数据和结果数据分析和处理；TMICM 在线影像识别系统服务器架构设计和搭建；撰写竞赛报告。
- 组员 1：** 模型训练；构建研究背景中的评价体系；建立数据分析模型辅助集成学习模型的设计和选择；TMICM 在线影像识别系统的设计和搭建；设计报告中全部可视化数据图像的代码编写。
- 组员 2：** 撰写竞赛报告；文献数据收集和整理；提出设计优化策略。

## 九、参考文献

- [1]国家卫生健康委员会. 2022 中国卫生健康统计年鉴[M]. 北京:中国协和医科大学出版社, 2022.
- [2]Wang, J., Yu, F., Shang, Y., Ping, Z., & Liu, L. (2020). Thyroid cancer: incidence and mortality trends in China, 2005-2015. *Endocrine*, 68(1), 163–173.  
<https://doi.org/10.1007/s12020-020-02207-6>
- [3]Mingxing Tan, Quoc V. Le. (2019,May 28). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. <https://arxiv.org/abs/1905.11946>
- [4]Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. (2015,December 10). Deep Residual Learning for Image Recognition. <https://arxiv.org/abs/1512.03385>
- [5]Diederik P. Kingma, Jimmy Ba. (2014, December 22). Adam: A Method for Stochastic Optimization. <https://arxiv.org/abs/1412.6980>