****

****

**人工智能导论**

**课程报告**

**学 院 信息学院**

**专 业 软件工程**

**成 员 任宇、左泽宇、范周喆、赵志强**

**日 期 2024年7月2日**

目录

[一、 研究背景 3](#_Toc170829693)

[二、 技术文献综述 4](#_Toc170829694)

[三、 创新方法 6](#_Toc170829695)

[RLANet(ResNext-LSTM-Attention Network) 6](#_Toc170829696)

[MedNet 9](#_Toc170829697)

[四、 实验结论 12](#_Toc170829698)

[五、 讨论 13](#_Toc170829699)

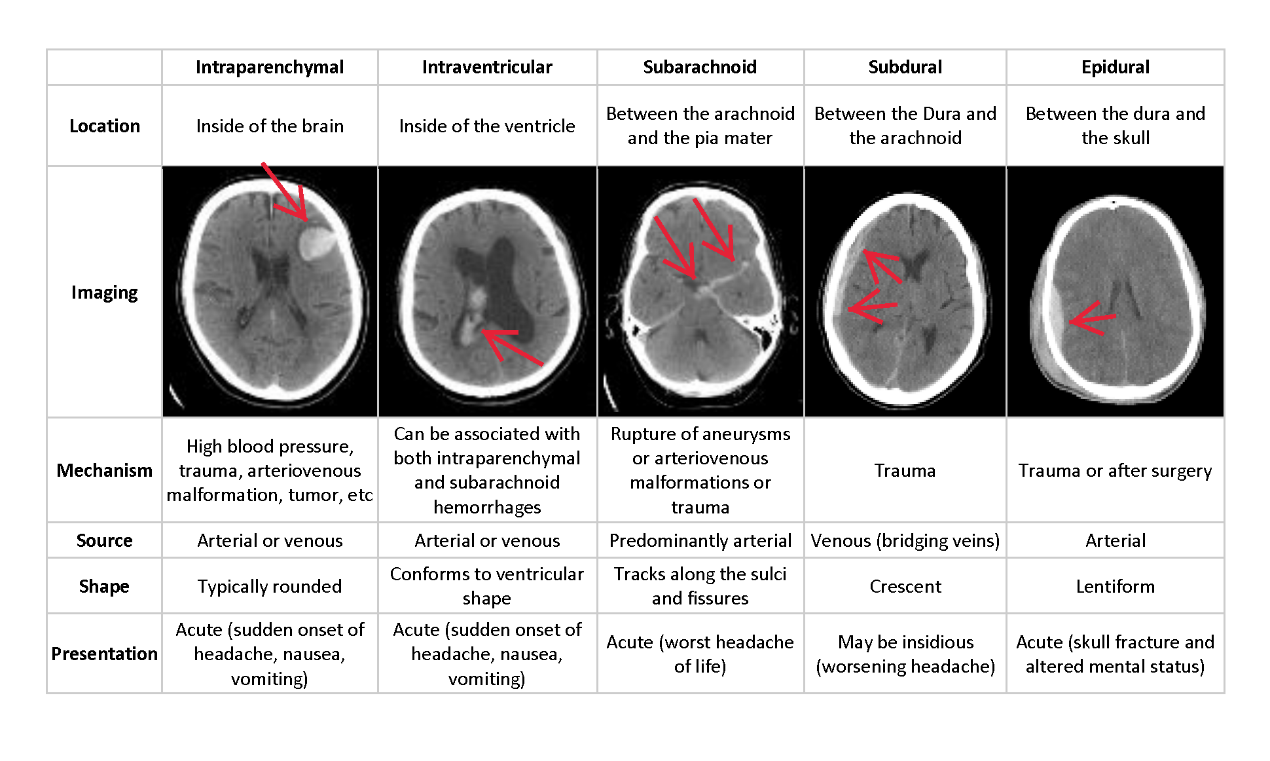
[六、 未来展望 15](#_Toc170829700)

[七、 参考文献 15](#_Toc170829701)

# 研究背景

颅内出血是一种常见的严重神经系统疾病，涉及到脑组织内或其周围血管的破裂，导致血液积聚在脑内或脑膜下空间。由于其快速发展的特性，颅内出血常常迅速恶化，对患者的生命健康构成了严重威胁。全球范围内，颅内出血的高致残率和高死亡率使其成为公共卫生领域亟需解决的问题之一。

颅内出血的类型主要包括脑室内出血、脑实质性出血、蛛网膜下腔出血、硬膜外出血和硬膜下血肿。这些类型的区分对于治疗方案的制定和预后的评估至关重要。尽早准确诊断出出血类型和准确位置可以大大提高治疗的效果，降低致残率和死亡率。



**图1 颅内出血类型图**

传统的诊断方法主要依靠医学影像技术，如计算机断层扫描（CT）和磁共振成像（MRI）。这些技术可以提供关于出血位置、大小和可能的原因的重要信息。在急诊医学中，CT因其速度快和普遍可用而成为首选。然而，CT和MRI图像的解析需要高度专业化的知识和经验，而且人工诊断存在主观性，可能导致误诊或漏诊。

随着人工智能尤其是深度学习技术的发展，越来越多的研究开始探索其在医疗影像分析中的应用。深度学习，特别是卷积神经网络（CNN），在图像识别和分类任务中已显示出极高的效能。在颅内出血的诊断中，深度学习模型能够自动识别和定位出血，这对于快速准确地提供诊断信息具有重大意义。

应用深度学习进行颅内出血诊断的主要优势包括：

* 提高诊断速度：深度学习模型能够在几秒钟内分析成百上千的图像，大大减少了从影像获取到诊断结果所需的时间。
* 提升诊断准确性：通过训练大量数据，深度学习模型可以识别出微小的或非典型的出血，提高了诊断的准确性和一致性。
* 辅助决策支持：深度学习模型可以作为医生的辅助工具，提供第二意见，减少医生的工作负担，尤其在医疗资源紧张的情况下显得尤为重要。

尽管深度学习在颅内出血诊断中展现出巨大潜力，但在实际应用中仍面临一些挑战。首先，高质量的训练数据是深度学习模型成功的关键。医疗影像数据的获取和标注需要大量的时间和专业知识，而且数据的隐私问题也需要被妥善处理。此外，模型的解释性是另一大挑战，医生和患者通常希望了解模型的决策过程，这对于提高深度学习模型的透明度和可信度是必要的。

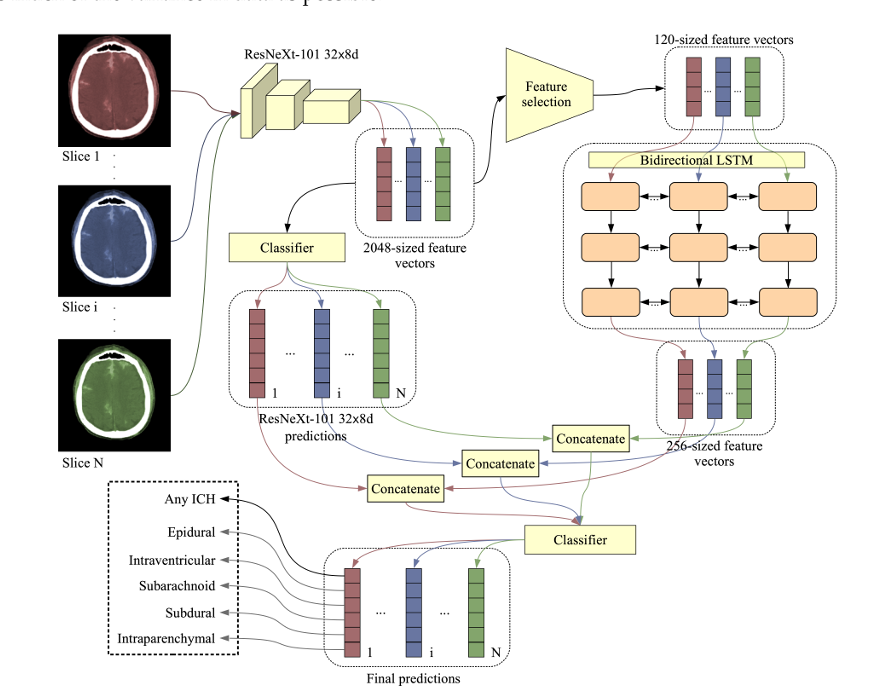
未来，随着技术的进步和更多临床验证的完成，深度学习有望在颅内出血的诊断中扮演更加重要的角色。通过与临床医生的紧密合作，开发更加精准、高效的深度学习工具，将进一步提升急性颅内出血的诊疗质量，最终提高患者的生存率和生活质量。此外，随着个性化医疗的发展，深度学习模型还可以根据患者的具体情况推荐个性化的治疗方案，这将是未来医疗技术发展的重要方向。

# 技术文献综述

在2019年的论文中[1]，Hesamian等人探讨了应用深度学习技术于医学图像分割的最新进展、成就与挑战。文章系统地回顾了当前医学图像分割中使用的几种主要网络结构，包括卷积神经网络（CNN）、全卷积网络（FCN）、U-Net及其变体等。这些网络结构被广泛应用于从医学图像（如CT或MRI图像）中识别出器官或病变区域的任务。通过深入分析这些网络的构建和训练方法，文章指出虽然深度学习在图像分割领域取得了显著的成果，但仍面临着一些挑战，例如数据量需求大、训练时间长、对特定特征的敏感性等。同时，作者也提出了可能的解决方案，例如使用转移学习、数据增强技术和改进的网络架构来提高模型的泛化能力和效率。

在2020年的研究中[2]，Mihail Burduja、Radu Tudor Ionescu 和 Nicolae Verga 提出了一种用于3D CT扫描中颅内出血检测和亚型分类的准确高效系统。该系统结合了卷积神经网络（CNN）和长短时记忆网络（LSTM），采用了一个轻量级的深度神经网络架构。CNN负责处理单独的CT切片，而LSTM则处理由CNN提供的多个特征嵌入。此外，为了提高处理效率，研究者们采用了不同的特征选择方法来优化LSTM的输入特征，并将CT切片的尺寸减少了一半，以加快模型训练速度。

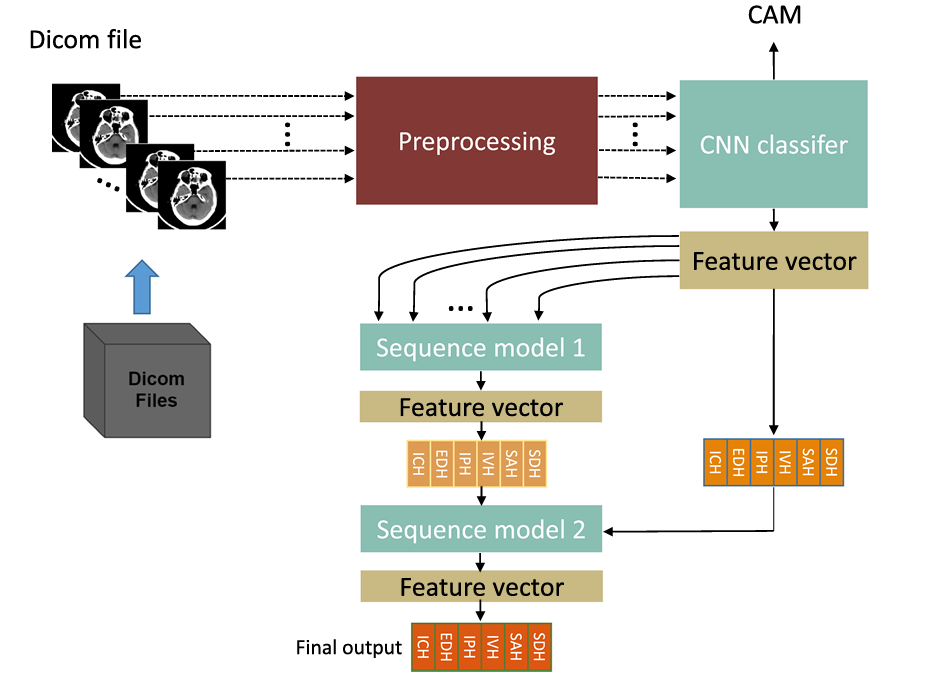
系统的设计旨在平衡速度和准确性，研究结果显示，该模型在最终测试集上的加权平均对数损失为0.04989，位于1345名参与者中的前30名（前2%）。实验后，通过放射科医师进行的主观评估也显示，该深度学习模型的表现与专业放射科医生相当。此外，研究还整合了Grad-CAM可视化技术，为模型的预测提供了有用的解释，使其成为快速诊断或提供颅内出血检测第二意见的可行选择。



**图2 CNN+LSTM模型架构图**

在2021年的论文中[3]，Wang等人开发了一种用于自动检测和分类头部CT扫描中急性颅内出血（ICH）的深度学习算法。该研究通过集成了二维卷积神经网络（CNN）和两个序列模型来模仿放射科医师的解读过程，实现了对急性ICH及其亚型的精确诊断。此算法使用了超过25,000份CT扫描的2019-RSNA脑部CT出血挑战数据集进行开发和验证，展示了出色的性能和鲁棒的泛化能力。

算法的独特之处在于其多阶段设计，其中每个阶段均对模型的预测结果进行优化。第一阶段的CNN分类器对每个CT切片进行初步的ICH存在性评估和特征提取。随后的两个序列模型使用这些特征来考虑3D上下文信息，从而精细化对每个切片的ICH检测和分类。最终，该系统在国际比赛中荣获第一名，并在两个独立的外部验证数据集上保持了高准确率，验证了其作为日常临床应用中的二次阅读或分流工具的实用性。



**图3 CNN和两个序列模型结构图**

# 创新方法

## RLANet(ResNext-LSTM-Attention Network)

受到阅读的文献的启发，我们构建了RLANet。RLANet 是一种用于医学图像分析的深度学习模型，结合了ResNext101图像分类器和LSTM网络，并引入了注意力机制。该模型通过提取图像特征嵌入来处理序列数据，从而实现对颅内出血的检测。通过分层次的训练和推理方法，RLANet 在提高分类准确性和鲁棒性方面表现出色。

* **输入**

图像输入：

主要输入是处理后的医学图像（DICOM格式），这些图像在预处理阶段被裁剪、缩放，并转换为JPEG格式。

图像大小为512x512，经过归一化处理。

图像数据通过DataLoader加载，批量输入到模型中。

标签输入：

每张图像对应的标签是一个包含六个分类的向量，分别对应于六种不同类型的颅内出血：

Epidural（硬膜外）

Intraparenchymal（脑内出血）

Intraventricular（脑室内出血）

Subarachnoid（蛛网膜下腔出血）

Subdural（硬膜下）

Any（任意一种出血）

* **输出**

模型输出一个形状为(batch\_size, 6)的张量，每个元素表示对应类别的预测概率。通过torch.sigmoid函数将输出值转换为概率值（0到1之间）。

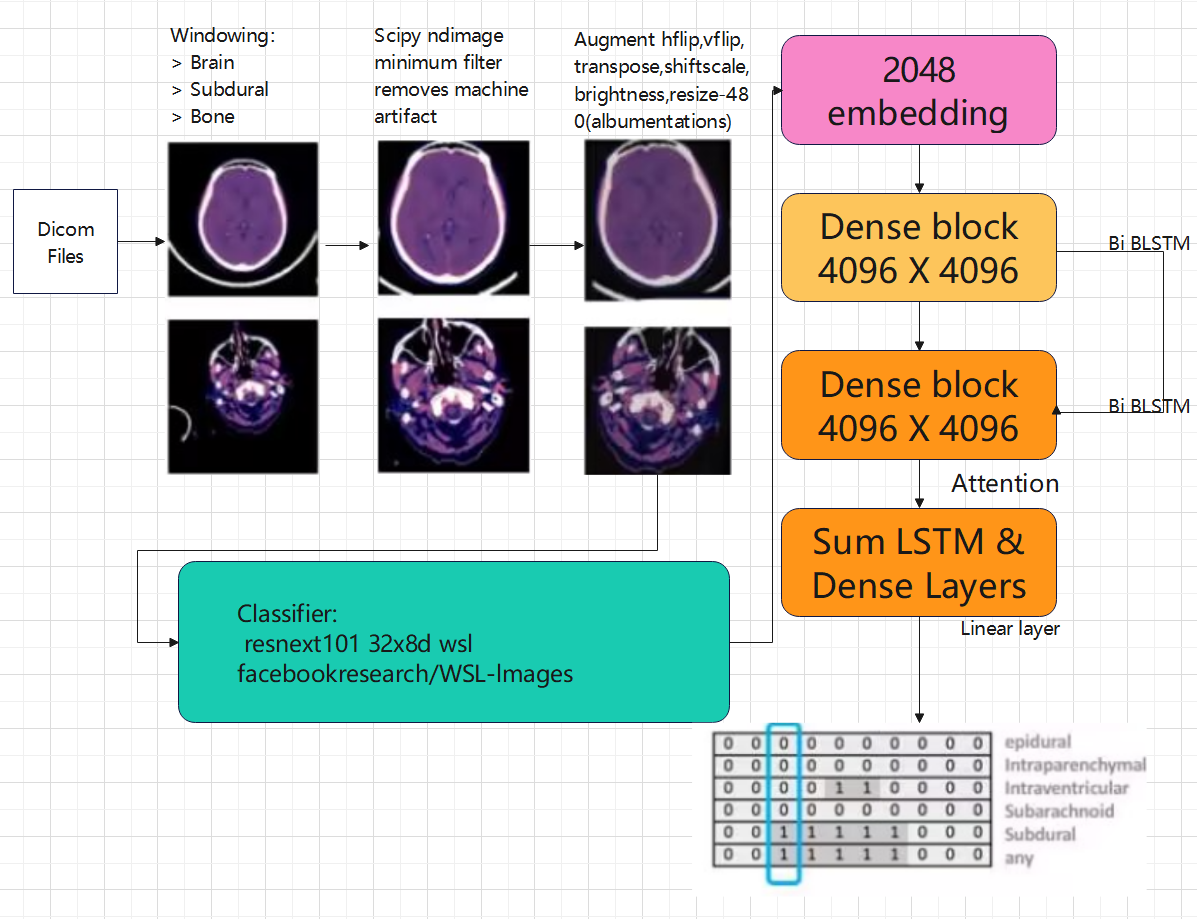
* **网络结构**

该模型主要是基于卷积神经网络（CNN）和长短期记忆网络（LSTM）构建的。

* 图像分类器：用于从图像中提取高维特征（嵌入）。
* LSTM网络：用于处理序列数据（特征嵌入序列），结合注意力机制来增强对关键特征的关注。

**项目流程包括**

1. DICOM图像预处理和增强
2. 图像分类器特征提取
3. 序列创建
4. LSTM模型训练与推理
5. 通过优化和组合两种不同管道的预测结果来提高模型的整体性能。



**图4 RLANet模型架构图**

**CNN部分：**

使用预训练的ResNeXt101模型作为特征提取器，将它的最后一个全连接层之前的特征图作为输入到LSTM模型的特征嵌入。

模型结构：

输入层：接收大小为(batch\_size, 3, 224, 224)的RGB图像输入。

初始卷积和池化层：7x7卷积层，步幅为2，输出通道数为64。最大池化层（Max Pooling），步幅为2。

残差块：每个残差块由多个Bottleneck层组成，包含1x1卷积、3x3卷积和另一个1x1卷积。

全局平均池化层（GAP层）：通过全局平均池化，将每个特征图的大小缩小到1x1。

全连接层：最后的全连接层将特征图映射到类别预测。将最后的全连接层替换为Identity层，以便直接输出GAP层的特征嵌入。

补充说明：图像经过预处理传入ResNeXt101，预处理包括DICOM图像窗口处理、裁剪黑色空白区域、数据增强等步骤。



**图5 RLANet模型-CNN部分代码**

**LSTM部分：**

LSTM模型的输入和输出

输入：

从ResNext101提取的特征嵌入，形状为(batch\_size, 2048)。

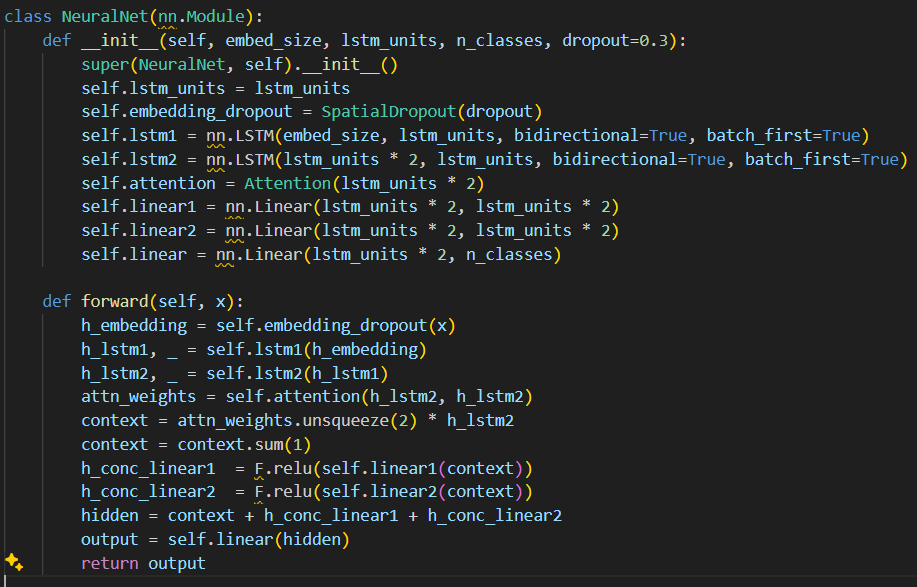
转换为LSTM需要的三维张量，形状为(batch\_size, sequence\_length, embed\_size)。

输出：

最终分类结果，形状为(batch\_size, n\_classes)。

LSTM网络结构：

* 空间Dropout层（SpatialDropout）：在LSTM之前对输入的特征嵌入进行随机丢弃，以减少过拟合。
* 双向LSTM层（BiLSTM）：用于处理序列数据。
* 注意力机制（Attention）：增强模型对关键特征的关注。
* 线性层：用于进一步处理LSTM的输出，并进行分类。



**图6 RLANet模型-LSTM部分代码**

**LSTM训练：**

每个epoch的每个折叠上提取嵌入层，使用TTA（测试时间数据增强）并平均嵌入，分别训练一个独立的LSTM模型12个epoch。总共训练了15个LSTM模型（3个折叠的图像模型 × 5个epoch），并对预测结果取平均值。

通过结合图像分类器和LSTM模型，该模型成功地实现了对颅内出血的检测。引入注意力机制后，模型能够更好地捕捉关键特征，从而提高分类的准确性。

## MedNet

图像预处理步骤：

1. 读取DICOM图像及其前后帧图像。

2. 对每个图像进行窗口调整和重缩放。

3. 将图像保存为JPEG格式。

4. 生成包含前后帧信息的图像组合。

处理原因：

1. 结合前后帧信息：在医学影像中，特别是脑出血检测，前后帧的信息可以提供重要的上下文，有助

2. 于提高模型的检测精度。例如，连续的图像可以帮助模型更好地理解出血的形态和位置变化。

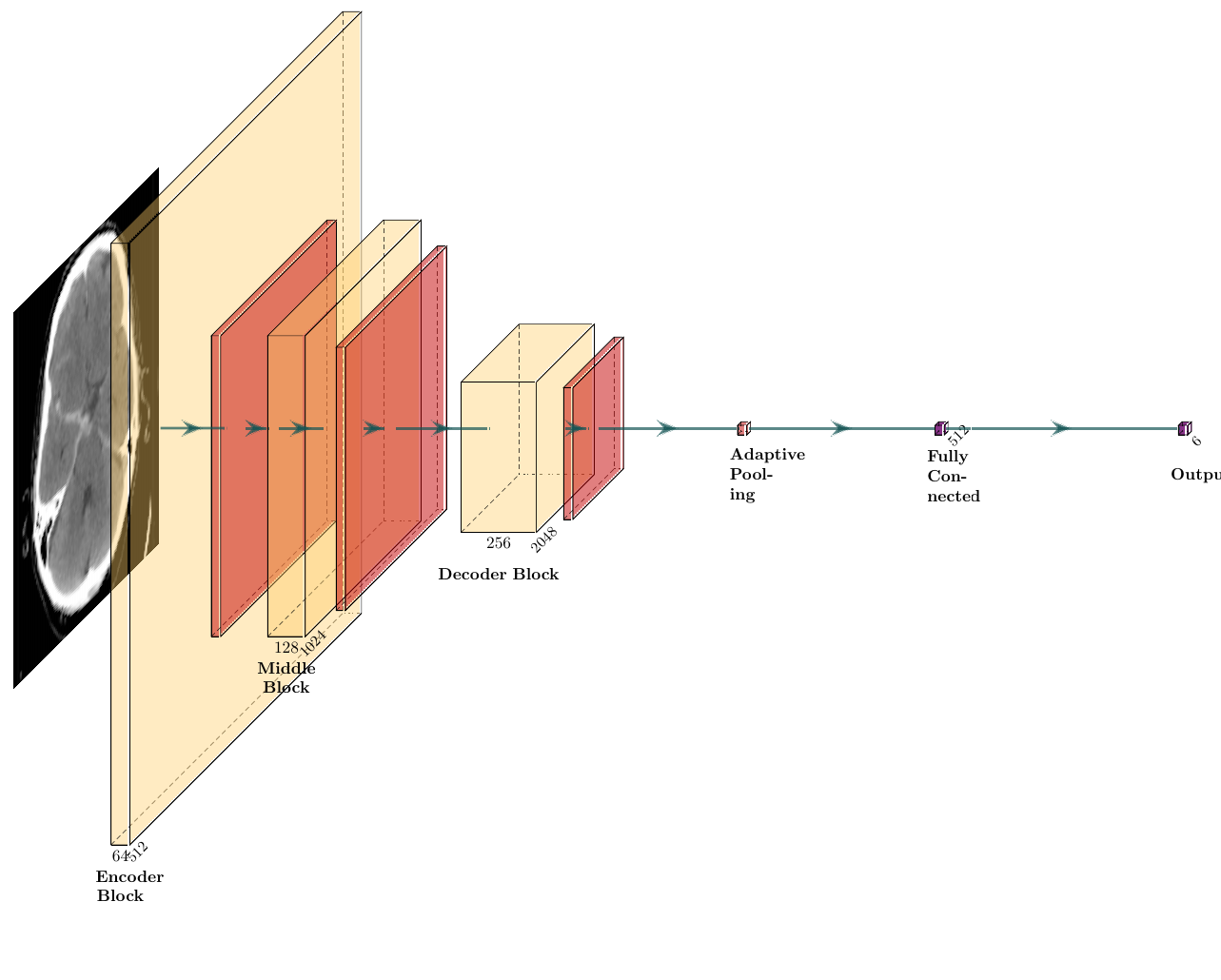
3. 增强数据：通过加入前后帧的信息，模型可以获得更多的训练数据和上下文信息，从而增强模型的

4. 学习能力和泛化能力。



**图7 MedNet模型-图像预处理流程**

**网络结构**



**图8 MedNet模型架构图**

* 1. 编码器（Encoders）

该项目使用多个预训练的编码器，从pretrainedmodels和torchvision 库中加载。不同的编码器具有不同的输出特征图尺寸。

se\_resnext50\_32x4d 和 se\_resnext101\_32x4d：

* SE-ResNeXt 网络，使用 Squeeze-and-Excitation 块来增强通道之间的特征。
* 输出特征图尺寸为 2048。

inceptionresnetv2：

* 混合了 Inception 模块和 ResNet 模块的网络。
* 输出特征图尺寸为 1536。

resnet34 和 resnet50：

* 标准的 ResNet 网络，使用基本的卷积块和残差连接。
* resnet34 的输出特征图尺寸为 512，resnet50 的输出特征图尺寸为 2048。

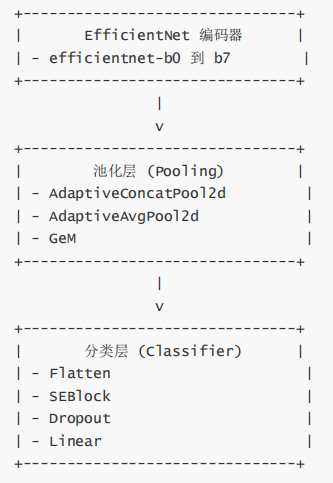
resnet50\_cbam：

* 在 ResNet-50 的基础上加入了CBAM（Convolutional Block Attention Module）模块。
* 输出特征图尺寸为 2048。
  1. CnnModel类

CnnModel 类初始化时选择不同的编码器，并根据选择的池化方式对特征图进行池化操作。支持的池化方式包括 concat （自适应拼接池化）、 avg （自适应平均池化）和 gem （广义均值池化）。

* AdaptiveConcatPool2d: 自适应拼接池化，将最大池化和平均池化的结果拼接在一起。
* Flatten: 展平层，将多维张量展平成一维。
* SEBlock: Squeeze-and-Excitation 块，用于通道注意力机制。
* Dropout: 随机丢弃部分神经元，防止过拟合。
* Linear: 全连接层，用于分类。
  1. Efficient类

Efficient 类使用 EfficientNet 编码器，并根据池化方式进行特征图池化和分类。



**图9 MedNet模型-Efficient类**

# 实验结论

使用Kaggle官方比赛指定的加权多标签对数损失进行评估，需要注意的是"any"标签的权重比具体的出血亚型更高。对于每个图像ID，将每个预测概率与其真实标签进行log\_loss计算。最后，损失在所有样本中取平均值。

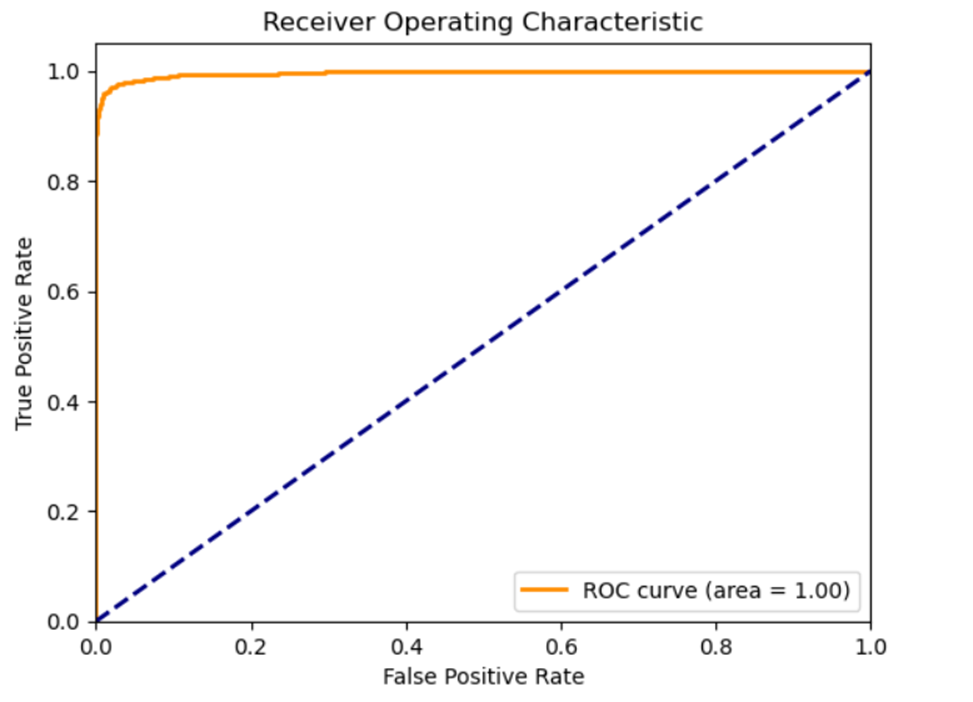
对数损失通过惩罚错误的分类，实现对分类器的准确度(Accuracy)的量化。最小化对数损失基本等价于最大化分类器的准确度。为了计算对数损失，分类器必须提供对输入的所属的每个类别的概率值， 不只是最可能的类别。对数损失函数的计算公式如下：

两个模型最终提交的结果如下表所示：

|  |  |
| --- | --- |
| 模型 | 得分（损失） |
| RLANet | 0.05434 |
| MedNet | 0.05825 |

若以有无颅内出血作为问题，我们可以把原先的多分类问题转为二分类问题，从而进行ROC计算、混淆矩阵、F1分数等指标的计算，如下所示：

RLANet：

****

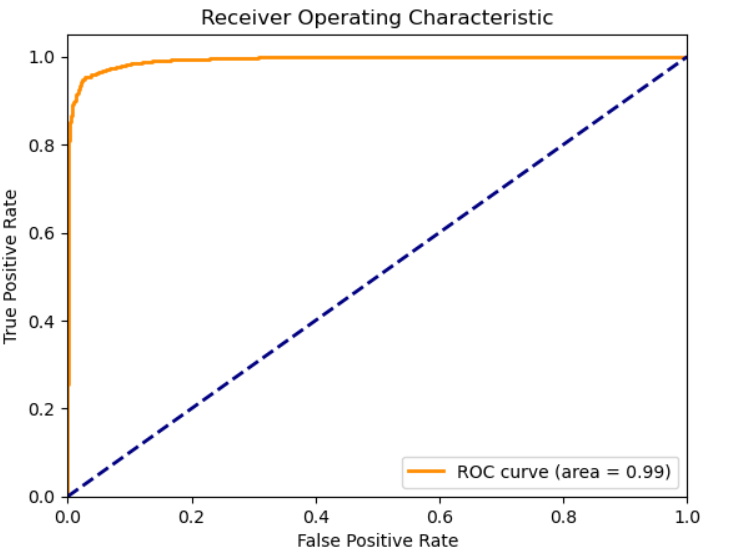
**图10 RLANet模型ROC曲线图**

混淆矩阵：

[[ 972 23]

[ 51 1732]]

MedNet：



**图11 MedNet模型ROC曲线图**

混淆矩阵：

[[ 944 51]

[ 64 1719]]

其他指标：

|  |  |  |
| --- | --- | --- |
| **性能指标** | **MedNet** | **RLANet** |
| 准确率 (Accuracy) | 0.9586 | 0.9733 |
| 精确率 (Precision) | 0.9711 | 0.9869 |
| 召回率 (Recall) | 0.9641 | 0.9714 |
| F1分数 (F1 Score) | 0.9676 | 0.9791 |

# 讨论

分析RLANet模型性能优于MedNet模型的原因：

1. **模型架构差异**

**RLANet**

* 图像分类器（ResNext101）：ResNext101是一个强大的图像分类器，具有高效的特征提取能力。
* LSTM网络：LSTM（长短期记忆网络）擅长处理序列数据，能够捕捉图像之间的时间依赖性。
* 注意力机制：通过注意力机制，模型可以更好地关注关键特征，提高对重要信息的捕捉能力。

**MedNet**

* 多编码器结构：MedNet使用多个预训练的编码器，如SE-ResNeXt、InceptionResNet等，每个编码器提取不同的特征。
* 自适应池化层：使用自适应池化层进行特征提取和融合。
* 前后帧信息：通过引入前后帧信息，MedNet可以获取更多的上下文信息。

1. **特征提取与处理能力**

**RLANet**

* 专注于单个强大的编码器：RLANet使用单个强大的ResNext101编码器，这样可以确保特征提取的一致性和高质量。
* LSTM处理序列数据：LSTM网络擅长处理序列数据，能够有效地捕捉图像序列中的时间依赖性。
* 注意力机制：注意力机制可以增强模型对关键特征的关注，提高特征提取的有效性。

**MedNet**

* 多编码器特征融合：虽然多编码器可以提取更多样的特征，但不同编码器的特征可能存在冗余或冲突，增加了融合的复杂性。
* 前后帧信息引入：前后帧信息的引入虽然可以提供更多的上下文信息，但也增加了模型的复杂性和训练难度。

1. **模型优化与组合**

**RLANet**

* 优化与组合：通过优化和组合两种不同管道的预测结果（有无预处理滤波器），提高了模型性能。
* 分层次的训练和推理：分层次的训练和推理方法可以更好地利用图像分类器和序列模型的优势，提高整体性能。

**MedNet**

* 复杂性与训练难度：多编码器结构和多种池化方式的组合增加了模型的复杂性，训练难度较大。
* 模型一致性：不同编码器特征的融合可能导致模型的一致性较差，影响最终的预测效果。通过我们的验证，去除ResNet34的模型性能反而得到提高，我们分析原因如下：每个编码器都会从输入数据中提取特征。如果多个编码器提取的特征存在较大重叠，那么这些编码器可能并没有增加额外的有用信息，反而增加了模型的复杂度和运行时开销。去除重复或效率低的编码器会让剩余的编码器工作更有效率。

1. **实际效果与评估**

RLANet通过结合图像分类器和LSTM网络，并引入注意力机制，在处理颅内出血检测任务时表现出色，具有较高的检测精度。而MedNet虽然采用了多编码器和前后帧信息融合的策略，但由于其结构复杂性和特征融合的难度，可能导致模型在实际效果上不如RLANet。

综上所述，我们认为RLANet在特征提取的一致性、序列数据处理能力以及注意力机制的引入方面具备优势，从而在颅内出血检测任务中优于MedNet。

# 未来展望

* 1. **模型优化与改进**

模型架构改进：

* 探索Transformer网络： Transformer架构由于其内置的自注意力机制，在捕捉长距离数据依赖方面表现卓越。这对于解析连续CT图像序列中的复杂模式特别有价值。将Transformer应用于医学图像分析，可以更有效地理解图像序列间的内在联系。
* 混合模型： 结合3D CNN和循环神经网络（如LSTM或GRU）可以利用3D CNN处理空间信息的能力和LSTM处理时间序列数据的优势。例如，3D CNN可以首先用于提取每个CT切片或一组切片的空间特征，然后通过LSTM来分析这些特征随时间的变化，以捕捉病变的动态发展。
  1. **数据增强与预处理**

更多数据增强技术：

* 图像混合（Mixup）： 通过在训练阶段将来自两个不同图像的像素值进行加权混合，可以生成新的训练样本。这种技术能增强模型对输入数据变化的鲁棒性，有助于模型学习更加泛化的特征表示。
* 随机擦除（Random Erasing）： 在训练过程中随机选择图像的一部分并擦除其内容，可以模拟遮挡和丢失数据的情况，增强模型对不完整数据的处理能力。
  1. **序列建模与注意力机制**

改进注意力机制：

* 自注意力和多头注意力机制： 这些机制能够让模型在处理每一个图像或序列时，更加聚焦于关键区域，例如ICH的具体位置。通过并行地计算多个注意力头，不同的头可以捕捉到图像不同的特征，从而丰富模型的表征能力。
  1. **模型训练与推理**

结合强化学习和自监督学习：

* 强化学习： 在医学影像分析中应用强化学习，可以通过设计奖励机制来指导模型自主学习更有效的扫描策略，例如优化扫描路径或调整参数设置，以最大化疾病检测的准确率。
* 自监督学习： 利用大量未标注的医学图像数据，可以通过自监督学习方法预训练模型，例如通过预测图像的某部分或对图像进行重构，从而无需大量标注数据也能有效提升模型性能。

# 参考文献

1. Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. Journal of Digital Imaging, 32(4), 582–596. https://doi.org/10.1007/s10278-019-00227-x
2. Burduja, M., Ionescu, R. T., & Verga, N. (2020). Accurate and efficient Intracranial Hemorrhage Detection and subtype classification in 3D CT scans with convolutional and Long Short-Term Memory neural networks. Sensors (Basel, Switzerland), 20(19), 5611. https://doi.org/10.3390/s20195611
3. Wang, X., Shen, T., Yang, S., Lan, J., Xu, Y., Wang, M., Zhang, J., & Han, X. (2021). A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. NeuroImage. Clinical, 32(102785), 102785. https://doi.org/10.1016/j.nicl.2021.102785