**长　沙　学　院**

智慧医疗

学 院： 计算机科学与工程学院

班 级： 21智能01

学 号： B20210307110

姓 名： 刘奇

2025年 1月 4日

# 选题与数据集选择

在医疗领域，胸部X光片一直是诊断肺部疾病的重要手段。从早期的医学研究到如今的临床实践，胸部X光片在疾病的诊断、治疗以及预防等方面都发挥着不可替代的作用。传统的X光片诊断主要依赖于医生的经验和专业知识，医生们凭借自身所掌握的医学知识和丰富的临床经验，对X光片进行分析和判断。然而，人工诊断存在着诸多不可避免的局限性。一方面，主观因素对诊断结果的影响较大。不同医生对同一X光片的解读可能存在差异，这种差异往往与医生个人的认知水平、工作经验以及情绪状态等有关。另一方面，诊断效率相对较低，在面对大量的X光片时，医生需要花费较多的时间和精力去逐一分析，这不仅耗费了大量的人力和物力资源，还可能导致患者等待时间过长，延误病情。但随着科技的飞速发展，人工智能技术逐渐渗透到医疗领域。其中，深度学习算法的出现为X光片的分类带来了新的契机。深度学习算法通过对大量的X光片数据进行学习和训练，能够自动提取出其中的特征信息，并根据这些特征信息进行分类。这种方法不仅能够大大提高诊断的准确性，还能有效提升诊断效率。

在此背景下，本次大作业的选题为 “基于PyTorch的医学图像处理：X光片分类任务”。

这一选题具有十分重要的意义，因为准确的X光片分类能够帮助医生迅速、准确地判断患者是否患有肺炎，这对于及时采取相应的治疗措施、降低患者的痛苦以及提高治疗效果都有着积极的作用。同时，通过对X光片分类的研究和应用，还可以进一步优化医疗资源的配置，提高医疗服务的质量和水平。

本次任务主要利用PyTorch框架构建一个高效的X光片分类模型，使用的是Chest X - Ray Images (Pneumonia) 数据集。Chest X - Ray Images (Pneumonia)数据集包含了大量的胸部X光片图像。这些图像具有丰富的特征信息，为模型的训练提供了充足的素材。数据集分为训练集、测试集和验证集，该数据集数据量较大，能为模型提供了足够的样本进行训练，有助于提高模型的泛化能力。而且标注信息明确，将每张X光片都进行了分类，便于模型进行学习。同时图像质量较高，能够清晰地展现肺部的结构和特征，为模型训练提供了良好的基础。

通过对该数据集的分析和处理，能够利用PyTorch框架构建一个性能良好的X光片分类模型。在后续的工作中，可以进一步优化模型，提高分类准确率。同时，还将探索更多的应用场景和方法，为医疗诊断提供更加可靠的支持。相信在不久的将来，基于PyTorch的X光片分类技术将在医疗领域发挥更大的作用，为人类健康事业做出更大的贡献。

# 数据预处理

在利用Chest X - Ray Images (Pneumonia)数据集构建X光片分类模型时，数据预处理是一个至关重要的环节。合理的数据预处理能够提高模型的性能、加快训练速度，并增强模型的泛化能力。

在本实验中，采用了一系列的预处理方法。

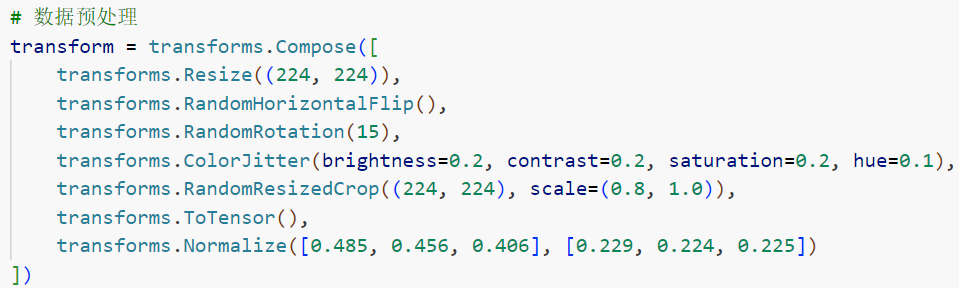


图 1 数据预处理

使用transforms.Resize((224, 224))将图像大小调整为统一的(224, 224)尺寸。这主要有两个重要原因：第一，模型输入通常要求图像具有固定的尺寸，不同的X光片可能在原始拍摄时有不同的大小，如果不进行统一处理，模型将无法正确接收和处理这些输入；第二，将图像大小调整为合适的尺寸可以减少计算量。在深度学习模型中，过大的图像会导致过多的参数和复杂的计算，这不仅会增加训练时间，还可能导致过拟合。而(224, 224)这个尺寸在平衡图像细节和计算效率方面是一个比较合适的选择，它既能够保留足够的肺部特征信息，又不会使模型的计算负担过重。

transforms.RandomHorizontalFlip()方法用于对图像进行随机水平翻转。在医学影像中，肺部的左右结构是相对对称的。通过随机水平翻转图像，可以增加数据集的多样性。例如，一张正常的X光片经过水平翻转后，其肺部的基本结构和特征依然是正常的，但对于模型来说，这相当于增加了一张新的训练样本。这种数据增强的方式有助于模型学习到更具鲁棒性的特征，避免模型过度依赖图像的方向等特定因素，从而提高模型对不同形态X光片的分类能力。

transforms.RandomRotation()允许图像在一定角度范围内（这里为15度）进行随机旋转。与随机水平翻转类似，随机旋转也是一种数据增强的手段。在实际的X光片拍摄过程中，由于患者的体位、拍摄设备的角度等因素，X光片可能会有轻微的旋转。通过对数据集中的图像进行随机旋转，可以使模型更好地适应这些变化。而且，这种旋转操作在一定程度上模拟了实际场景中的各种情况，让模型能够学习到不受旋转角度影响的肺部特征，使其在面对各种角度拍摄的X光片时都能准确地进行分类。

transforms.ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.1)是随机调整图像的亮度、对比度、饱和度和色调的操作。这种变换可以模拟不同光照条件下的X光片，增加模型对不同条件下图像的适应能力，提高模型的泛化性。

transforms.RandomResizedCrop((224, 224), scale=(0.8, 1.0))是随机裁剪并调整图像大小的操作。这个变换首先从原始图像中随机裁剪出一个区域，裁剪区域的大小是原始图像的80%到100%（按照面积计算），然后调整这个裁剪区域的大小到224x224像素。这也是一种数据增强技术，可以模拟X光片在实际应用中可能出现的局部遮挡或裁剪情况，使模型更加健壮。

transforms.ToTensor()是将图像数据转换为张量的操作。在PyTorch框架中，张量是基本的数据结构，模型的输入必须是张量形式。张量可以提高深度学习模型的训练速度，而且具有高效的存储和操作方式，能够更好地与PyTorch的各种计算模块和优化算法进行集成，便于模型进行各种数学运算和参数更新。

transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])是对图像数据进行归一化的操作。归一化的主要目的是将数据的分布调整到一个合适的范围。在图像数据中，不同的X光片可能由于拍摄设备、光照等因素导致像素值分布差异较大，通过归一化，可以将数据的均值和标准差调整到特定的值，使得数据分布更加稳定和一致，这有助于模型更快地收敛，同时，还可以减少数据中的异常值对模型训练的影响，提高模型的鲁棒性和准确性。

鉴于该数据集的规模较大，直接使用全部数据进行模型训练可能会导致训练时间显著延长，并对计算资源造成较大压力。为了平衡训练效率和模型性能，我从各个数据集中随机筛选出一部分的数据用于模型的训练和验证，在保证数据多样性和覆盖面的同时，减少训练数据的总量。

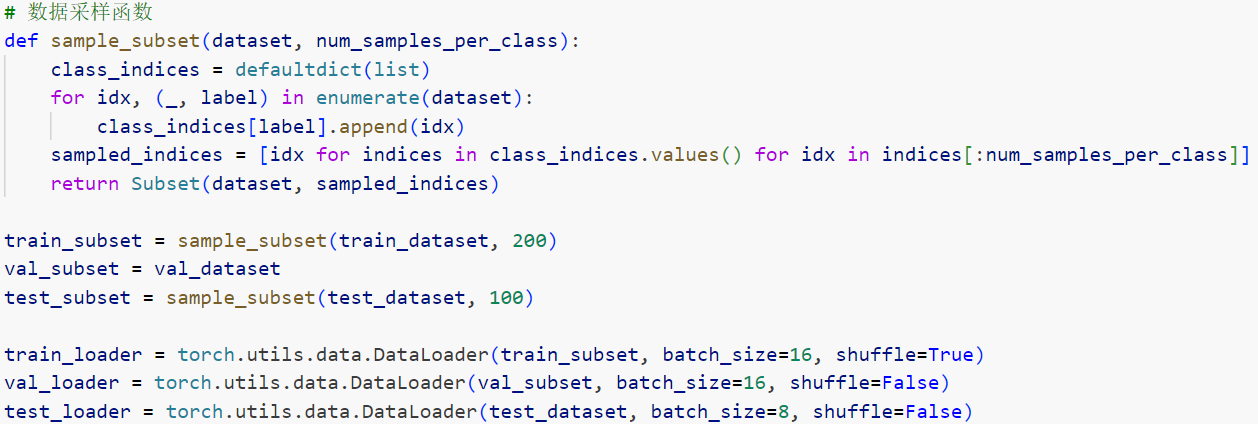


图 2 数据集的选择

这样不仅能够缩短模型训练的时间，降低对计算资源的需求，还能够确保模型训练的效率和质量。此外，使用部分数据还有助于减少过拟合的风险，而且模型能够在更广泛的数据场景上进行泛化能力的测试。

# 模型构建

在本次X光片分类任务中，我选用基于预训练的ResNet18模型。ResNet作为一种深度神经网络架构，在图像识别领域表现出色，它通过引入残差块有效解决了深度神经网络训练过程中的梯度消失和梯度爆炸问题，使得网络能够构建得更深，学习到更复杂的特征。ResNet18在模型复杂度和计算资源需求之间达到了良好的平衡，既能提取足够的图像特征，又不会因过于复杂而导致训练时间过长或过拟合。这个预训练模型在大规模图像数据集上进行了训练，已经学习到了丰富的通用图像特征，这有助于在X光片分类任务中快速收敛并获得更好的性能。

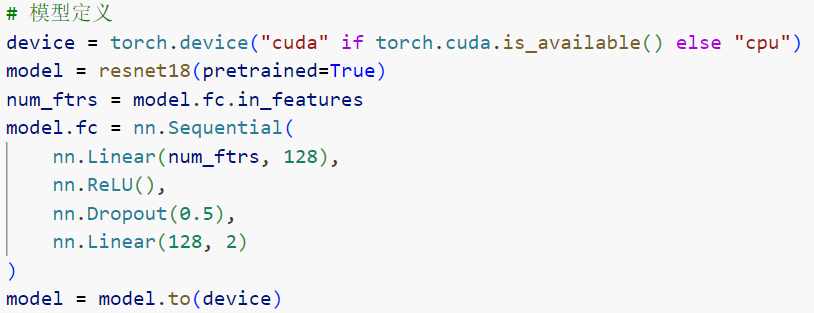


图 3 模型的定义

在获取原始ResNet18模型最后全连接层的输入特征数量为num\_ftrs，这一数值记录了展平后的特征向量维度，紧接着构建新的全连接层序列，第一层全连接层将输入特征向量从num\_ftrs维映射到128维，这样可以降低模型复杂度，减少计算量，同时有助于提取更具代表性的特征。接着引入ReLU激活函数，ReLU函数能够引入非线性因素，打破线性变换的局限性，增强模型的表达能力。随后是Dropout层，它以一定概率（这里为0.5）随机将输入神经元输出设置为0，防止模型过度依赖某些特定神经元，从而增强模型的泛化能力避免过拟合。由于本次任务是二分类，所以第二层全连接层将128维的特征向量进一步映射到2维。最后通过Softmax层将原始分数转换为概率分布，使得每个类别对应的输出值在0到1之间并且所有类别输出值之和为1，从而实现对X光片的分类预测。

# 模型训练与调优

在建立好模型后还要对损失函数、优化器以及学习率调度器进行设置，以便能够很好的训练模型。

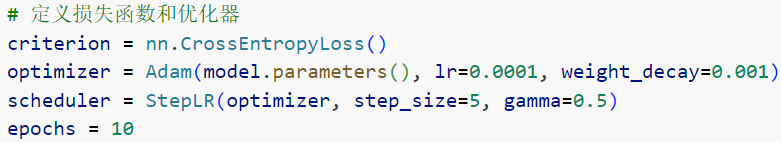


图 4 模型参数的设置

在本次任务中选择交叉熵损失函数作为模型的损失函数。交叉熵损失函数是多分类问题中最常用的损失函数之一，它能够衡量模型预测的概率分布与真实标签的概率分布之间的差异。在二分类任务中，交叉熵损失能够引导模型学习到更加精确的类别边界，从而提高分类的准确性。

模型选用了Adam优化器来更新模型的权重。Adam优化器结合了Momentum和RMSprop的优点，能够自适应地调整学习率，并且对于稀疏梯度具有较好的鲁棒性。在优化器中，设置了初始学习率lr=0.0001，这是一个较小的学习率，有助于模型在训练初期稳定地收敛。同时，引入了权重衰减weight\_decay=0.001，这有助于防止模型过拟合，提高模型的泛化能力。

随后定义了学习率调度器StepLR来用于在训练过程中动态调整学习率。StepLR调度器会在每5个epoch后，将学习率乘以一个因子gamma=0.5，即学习率会按照0.5的比率衰减。这种策略有助于在训练的后期降低学习率，使得模型能够更加精细地调整权重，避免过大的步长导致的震荡。在本任务中，选择了10个epoch，这是一个适中的轮数，既能够保证模型有足够的时间学习数据特征，又能够避免过多的训练导致的过拟合。

经过上面的设置来训练模型，得到下面结果。

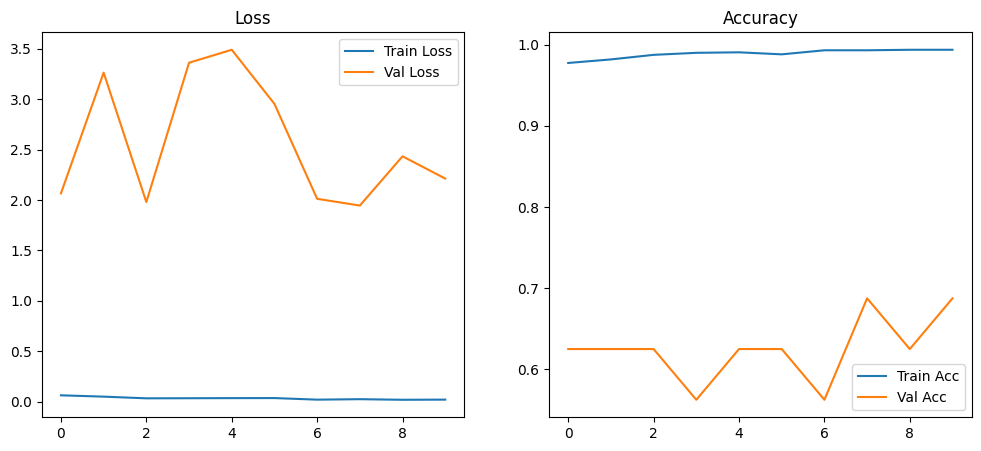


图 5 初始训练结果

从训练结果来看，在整个训练过程中，训练损失呈现出逐渐下降的趋势。在第1个训练周期中，训练损失为0.0621，随着训练的推进，训练损失下降至0.0194。这表明模型在不断地学习和调整参数，对训练数据的拟合程度越来越好。训练准确率也在逐步提升，也从开始的的0.9775到最后的0.9938。这显示出模型对训练数据的预测能力不断增强。

在验证阶段，验证损失和验证准确率呈现出一定的波动。在第1个训练周期，验证损失为2.0658，验证准确率为0.6250。后续的过程中验证损失和验证准确率在不同范围内波动。例如第4个训练周期验证损失为3.3623，验证准确率为0.5625；第8个训练周期验证损失为1.9451，验证准确率为0.6875。

综合来看，模型在训练过程中表现出了较好的学习能力和稳定性。然而，该模型有着明显的过拟合现象，并且在验证阶段的波动反映出模型在泛化能力方面还有一定的提升空间。

在对模型的改进过程中，我发现验证集的数据量太少，导致验证损失和准确率一直处于波动状态，正确率也比较低。所以我尝试着不用原来的验证集，将测试集的20%作为验证集，并且增加其他数据的数量。同时减少Dropout层的概率、减少学习率、增加训练周期等一系列优化操作，以便更好的训练模型。

通过改进，模型的训练结果如下。

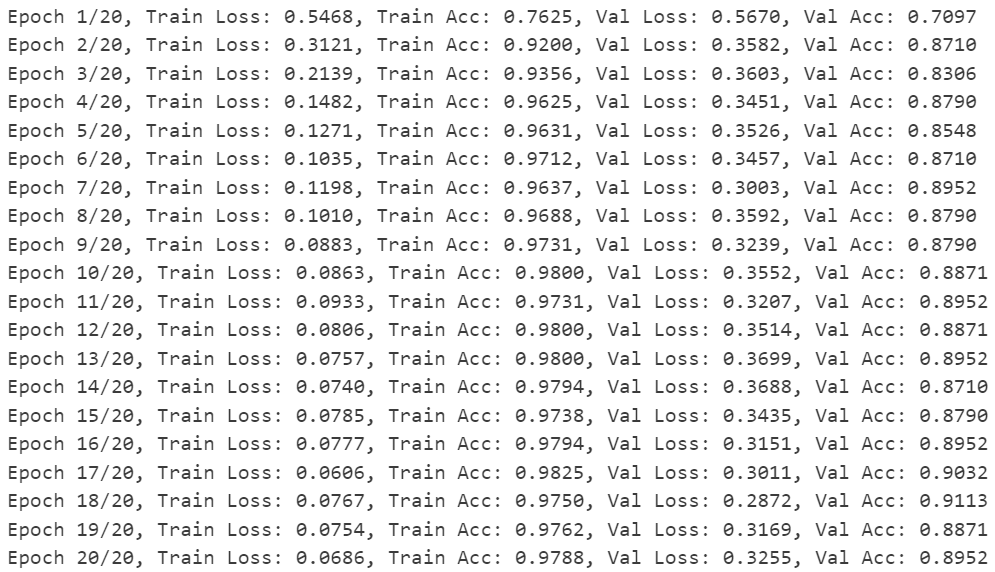


图 6 改进后的模型运行结果

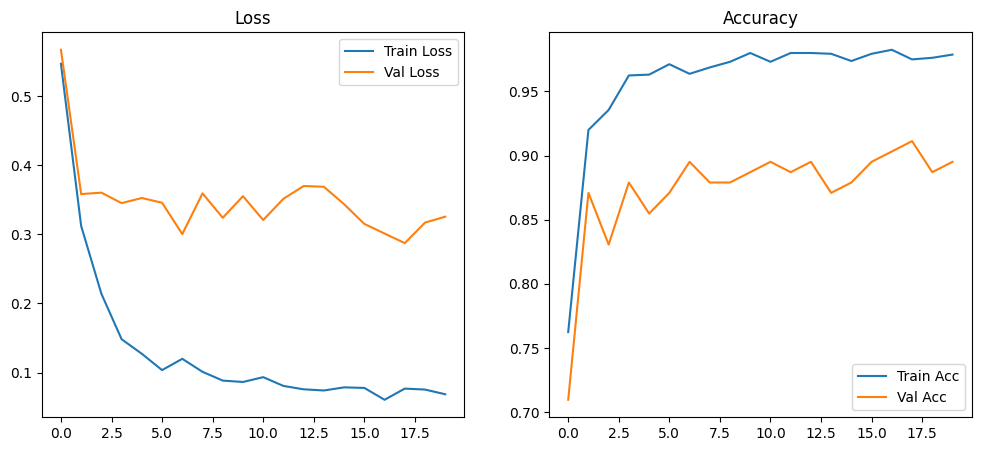


图 7 运行结果可视化

# 模型评估

对比分析模型改进前后的表现时，可以明显看出改进后的模型在多个方面都有了显著的提升。从训练损失的角度来看，改进后的模型从0.5468降低到了0.0686，这表明模型在训练集上学习得更加有效，能够更好地拟合训练数据。同时，训练准确率也从76.25%提高到了97.88%，虽然起始准确率较低，但最终达到了一个很高的水平，显示出模型在训练过程中的学习能力得到了加强。最显著的改进成果体现在验证集上的表现。改进前的模型在验证集上的准确率始终在62.5%到68.75%之间，远低于训练准确率，这表明模型可能严重过拟合。相比之下，改进后的模型在验证集上的准确率从70.97%提高到了91.13%，且大部分时间保持在87%以上，这不仅显示出模型的泛化能力得到了显著提升，也表明模型能够更好地识别未见过的数据。此外，改进后的模型在验证损失上也表现出了更好的稳定性，波动较小，最高为0.3699，最低为0.2872，这与改进前模型的验证损失波动形成鲜明对比，后者最高达到3.4910，最低为1.9800。这种稳定性的提升意味着模型在不同验证集上的表现更加可靠，减少了过拟合的风险。

表 1 改进前后模型损失与准确率的对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 训练准确率 | 验证准确率 | 训练损失值 | 验证损失值 |
| 改进前 | 0.9938 | 0.6875 | 0.0194 | 2.2130 |
| 改进后 | 0.9788 | 0.8952 | 0.0686 | 0.3255 |

尽管改进后的模型在验证准确率上有所提高，但与训练准确率相比，仍有提升空间。这意味模型在某些方面还有待进一步优化，例如通过调整超参数、增加正则化或尝试更复杂的模型架构来进一步提高模型的性能。此外，集成学习等技术也可以考虑，以进一步提高模型的稳定性和准确性。

另外在肺炎检测模型的分类报告中，模型在不同类别表现各异。正常（NORMAL）类别精确度达0.97，但召回率仅0.72，意味着模型识别正常样本的能力有待提升。肺炎（PNEUMONIA）类别表现更优，精确度0.86，召回率高达0.99，F1分数为0.92，在精确度与召回率间平衡良好。而且模型整体准确率为0.89，宏平均精确度、召回率、F1分数分别为0.91、0.86、0.87，加权平均则为0.90、0.89、0.88。

表 2 分类报告

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classification Report | | | | |
|  | precision | recall | f1-score | support |
| NORMAL | 0.97 | 0.72 | 0.83 | 189 |
| PNEUMONIA | 0.86 | 0.99 | 0.92 | 311 |
| accuracy |  |  | 0.89 | 500 |
| macro avg | 0.91 | 0.86 | 0.87 | 500 |
| weighted avg | 0.90 | 0.89 | 0.88 | 500 |

通过分类报告可以看出模型检测肺炎效果不错，但正常类别召回率低，可能导致部分正常病例误判为肺炎。为提升整体性能，需优化模型，减少误分类，尤其要提高正常类别的召回率，可通过数据平衡、改进特征提取或调整超参数等方式实现，以此增强模型实用性。

在混淆矩阵中提供了模型预测结果的详细视图，显示了模型在分类任务中的表现。在本次分析中，矩阵揭示了模型在正常（NORMAL）和肺炎（PNEUMONIA）两个类别上的预测准确性。具体来说，模型正确地将137个正常样本分类为正常，同时有52个正常样本被错误地识别为肺炎。这表明，尽管模型在大多数情况下能够准确识别正常样本，但仍有一定的误判率。在肺炎类别方面，模型的表现更为出色，正确分类了307个肺炎样本，仅有4个肺炎样本被误判为正常。这种高准确率的分类对于肺炎的早期诊断和治疗至关重要。

表 3 混淆矩阵

|  |  |
| --- | --- |
| Confusion Matrix | |
| 137 | 52 |
| 4 | 307 |

总体而言，混淆矩阵显示模型在肺炎检测上具有较高的准确性，但在正常类别的识别上存在一定的改进空间。为了提高模型的整体性能，减少正常类别的误判，可以考虑采用更多的数据增强技术、优化模型结构或调整超参数，从而提高模型的可靠性。

# 代码链接

将完整的项目代码上传至 GitHub ，并提供链接。链接中同时还要包括代码测试过程中的屏幕录屏。无录屏一律不及格。

# 总结

本次实验的核心任务是基于PyTorch框架构建一个高效的X光片分类模型，以实现对肺部疾病的快速诊断。实验过程中，我首先对Chest X-Ray Images (Pneumonia)数据集进行了深入的分析和处理，通过一系列的数据预处理步骤，包括图像大小调整、随机翻转、旋转等，有效地提高了数据的多样性和质量，增强了模型的泛化能力。这些预处理步骤模拟了实际应用中可能出现的各种情况，使模型能够更好地适应不同条件下的X光片图像。而且归一化操作进一步稳定了数据分布，减少了异常值对模型训练的影响，加快了模型的收敛速度。

在模型构建方面，我选择了预训练的ResNet18模型作为基础架构，充分利用了其在大规模图像数据集上学习到的丰富特征。通过调整全连接层和引入Dropout层，我们有效地防止了过拟合现象，提高了模型的泛化能力。在模型训练过程中，通过调整优化器、学习率调度器等参数，模型在训练集上的表现逐渐趋于稳定，训练损失不断下降，准确率稳步提升。

初步训练结果显示，模型存在一定的过拟合现象。但通过增加验证集的数据量、调整Dropout层概率、降低学习率等措施，模型的泛化能力得到了显著提升，验证集上的准确率和损失值均表现出更好的稳定性。模型评估结果显示，改进后的模型在训练集和验证集上的表现均有所提高，尤其是在验证集上的准确率达到了89.52%，显示出较强的泛化能力。分类报告和混淆矩阵进一步揭示了模型在不同类别上的表现，指出了正常类别召回率较低的问题。

尽管模型在整体上取得了较好的效果，但在正常类别的识别上仍存在一定的误判率。这可能由于数据集中正常样本的特征不够明显，或者模型在特征提取方面还有待进一步优化。实验过程中，对模型超参数的调整主要依赖于经验，缺乏系统的优化方法，可能未能充分挖掘模型的潜力。

在未来的研究中，我将继续深入探索智慧医疗领域的相关技术，努力提高模型的性能和实用性，并且尝试更多的数据增强技术，以增加数据的多样性和复杂性。同时，探索更先进的特征提取方法，如注意力机制等，以更好地捕捉X光片中的关键特征。此外，可以尝试更复杂的模型架构，如ResNet的更深版本或其他先进的卷积神经网络架构，以进一步提升模型的性能。还可以考虑集成学习技术，如随机森林、梯度提升树等，将多个模型的预测结果进行融合，以提高模型的稳定性和准确性。通过这些努力，我希望能够为医疗诊断提供更加可靠的支持，为人类健康事业做出更大的贡献。