

多模态在阿尔兹海默症诊断的进展调研报告

方桂安, 刘梦莎, 钟意, 张焯霖, 黄杰辉[†], 唐振超[†], 陈语谦^{*}

摘要—阿尔兹海默症 (AD) 作为主要的神经退行性疾病之一, 已成为导致痴呆问题最常见的原因。随着计算机技术的不断发展, 将计算机辅助诊断技术工具用于 AD 早期分类研究将为临床医生提供重要帮助。我们科创的项目名称为《基于多模态图像和病人姿态以及声音辨识阿兹海默症》, 目标是发表高质量的综述论文。张焯霖负责引言部分, 钟意负责实验部分, 刘梦莎和方桂安负责数据与模型。为了实现这一目标, 在前五周中我们共同进行了充分的调研与分析, 相关内容整理如下; 并共收集了 84 篇相关文献, 为接下来的写作奠定的基础。

关键词—阿尔茨海默症; 轻度认知障碍; 机器学习; 深度学习; 计算机辅助诊断

1. 概述

AD, 即阿尔兹海默症 (Alzheimer's Disease) 为一种不可逆转的进行性神经退行性疾病, 约占目前所有痴呆病例的 70%~80%。其表现为大脑记忆功能逐渐退化, 认知功能逐步降低, 语言表达能力出现障碍, 甚至人格方面发生变化。晚期 AD 患者, 往往个人生活不能自理, 甚至不能识别出家人, 最终记忆稳步恶化。该疾病一般开始于中年或者老年时期, 可能是由神经元和周围蛋白质的积累而引发的, 与神经细胞突触功能障碍、脑部萎缩、脑神经元细胞死亡有关。有研究人员称, 由 AD 所引起的脑部相关变化有可能早于出现 AD 症状至少 20 年。

按照发病过程情况, 可将 AD 分为: 正常对照组 NC(Normal Controls)、主观记忆抱怨 SMC(Subjective Memory Complaints)、轻度认知功能障碍早期 EMCI(Early Mild Cognitive Impairment)、轻度认知功能障碍晚期 LMCI(Late Mild Cognitive Impairment) 和 AD。其中, 对于早期的 AD 患者, 被称为轻度认知障碍阶段 MCI(Mild Cognitive Impairment)。这一阶段被认为是从健康者到 AD 患者的过渡。统计显示, 患有

MCI 的 65 周岁及以上的人员中, 有近 40% 的患者会在接下来的 5 年内发展成为 AD 患者。

目前关于该疾病的发病原因还有待明确, 同时由于 AD 的病情多样化, 这给疾病的预防、诊断和治疗造成了非常大的困难。研究发现, 从 MCI 到 AD 的年转化率高达 10%。AD 患者提早被确诊难度大, 且晚期患者治疗效果不明显, 对 AD 进行早期分类研究具有重要临床意义。

由临床可知, AD 的确切诊断需要依靠患者死亡后的脑组织检查。随着计算机技术的不断发展, 计算机辅助 AD 早期诊断的分类研究成为可能。机器学习 ML(Machine Learning) 在医学诊断中得到了广泛应用, 特别是近些年来, 一些先进的脑部神经成像技术, 如核磁共振成像技术 MRI(Magnetic Resonance Imaging)、正电子发射断层扫描技术 PET(Positron Emission computed Tomography) 等不断发展成熟, 一些高维度的更大规模的神经影像图像为计算机辅助诊断 CAD(Computer Aided Diagnosis) 提供了更广泛的素材, 也极大地激发了广大科研人员将此用于 AD 早期诊断的研究热情, 同时也为研究者开发更好的诊断模型提供了有力支持。

本文旨在展示我们的调研结果, 基于计算机辅助诊断技术的 AD 早期分类研究, 把传统机器学习和深度学习应用于 AD 的早期诊断中。计算机辅助 AD 检测系统的典型框图如图 1 所示。

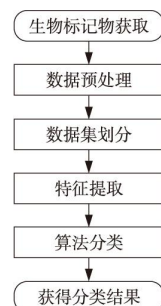


Fig. 1. 计算机辅助 AD 分类系统框图

方桂安, 20354027, (e-mail: fanggan@mail2.sysu.edu.cn)。

刘梦莎, 20354091, (e-mail: liumsh6@mail2.sysu.edu.cn)。

钟意, 20320229, (e-mail: zhongy276@mail2.sysu.edu.cn)。

张焯霖, 20354156, (e-mail: zhangzhlin8@mail2.sysu.edu.cn)。

本项目由陈语谦^{*}老师和黄杰辉[†]、唐振超[†]师兄进行指导。

本报告利用当下先进的复杂高维度脑部神经影像数据 (如 MRI、PET 等) 和其它有价值的生物学标记 (声音、脑电图、脑磁图、手写字迹、身体、精神检查、神经心理学测试) 等多种模态的数据, 对 AD 早期阶段 MCI 患者进行分类研究, 对比采用不同生物学标记和不同分类方法的计算机辅助诊断研究, 通过不同方式的分类方法, 总结了近些年来将计算机辅助工具用于人脑医学影像数据的 AD 早期诊断分类研究现状, 并总结了我们对此调研的参考文献。

II. 研究方法

采用计算机辅助手段进行 AD 早期诊断的研究方法目前主要有传统的机器学习方法和深度学习方法。前者在 AD 早期诊断方面已经有了十余年的发展历程, 有比较成熟的研究, 研究者们采用不同的机器学习算法对 AD 进行分类研究。

深度学习作为机器学习的一个新的发展分支, 在近年来取得了很大的发展, 已被广泛应用于语音识别、机器翻译、自动驾驶和机器视觉等多领域。凭借深度学习更深层次的网络模型, 可提取到更加抽象的图像特征, 进而得到更高的分类准确率。在深度学习方法的有关研究中, 特别是卷积神经网络 CNN(Convolutional Neural Networks) 的使用, 其分类结果已经被证明优于现有的机器学习方法。

本文将从传统机器学习和深度学习两方面来描述基于计算机辅助诊断的 AD 早期分类研究。

A. 传统机器学习方法

使用传统机器学习方法进行 AD 分类研究一般需经过以下步骤:

- 1) 数据预处理
- 2) 特征提取
- 3) 特征选择
- 4) 基于特征分类

1) 数据预处理: 在获取到医学影像学数据等 AD 生物学标记后, 需要对所获得的数据进行预处理, 主要的预处理环节包含了前连合 AC (Anterior Commissure) 和后连合 PC(Posterior Commissure) 对齐校正、颅骨分离、脑组织分割和配准等。AC-PC 线用于基准线校正, 防止因获得的影像数据未确定 AC-PC 基准线而给后续实验造成误差。颅骨剥离, 排除非脑部结构, 减少不必要的干扰。颅骨剥离后, 需对图像不均匀部分进行校正。

由于磁共振等影像采集设备的工作特点, 所采集的影像中难免混杂有噪声, 滤除影像中的噪声, 对于实验是非常必要的。图 2 展示了不同预处理过程的分类情况。

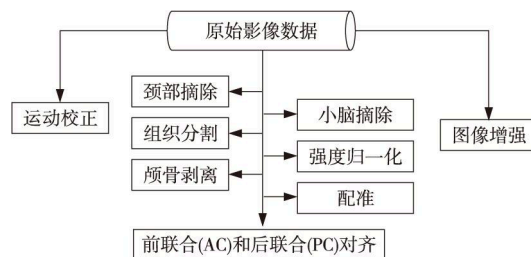


Fig. 2. 原始影像数据预处理

2) 特征提取: 特征提取是机器学习中模式识别和图像处理过程中的常见应用。旨在从原始数据中除去因特征过多而产生过拟合问题的冗余部分, 以便节省内存空间和提高计算能力, 改善分类效果, 同时也属于降维过程。

特征提取重点针对所提取特征的大小、形状和体积等要素来进行分类研究。AD 早期诊断研究将主要涉及到海马区、内嗅皮层结构等。常见的特征提取对象主要有以下几种形式:

- 1) **基于体积特征:** 已有研究指出, 有关海马区在 AD 早期已出现萎缩, 研究显示约 50% 的 AD 患者存在海马区萎缩。因此, 可将海马体积作为 AD 诊断的生物标志物特征。
- 2) **基于厚度特征:** 研究发现, AD 患者相对于正常对照组 NC 的整个脑部皮质厚度有明显差异, 反映为大脑多部位的皮质厚度减小。已有的临床研究发现, 脑部萎缩引起的脑部体积变化是 AD 早期阶段出现的一重要生物标记物。
- 3) **基于形态特征:** 海马区、内侧颞叶萎缩所造成的脑部体积变化, 已作为生物标记物特征用于 AD 早期检测中。研究发现, AD 早期患者的脑部海马区还未出现萎缩时就已经存在形态上的改变。以形态特征作为生物标记物特征用于 AD 早期诊断, 表现更加敏感的特性。
- 4) **基于纹理特征分析:** 纹理是一种视觉特征, 反映影像或图像的同质现象, 体现物体表面缓慢变化或者呈现周期性变化的组织结构特征属性。Freeborough 等采用纹理特征进行分类时发现, 可以通过纹理质地分析得到 AD 患者和正常对照组 (NC) 之间显著的差异。
- 5) **基于体素形态学分析:** 体素 (Volume Pixel) 即

体积元素, 相对于二维空间的最小单位——像素, 是数字数据于三维空间分割的最小单位。体素形态学分析 VBM(Voxel-Based Morphometry) 是基于体素单位针对 MRI 医学影像的综合、客观和自动分析方法, 能够用来对活体人脑组织进行形态学方面的科学研究。通过 VBM 定量计算, 分析 MRI 影像中不同体素所对应的脑灰质、白质密度和体积的相对变化, 可用于 AD 早期诊断研究。Karas 等采用 VBM 技术进行 AD 研究时发现, AD 患者和正常对照组存在 GM(Grey Matter) 体素形态学上的差异。

3) **特征选择**: 特征选择指的是从一组特征中选择出最有效的特征, 以达到降低维度的目的, 这是模式识别过程中的一个重要环节。特征选择通过对脑部感兴趣区域 ROI(Region of Interest) 进行分割获取。根据实验设定, 可以将 ROI 定位为全脑区、海马区、脑灰质 (GM)、脑白质 (WM)、脑脊液 (CSF) 和内侧颞叶区等区域。在所有的脑部 ROI 中, 大脑海马区是最早受影响的区域之一, 是 AD 病因学的重要解剖区域。Amoroso 等提出将双侧海马的形状和体积特征作为生物标记物用于 AD 早期分类研究。

4) **基于分类算法**: 采用传统机器学习手段用于 AD 早期诊断分类研究已不断趋于系统化、成熟化。机器学习辅助手段用于 AD 早期分类的算法主要有以下几类:

- 1) **支持向量机 SVM**: 支持向量机 SVM 是一类监督学习方式, 是能对数据进行二分类的广义线性分类器。决策边界是样本数据求解的最大超平面。在 AD 早期诊断分类研究中, SVM 用于 AD 分类的频率是最高的。Kim 等利用支持向量机提取特征后, 再进行 AD 分类。
- 2) **Logistic 回归**: Logistic 回归又被称为 Logistic 回归分析, 是一种广义线性回归分析模型, 在医学影像中多用于疾病诊断, 常用于二分类研究。当学习训练样本量大时, 可以取得较好的分类效果。Ieracitano 等采用 Logistic 回归算法, 基于单模态或多模态生物特征进行 AD 分类。
- 3) **线性判别分析 (LDA)**: 线性判别分析算法是对费舍尔的线性鉴别方法的归纳, 其目的在于寻找 2 类失误特征的线性组合来分类, 常用于评估患者病情程度和对疾病的预测判断等。当样本数据较少时, 也可取得较 Logistic 回归更好的分类效果。Ahmad 等利用生物标记物模式和静息状态下的

功能磁共振成像 (fMRI), 标记影响较大的海马区 (HP)、颞中回区 (MTG)、内嗅皮层和后扣带回 PCC (Posterior Cingulate Cortex) 等区域, 再采用 LDA 对 AD 进行分类研究。

- 4) **贝叶斯分类器**: 贝叶斯分类器, 通过被分类对象的先验概率, 用贝叶斯公式计算得到被分类对象属于某一类别属性的最大概率值。该分类器拥有较小的分类错误率, 常用于模式识别等领域。Yubraj 等提出将皮质厚度和皮质下体积作为生物学标记用于 AD 分类, 使用主成分分析 PCA(Principal Component Analysis) 方法进行降维, 并使用贝叶斯分类器进行分类研究。
- 5) **随机森林**: 随机森林是一种融合多个决策树算法的分类器, 该分类器输出由每一个决策树输出类别的众数来确定。相对于决策树算法有更高的准确率, 且能处理更多的输入数据。Dimitriadis 等在 AD 分类研究中, 提出使用随机森林算法从特征集的子集选择多个生物特征, 融合这些生物标志物特征采用随机森林算法进行分类。
- 6) **K 近邻 (KNN)**: KNN 算法属于机器学习算法中较为简单的一种, 分类依据是样本数据点特征空间中最近的 K 个样本是否属于某同一个类别。K 近邻算法常用于分类和回归分析中。Farouk 等在结构核磁共振图像 (SMRI) 中提取生物学标志物信息, 融合灰度共生矩阵提取纹理特征和基于体素的形态学特征, 采用 KNN 算法进行 AD 早期分类。

B. 深度学习方法

深度学习是机器学习方法的一种, 属于机器学习的子集。通过深层次的学习过程学习输入数据特征, 其结构类似于模仿人类大脑分析处理信息的神经网络。自 2006 年以来, 由于深度学习计算能力的不断提高和数据量的不断增大, 深度学习已经得到了充分发展, 特别是深度学习模型和算法的逐渐改进, 其输出准确率不断刷新。得益于以上优势, 深度学习已广泛应用于自然语言处理、无人驾驶、语音识别和机器视觉等领域。用于 AD 分类的深度学习网络模型主要有以下几类:

- LetNet-5
- AlexNet
- VGGNet(Visual Geometry Group Network)
- GoogLeNet

- ResNet(Residual Network)
- DenseNet(Dense convolutional Network)
- 3D-CNN(3D-Convolutional Neural Network)

研究者利用这些网络模型或者基于以上模型的改进模型进行 AD 分类研究。

1) 神经网络分类模型:

- **LetNet-5**: LeCun 等提出的卷积神经网络 (CNN) 是第 1 个真正意义上的多层次结构学习模型, LetNet-5 为卷积神经网络 (CNN) 的一种经典框架。LetNet-5 网络模型共有包含输入层在内的 8 层网络构成, 其中包含 2 个卷积层和 2 个下采样层, 属于经典的 CNN 结构。Sarraf 等基于 MRI 影像, 采用 LeNet 模型进行 AD 和 NC 的分类研究。LetNet-5 能够获得原始图像的有效表征, 仅使用极少的预处理也能识别原始图像规律, 但鉴于其模型的特点, 处理较大规模数据和复杂问题时结果欠理想。
- **AlexNet**: AlexNet 是由 Hinton 和 Krizhevsky 等在 2012 年提出的, 该模型首次在 CNN 中成功应用了 Dropout 策略、ReLU 函数和 LRN(Local Response Normalization) 层等, 解决了深层神经网络过拟合的问题, 同时使用 GPU 进行加速运算操作, 实现了在大规模图像数据集上的深层卷积神经网络结构。Afzal 等通过预训练的 AlexNet 网络模型, 选取开放获取系列成像研究数据集 OASIS(the Open Access Series of Imaging Studies) 数据集中 199 名受试者数据进行 AD 和 NC 的分类研究, 分类准确率达到 98.41%。
- **VGG**: VGG 模型是由牛津大学的 Oxford Visual Geometry Group 提出的, 其通过增加网络深度来提升网络最终性能。它的 2 种结构 VGG16 和 VGG19 除了深度不同外, 在本质上没有太大区别。相比 AlexNet 而言, VGG 采用连续的 3*3 卷积核代替 11*11, 7*7, 5*5 的较大卷积核, 实现了增加网络深度来学习更加复杂特征的同时减少模型参数的目的。Jain 等通过 VGG-16 网络模型, 采用 MRI 单模态影像进行 NC、MCI 和 AD 三分类研究, 实验数据来源于 AD 神经影像学计划数据集 ADNI, 选取 50 个样本用于实验, 其分类准确率能达到 99.14%。
- **GoogLeNet**: GoogLeNet 是由 Szegedy 等提出的, 不同于 AlexNet、VGG 等网络结构通过增加网

络层数来获得好的训练效果, GoogLeNet 采用 average pooling 代替参数量大的全连接层, 同时使用 inception 模块来代替一般的卷积层, 1*1 卷积层的使用对模型降维和参数的减少起了重要作用。这些方法, 提高了计算资源的利用率, 使其能够在计算量不变的情况下提取到更多的特征, 训练效果也得到提升。Sarraf 等分别基于 MRI 和 FMRI 模态影像数据, 从 ADNI 数据集中分别获取 144 例 FMRI 和 302 例 MRI 样本数据, 采用 GoogLeNet 网络模型从大量训练图像中提取低层到高层特征, 实验训练集和测试集占比为 3: 1, GoogLeNet 模型在 FMRI 影像下对 AD 和 NC 分类的平均准确率为 94.24%。基于 MRI 影像的 GoogLeNet 和 LeNet 模型的分类平均准确率分别为 98.74% 和 97.88%, GoogLeNet 模型表现更优秀。

- **ResNet**: 残差神经网络 ResNet 于 2015 年被提出, 它“简单与实用”并存, 一些研究方法也都建立在 ResNet50 或者 ResNet101 模型的基础上, 其在目标检测、分类、分割和识别等领域都得到了广泛的应用。ResNet 内部的残差模块使用了 skip connection(跳跃连接)的残差结构, 加深了网络层次, 同时提高了网络性能, 消除了因网络层次加深而出现的梯度爆炸、网络无法收敛和网络退化等问题。特点是易于优化, 并且可通过增加一定的深度来提高准确率。Ji 等分别在单模态 MRI 和多模态 MRI、FDG-PET 影像下, 基于 ResNet 网络模型进行 AD 分类研究。
- **DenseNet**: Huang 等提出了 DenseNet 模型, 虽然其结构并不算复杂, 但特征提取却很高效。ResNet 模型构建前面层和后面层的 skip connection, 有助于梯度反向传播和训练深层网络; GoogLeNet 构建了 inception 模块, 加大了模型宽度, 实现了降维和多尺度再聚合。DenseNet 模型不同于 ResNet 和 GoogleNet 模型, 它建立前面所有层和后面层的密集连接 (dense connection), 并通过特征在 channel 上的连接实现特征重用 (feature reuse), 在计算成本和参数更少的情况下, 得到了较 ResNet 网络更优秀的性能。
- **3D-CNN**: 3D-CNN 是在 2D-CNN 的基础上改变而来的, 是基于多场景中的多幅图像, 或者是在视频连续帧之间的关联信息背景下提出的。其目的在于通过增加新维度 (时间) 信息, 获取图像间

的关联信息,提取更具有表达性的特征,从而克服 2D-CNN 在捕捉时序信息上的缺陷。2D-CNN 的卷积应用于 2D 特征图,3D 卷积通过对堆叠连续帧组成的立方图进行 3D 卷积,使卷积层中的特征图均与其上一层中多个帧相连,以获取空间和时间维度特征。Kompanek 等使用 3D-CNN 基于 MRI 影像对 NC 和 AD 进行分类,采用体数据增强 (Volumetric Data Augmentation) 方法,提高分类准确率的同时也增强了网络模型的泛化能力。

2) 基于多模态深度学习方法:多模态深度学习是相对于单模态而言的,指将多个单模态的信息用于深度学习方法中,以达到实现不同模态间的信息融合的目的。

很多研究者采用多模态数据用于 AD 早期分类研究。Huang 等基于海马区 TI 加权核磁共振成像及 FDG-PET 影像下各种多模态特征信息,采用卷积神经网络进行 AD 分类研究。获得 NC 与 PMCI、SMCI 和 PMCI、AD 与 NC 的分类准确率分别为 87.46%、76.90% 和 90.10%,结果表明结合多模态成像数据的分类结果优于单模态形式的。

Forouzannzhad 等结合正电子发射断层扫描 (PET) 和核磁共振成像 (MRI) 多模态成像技术和标准神经心理学测试评分结果,采用深度神经网络 DNN(Deep Neural Networks) 对 AD 早期诊断进行分类研究。对正常对照组 NC 与早期轻度认知障碍 EMCI 的分类准确率高达 84.0%,对 NC 与晚期轻度认知障碍 LMCI、CN 与 AD、EMCI 与 LMCI、EMCI 与 AD、LMCI 与 AD 的分类准确率分别为 84.1%、96.8%、69.5%、90.3% 和 80.2%,而单独在 MRI 影像上,NC 与 EMCI 的分类准确率仅为 68.0%。研究表明多模态方法优于单模态影像分析。

Kang 等基于 SMRI 和扩散张量成像 DTI(Diffusion Tensor Imaging) 双模态数据,构建 VGG16 模型的迁移学习方法,用于 EMCI 和 NC 的分类研究。数据来源于 ADNI 数据集,实验采用多模态融合策略,将具有相同索引的切片合并成 RGB 切片,组成切片数据集输入模型进行训练,采用 LASSO(Least Absolute Shrinkage and Selection Operator) 算法提取与 EMCI 疾病相关部分特征,实验获得了 94.2% 的分类准确率,敏感度高达 97.3%。实验结果表明多模态数据可以为区分 EMCI 和 NC 提供更多更有用的信息,且从临床角度验证了 DTI。图像可作为 EMCI 的一个重要生物标志物。

Khvostikov 等通过对海马感兴趣区域进行 SMRI 和 DTI 成像模态的融合,并基于 3D-CNN 模型的 AD 分类算法,对比了单模态数据用于实验的情况,在 AD 与 MCI 分类中,获得了多模态情况下 93.3% 的准确率,相对于 SMRI 单模态情况下 65.8% 的准确率有较大优势。实验中为了平衡数据量大小不同的类,采用数据扩充方法进行数据增强,以消除不同大小的数据量对网络训练过程的影响。

III. 实验数据来源

AD 的早期分类研究是一项复杂的任务,所需要的数据样本的特点如下:

- 1) 需要具备一定的规模
- 2) 样本需由专业医生进行标注
- 3) 大量实验数据和标注的获得,往往需要大量的人力、物力

不同研究者所使用的数据样本主要来源于以下数据集:

- 阿尔兹海默症神经影像学计划数据集 (ADNI)
- 开放获取系列成像研究数据集 (OASIS)
- 澳大利亚成像、生物标记技术和生活方式研究数据集 (AIBL)
- 阿尔茨海默病最小间隔共振成像数据集 (MIRIAD)
- 其它合作性的医学影像学数据。

A. ADNI

ADNI 创建于 2003 年,是到目前为止较为成功也是应用最多的针对 AD 的综合研究数据集,它具有多中心、跨越多研究学科特点,主要研究人体脑部 MRI 神经影像、人脑 PET 影像、其他人体生物标记物如脑脊液 (CSF)、血液生物标记物,同时还包含有人体遗传学信息、临床资料、神经心理学评估等信息。已招募来自于全球 59 个国家和地区的 800 名成年人作为受试者,年龄跨度在 55~90 周岁。为全球的 AD 研究提供了非常有价值的临床医学数据,ADNI 已经成为研究人员采用的核心数据资源。

B. OASIS

OASIS 数据集由 2 大类构成,分别是横断面数据集和纵向面数据集。横断面数据集涵盖了 416 名年龄在 18~96 周岁受试者的 MRI 数据资源。纵向面数据集涵盖了 150 名年龄在 60~96 周岁的受试者的 MRI 数据

资源, 且每个受试者一般都有 2 次或 2 次以上间隔满一年的扫描数据。目前, OASIS 是仅次于 ADNI 的核心数据资源。

C. AIBL

AIBL 提供了 1000 名年龄在 60 周岁及以上的受试者有关基线人口统计学、诊断、认知功能、健康和生活方式等方面的调查统计信息。约 25% 的受试者参与了匹兹堡化合物 (PiB PET) 和 MRI 脑成像的淀粉样蛋白 PET 成像扫描。AIBL 对受试者进行超过 18 个月间隔的重复评估, 能更充分地确定不同的生物标志物, 并加强 AD 在认知参数和生活方式等因素参与下的预测性标准。AIBL 在 AD 的科学研究中发挥着重要作用。

D. MIRIAD

MIRIAD 数据集受试者由轻度阿尔兹海默症患者 46 人和正常对照组 23 人组成, 包含了上述受试者一系列的纵向体积 T1 加权 MRI 扫描医学影像图像。以上这些影像均由同一位放射科技师采用相同扫描设备采集的相同序列的 708 次扫描结果组成。在这些图像中, 包含从基线起始的 2 周、6 周、14 周、26 周、38 周、52 周、18 个月和 24 个月不等的扫描间隔。数据集包含了受试者关于性别、年龄和简易精神状态检查得分等情况记录。

E. 其它合作性的医学影像学数据

一些研究人员倾向于使用本地数据集。如在研究 AD 等相关课题时, 使用有些医院的存档数据或与医院合作课题研究的就诊者检查数据。吕鸿蒙等通过改进的 AlexNet 神经网络模型, 对阿尔兹海默症进行多分类研究。为了研究出符合中国居民实际的分类诊断系统, 除采用 ADNI 数据影像之外, 还采用了首都医科大学附属北京天坛医院提供的数据。

IV. 常用评价指标

基于计算机辅助技术的 AD 分类研究, 判断其方法是否具有有效性、可靠性和普遍适用性, 能否用于临床辅助诊断, 主要用以下几个衡量标准来评价:

- 准确率 (Accuracy)
- 敏感度 (Sensitivity)
- 特异度 (Specificity)
- 精确度 (Precision)

- ROC(Receiver Operating Characteristic curve) 曲线下的面积 AUC(Area Under Curve)。

准确率是在给定测试数据集上, 分类器能正确分类的样本数和总样本数的比值, 用 ACC 表示, 反映了分类器准确分类出 AD、MCI 和 NC 的概率, 通常其值越高, 分类器性能越好; 敏感度又称真阳性率, 表示所有真阳性 (正例) 能被正确检测的概率, 用 TPR 表示, 通常其值越大, 表示 AD 病例漏诊就越少; 特异度一般用 TNR 表示, 又被称为真阴性率, 表示所有真阴性 (负例) 被正确检测的概率, 衡量分类器对真阴性的识别能力, 在 AD 分类中体现分类器能准确分类出 NC 的能力, 一般其值越大, 表明正常人被误诊为 AD 的几率就越小; 精确度即 PPV, 指被分为真阳性 (正例) 的样本实际为真阳性的比例; ROC 曲线下的面积 AUC 是指以真阳性率 (TPR) 为纵坐标, 假阳性率 (FPR) 为横坐标所绘制图形的曲线下面积, 通常其值越大, 分类器分类性能越好。计算公式如下:

$$\text{准确率 ACC(Accuracy): } ACC = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{敏感度 TPR(Sensitivity): } TPR = \frac{TP}{TP + FN}$$

$$\text{特异度 SPC(Specificity): } SPC = \frac{TN}{TN + FP}$$

$$\text{精确度 PPV(Precision): } PPV = \frac{TP}{TP + FP}$$

其中, TP(True Positive) 表示预测为阳性, 实际也为阳性的样本数量; FP(False Positive) 表示预测为阳性, 实际不是阳性的样本数量; FN(False Negative) 表示预测为阴性, 实际为阳性的样本数量; TN(True Negative) 表示预测为阴性, 实际也为阴性的样本数量。

V. 文献综述

当涉及到撰写一篇综述文章时, 参考文献的重要性无法被低估。一篇综述文章的目标是为读者提供一个有关特定话题的全面和准确的概述。这对于提高对某一特定话题的了解, 以及做出更明智的决策都是至关重要的。

通过引用相关的研究文献, 读者可以了解一个给定话题的历史、现状以及未来趋势。此外, 参考文献也可以帮助综述作者确定哪些方面需要给予更多关注, 以及哪些候选措施已经被探索过和研究过了。

对于阿尔兹海默症诊断多模态方法的发展, 参考文献的重要性更加凸显。这种疾病是一个通过多种手段进

行诊断的疾病，因此，了解不同类型的数据集和评估标准至关重要。通过查阅先前的研究文献，我们可以更好地了解研究领域的发现，并深入探讨当前研究中的缺陷，为未来的工作提供思路。在这一部分中，将展示我们是如何调研相关工作，收集相关文献的。

A. 文献收集

首先我们在论文阅读平台 ReadPaper 成立了研究小组，所有人都将在各个搜索引擎中找到的高质量论文上传，包括图像、姿态、声纹、临床数据、多模态等多个领域以及传统机器学习（统计学、特征工程等）和现代深度学习（CNN、transformer 等）两个方向。使用到的论文检索引擎如表 I 所示。

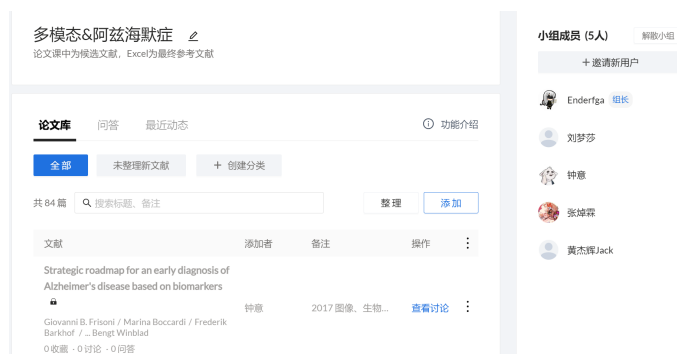


Fig. 3. 小组论文库

如图 3 所示, 我们目前共收集了 84 篇文献, 在完成该项目的过程中这个数字可能会继续上升。利用论文库我们可以快速得和组员同步论文信息, 防止重复工作, 便于后续整理期刊、影响因子等相关信息。

B. 文献分析

虽然通过搜索引擎检索到了一定量的参考文献，但对于其质量还需要进一步筛选，我们通过以下步骤进一步在共享文档中整理了 Excel 表格：

- 1) 首先, 浏览文献的摘要和标题, 了解文献涉及的主题和内容。
- 2) 然后, 查看文献的作者和出版物的来源。如果文献出自权威杂志或被权威学者写作, 则可能可信度更高。
- 3) 仔细阅读文献, 注意其中提到的引用和参考文献。如果其中有许多权威的引用或参考, 那么该文献可能更可靠。
- 4) 注意文献中使用的语言和术语。如果语言和术语正确并且明确, 那么文献可能更可信。

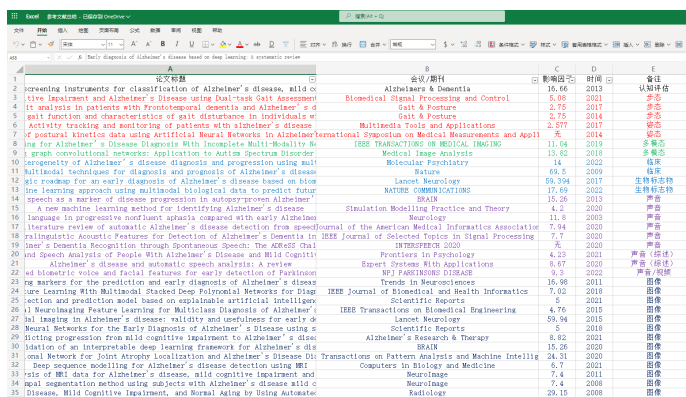


Fig. 4. 参考文献统计

- 5) 查阅该文献的影响因子, 被引频次、H 指数及其分区等相关信息。

最终，我们整理出了预计会使用在综述论文写作中的参考文献，利用不同的颜色标出不同的模态，如图 4 所示。

影响因子指的是某一科学期刊被引用的频率,通常被认为是该期刊学术水平或重要性的评价标准。然而,影响因子并不是完美的评价指标,它有一定的局限性和缺陷,导致它并不能完全反映一个期刊的真实学术质量。比如它只反映了短期内的引用情况,而对于一些学科领域而言,引用存在滞后的现象。同时,影响因子过度重视期刊文章的数量,而非质量,因此可能助长了“以量取胜”的不良趋势。因此,在评估期刊,特别是评价特定领域的期刊时,我们应该同时考虑多个指标,而不仅仅是影响因子。所以我们目前利用影响因子作为对文献的评判标准还是比较片面的(如图 5 所示),在后续的写作中要考虑作者、期刊等多方面的内容。

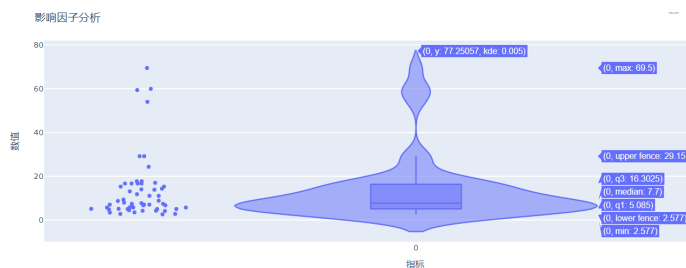


Fig. 5. 影响因子分布分析

可以看出,阿尔兹海默症在 Nature 等著名期刊上也时常发表相关文献,部分文章的影响因子高达 70,这表明该领域的研究深度和广度都非常的不错,我们也在尽可能收集这类高质量文章。大部分文献的影响因子分布在 5 到 20 之间,符合我们的初步判断,说明阿尔兹

TABLE I
科研搜索引擎

| 论文检索 | 网址 | 简介 |
|------------------|-------------------------|-------------------------------------|
| web of science | www.webofscience.com | 大型综合性、多学科、核心期刊引文索引数据库 |
| 中国知网 | www.cnki.net | 中国最大的数字图书馆之一，提供多种论文资源的数字化服务平台 |
| ReadPaper | readpaper.com | 一个专注于学术论文分享和交流的社区平台 |
| 百度学术 | xueshu.baidu.com | Baidu 推出的学术搜索引擎平台 |
| 谷歌学术 | scholar.google.com | Google 推出的学术搜索引擎平台 |
| Semantic scholar | www.semanticscholar.org | 一个由 AI 驱动的学术文献搜索引擎和数据库 |
| paper with code | paperswithcode.com | 旨在提供最新技术研究论文和相应公开代码的在线平台 |
| PubMed | pubmed.ncbi.nlm.nih.gov | 由美国国立卫生研究院 (NIH) 提供的免费生命科学和医学领域的数据库 |
| Arxiv | arxiv.org | 一个收集物理学、数学、计算机科学与生物学的论文预印本的网站 |
| ArxivXplorer | arxivxplorer.com | 一个利用 ChatGPT 的可视化论文检索引擎 |
| AMiner | www.aminer.cn | 科技情报分析与挖掘平台，利用 AI 为用户高效获取和理解科研 |

海默症的研究进展正朝着积极的方向发展。

VI. 结语

对近些年文献的研究发现，基于计算机辅助诊断技术的 AD 早期分类研究已经取得了前所未有的发展，利用人工智能手段进行生物医学研究也逐步趋于成熟，采用传统机器学习分类算法和深度学习方法进行 AD 分类的准确率不断提高，但即便如此，也仍然存在一些问题，这对未来用于 AD 早期阶段的研究带来了挑战。这些问题主要集中在构建数据集、数据处理和模型推广泛化能力等方面。我们希望我们的综述论文能够帮助科研工作者充分做好迎接未来挑战的准备。

在数据集方面，依然存在数据集资源有限和部分数据缺失的情况。利用计算机辅助方法进行 AD 早期分类研究时，分类准确率在很大程度上取决于数据样本量的大小和受试者不同模态的生物标记物是否缺失。小的样本集在网络训练过程中，易出现过拟合现象，同时训练集和测试集数据相互混杂、破坏相互独立性等情况都会对最终分类准确率带来负面影响。与单模态神经成像方法相比，多模态成像方法在分类上有明显优势，表现为时间和空间分辨率的提高，多样化生物信息功能互补等。由于在最初采集受试者数据时，可能存在成像设备的限制、受试者人员关于隐私的要求、运行成本、受试者纵向数据研究的缺失等因素，并不能充分保证数据集包含满足实验和临床需要的所有模态数据。时代发展的同时，医学影像学也取得了很大的进步，是否可以在患者或是受试者进行影像采集时同时获得独立且具有互补性的多模态数据将是未来的一个挑战。

针对数据处理问题，基于 ROI 的研究需要加强。图像感兴趣目标区域或者对病灶的检测是诊断的关键部

分，加强对 ROI 的研究，可以提高检测准确率或者减少临床医生的读片时间。可以设法实现在全图像空间中定位出 ROI，再在感兴趣区域进一步识别出特定病灶区；而影像在输入网络前，需要进行一系列标准化的预处理，这个过程并未实现完全的自动化。数据预处理能在模型训练时加快收敛速度，且能提高模型性能，但是预处理的过程又太繁琐，同时需要一定的专业背景知识，如何使用未处理的影像数据进行模型训练并能取得良好的分类效果，是未来需要应对的挑战。

在模型的推广和泛化能力等方面，依然存在着问题和挑战。对于已经训练好的模型，能否在其他新出现的患者数据上依然具有好的效果，这是一个需要面对的挑战。可推广性也同样体现在所使用的数据集和实验研究方案的可重复操作性上，尽可能地选择公共的大型数据集，使用更加详尽的数据处理方式和更优秀的分类算法，这对再现实验、模型推广、算法改进、提高科研成效等将产生强有力的推进作用。

上述内容是我们经过五周调研对该项目的初步认识，还不能概括该工作的全貌，在接下来的研究中我们会继续拓展其深度和广度。目前前期调研工作已经告一段落，我们开始在 overleaf 上共同编写文章。尽管我们正在撰写的是一篇综述论文，而不是一篇研究型论文，但我们对于推动多模态在阿尔兹海默症诊断中的应用充满信心和热情。在项目结束后，我们希望我们的成果文章能够帮助科研人员在未来的研究中，从多个角度和多种模态出发来进行全面的研究，寻找最佳的生物标记物组合。并且将不断注重研究方法和模型的选择，来优化模型参数，以构建最优模型，实现计算机辅助诊断技术在阿尔兹海默症研究中的综合意义。我们的目标是阿尔兹海默症的诊断和治疗贡献我们的智慧和努力。