

◆ 医疗场景 ◆

基于深度神经网络的阿尔茨海默病早期诊断算法*

刘星毅

(广西工业职业技术学院, 广西南宁 530001)

摘要: 由于阿尔茨海默病(Alzheimer's Disease, AD)目前无法治愈, 只能通过早期干预的方式缓解其恶化, 因此使用神经影像学技术对阿尔茨海默病进行早期诊断已经在临床中得到广泛应用。本文利用深度学习技术, 对基于磁共振成像(Magnetic Resonance Imaging, MRI)数据的阿尔茨海默病早期诊断进行了研究。首先, 引入样本重加权策略来缓解疾病诊断数据集中普遍存在的类不平衡问题对模型的影响。其次, 设计监督对比学习来提升模型的特征提取能力。在4个神经退行性疾病诊断数据集上的消融实验结果表明, 本文提出的方法取得了比现有算法更好的性能。

关键词: 阿尔茨海默病; 神经退行性疾病; 疾病诊断; 深度学习; 对比学习; 多层感知机

中图分类号: TP183 文献标识码: A 文章编号: 1005-9164(2024)05-0864-09

DOI: 10.13656/j.cnki.gxkx.20241127.005

阿尔茨海默病(Alzheimer's Disease, AD)是一种常见的神经退行性疾病, 俗称老年痴呆^[1,2]。它会导致大脑中神经细胞的退化和死亡, 进而影响记忆、思维和行为的能力, 也是导致老年人死亡的主要原因之一。目前, 尚无治愈AD的方法, 进行早期干预是减缓其恶化的有效途径^[3], 因此对AD的早期诊断极为重要。神经影像学技术如磁共振成像(Magnetic Resonance Imaging, MRI)在AD的早期诊断中发挥着关键作用。由于人工检查神经影像学数据既费力又耗时, 机器学习技术已经被广泛应用于神经影像学数据的处理以减轻神经科医生的负担。

在利用机器学习技术对AD进行早期诊断的起

步阶段, 主要是通过传统的机器学习方法来处理神经影像学数据。例如, 随机森林^[4]通过随机选择特征子集和样本子集构建多个弱分类器(如决策树), 最后集成多个弱分类器的预测(如投票法、平均法等)作为最终分类结果。L1SVM(L1-Support Vector Machine)^[5]利用支持向量机(Support Vector Machine, SVM)完成分类任务, 并引入L1正则化对重要特征进行选择。近年来, 随着机器学习技术的不断发展^[6,7], 深度学习方法已经被广泛应用于疾病诊断领域, 并因其强大的特征提取能力取得了比传统机器学习方法更好的效果。现有的深度学习方法通常把患者作为节点, 并通过K最近邻算法(K-Nearest

投稿日期: 2024-04-15

修回日期: 2024-06-02

*广西自然科学基金项目(2023GXNSFBA026010)资助。

【作者简介】

刘星毅(1972—), 男, 副教授, 主要从事机器学习、数据库技术和多媒体技术研究, E-mail: qzk@163.com。

【引用本文】

刘星毅. 基于深度神经网络的阿尔茨海默病早期诊断算法[J]. 广西科学, 2024, 31(5): 864-872.

LIU X Y. An Algorithm for Early Diagnosis of Alzheimer's Disease Based on Deep Neural Network [J]. Guangxi Sciences, 2024, 31(5): 864-872.

Neighbors, KNN)^[8], 根据节点间的相似性构建图结构数据, 然后使用图卷积网络(Graph Convolutional Network, GCN)^[9]进行特征学习, 最后根据学习到的特征嵌入进行疾病诊断。例如 IDGCN (Interpretable learning based Dynamic Graph Convolutional Network)^[10]在 GCN 的基础上加入可解释特征学习和动态图学习模块, 获得了比传统方法更好的性能, 并且能够解释诊断结果与大脑的特定区域之间的关系。DGLCN (Dual-Graph Learning Convolutional Network)^[11]使用原始特征矩阵及其转置矩阵分别构建目标图和特征图, 并使用它们进行双图学习从而更好地对 MRI 数据进行特征学习。FSNet^[12]使用 FGCN (Feature Graph Convolutional Network) 和 SGCN (Simple Graph Convolutional Network) 分别捕捉特征和样本的重要性, 并根据重要性信息提取特征嵌入, 最后使用特征嵌入进行分类。Multi-Modal GNN (Multi-Modal Graph Neural Network)^[13]分别根据 MRI 和 PET (Positron Emission Tomography) 数据构建图结构, 并通过图融合策略来利用它们之间的互补信息。UNB-GCN (Univariate Neurodegeneration Biomarker Graph Convolutional Network)^[14]

根据 MRI 数据的重要区域生成单变量神经退行性生物标志物 (Univariate Neurodegeneration Biomarker, UNB), 并设计了一个注意力模块使 GCN 更关注产生重要 UNB 的脑部区域。

由于不同疾病患者的数量有差异, 数据集中往往存在类不平衡问题, 现有方法均缺乏对这一问题的考虑。另外, 现有方法特征学习的效果依赖于图结构的质量, 但在神经系统疾病的早期阶段, 患者脑部的病理变化较小, 神经影像学数据与正常人极为相似, 很难构建出高质量的图结构。因此, 本文使用多层感知机 (Multilayer Perceptron, MLP) 对神经影像学数据进行特征学习与疾病诊断, 旨在更有效地使用 MRI 数据对阿尔茨海默病进行早期诊断。

1 方法

本文方法包括标签学习 (蓝色路径) 和对比学习 (红色路径) 两个部分 (图 1)。在标签学习部分, 引入加权交叉熵损失来降低类不平衡问题对模型的影响。在对比学习部分, 设计监督对比学习来提升模型的特征提取能力。

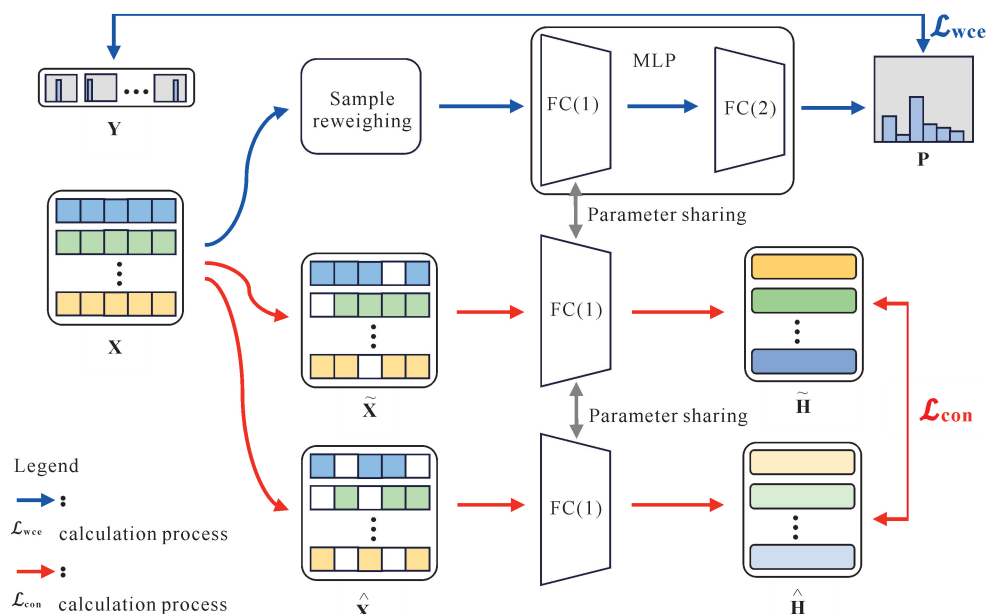


图 1 本文所提出算法的整体框架

Fig. 1 Overall framework of the algorithm proposed in this paper

1.1 模型结构

MLP 是一种经典的前馈人工神经网络模型, 由输入层、隐藏层和输出层组成, 隐藏层和输出层中的每个神经元都与前一层的所有神经元相连接, 因此也

被称为全连接层 (Fully Connected layer, FC)。其中, 隐藏层负责对输入数据进行一系列的非线性变换从而实现特征提取, 输出层负责根据隐藏层提取的特征嵌入对样本类别进行预测。本文采用两层网络结

构的 MLP 作为预测模型, 并将第 1 层输出的特征嵌入 \mathbf{H} 和第 2 层输出的预测概率矩阵 \mathbf{P} 表示如下:

$$\mathbf{H} = \mathbf{W}^{(1)} \mathbf{X} + \mathbf{b}^{(1)}, \quad (1)$$

$$\mathbf{P} = \text{softmax}[\sigma(\mathbf{W}^{(2)} \mathbf{H} + \mathbf{b}^{(2)})], \quad (2)$$

式中, $\mathbf{W}^{(i)}$ 和 $\mathbf{b}^{(i)}$ 表示第 i 层的可训练参数, σ 表示激活函数(如 ReLU、Sigmoid 等)。

1.2 加权交叉熵损失

交叉熵损失是分类任务中常用的损失函数之一^[15,16], 其有效性已经在实践中得到广泛验证。本文将交叉熵损失函数表达如下:

$$\mathcal{L}_{\text{ce}} = - \sum_{i \in T} y_i \log p_i, \quad (3)$$

式中, $T = v_1, v_2, \dots, v_t$, 表示训练数据集; t 表示训练样本总数。 y_i 和 p_i 分别表示样本 v_i 的标签和模型对样本 v_i 的预测概率分布。交叉熵损失可以衡量模型输出的概率分布与真实标签之间的差异, 当两者相差较大时, 交叉熵损失的值会较大, 反之则较小。因此, 通过最小化交叉熵损失, 可以使模型获得对不同类别的区分能力。然而, 当数据集中不同类别的样本数量差距较大时, 使用交叉熵损失可能会导致模型在训练过程中更关注数量较多的类别而忽略了数量较少的类别, 原因是传统的交叉熵损失会公平地看待所有样本。

为解决上述问题, 对式(3)进行了一些改进, 使其能够为不同类别样本给予不同程度的关注度。具体来说, 引入样本重加权策略, 使样本数量较少的类别获得更高的权重, 而样本数量较多的类别则相反。为此, 先计算训练集中每个类别样本的数量:

$$M_i = \sum_{j=1}^t \mathbf{1}(y_j = i), \quad (4)$$

式中, M_i 表示类别 i 的样本数量; $\mathbf{1}(\ast)$ 表示指示函数, 当括号内的条件成立时, $\mathbf{1}(\ast) = 1$, 否则 $\mathbf{1}(\ast) = 0$ 。计算每个类别的样本数量后, 根据式(5)计算每个类别的样本在训练中的权重:

$$\omega_i = 1 - \frac{M_{y_i}}{t}, \quad (5)$$

式中, ω_i 表示样本 v_i 的权重, M_{y_i} 表示与节点 v_i 类别相同的样本数量。根据式(5), 样本的权重将与其同类样本的数量占总样本数量的比例成反比。最终, 结合式(3)和式(5), 得到加权交叉熵损失 \mathcal{L}_{wce} :

$$\mathcal{L}_{\text{wce}} = - \sum_{i \in T} \omega_i \cdot (y_i \log p_i). \quad (6)$$

式(6)通过为不同类别的样本赋予不同的权重, 使得模型在训练中公平地对每个类别进行学习, 从而降低类不平衡问题对模型训练的影响, 使模型获得更好的

泛化能力。

1.3 监督对比损失

加权交叉熵损失能够通过最小化模型的预测结果与真实标签之间的差异来使模型从样本标签中学习类别信息。然而, 它缺乏对样本间相互关系的学习能力, 而学习样本间相互关系往往能够帮助模型更好地捕捉数据的内在结构和关联性, 从而获得更高质量的特征嵌入。基于上述思考, 本文设计监督对比损失来使模型能够更好地学习样本之间的相互关系。具体来说, 为每个样本选择 k 个正样本和 k 个负样本, 并使其与正样本在嵌入空间中互相靠近, 与负样本在嵌入空间中互相远离。

1.3.1 选择正样本

为了使模型能够学习到同类样本间的联系, 为每个样本选择与其类别相同且特征最为相似的部分样本作为其正样本, 并在嵌入空间中使它们互相靠近。具体来说, 先将训练集按照不同类别分为多个子集:

$$C_i = \{v_j \mid \forall j \in (1, n), y_j = i\}, \quad (7)$$

式中, C_i 表示由类别为 i 的样本组成的集合。接下来, 计算每个样本与其余同类样本的特征相似度:

$$P_i = \{S_{ij} \mid \forall v_j \in C_{y_i}\}, \text{ where } S_{ij} = \frac{x_i \cdot x_j}{|x_i| |x_j|}, \quad (8)$$

式中, P_i 表示由样本 v_i 与其余同类样本的相似度组成的集合, S_{ij} 表示节点 v_i 与 v_j 的相似度, x_i 和 x_j 分别表示节点 v_i 与 v_j 的特征向量。最后, 为每个样本选择 k 个相似度最高的同类样本作为其正样本:

$$\text{Pos}_i = \text{top}(P_i, k), \quad (9)$$

式中, Pos_i 表示为样本 v_i 选择的正样本组成的集合, $\text{top}(P_i, k)$ 表示从 P_i 中选取前 k 个具有最大值的样本。通过减小样本与其正样本在嵌入空间中的距离, 模型能够更好地学习到同类样本间的联系, 从而提升模型的特征提取能力。

1.3.2 选择负样本

在让模型学习同类样本间联系的同时, 也需要学习不同类别样本间的差异。因此, 本文将为每个样本选择与正样本数量相同的负样本。首先, 计算每个样本与非同类样本的特征相似度:

$$N_i = \{S_{ij} \mid \forall v_j \notin C_{y_i}\}, \quad (10)$$

式中, N_i 表示由样本 v_i 与非同类样本的相似度组成的集合。接下来, 为了使模型在提升对特征相似的非同类样本区分能力的同时, 保持对特征不相似样本的区分能力。为每个样本选择 $k-1$ 个相似度最高的

非同类样本以及一个相似度最低的非同类样本作为负样本:

$$\text{Neg}_i = \text{top}(N_i, k-1) + \text{top}(N_i, -1), \quad (11)$$

式中, Neg_i 表示为样本 v_i 选择的负样本组成的集合, $\text{top}(N_i, k-1)$ 表示从 N_i 中选取前 $k-1$ 个相似度最高的样本, $\text{top}(N_i, -1)$ 表示从 N_i 中选取相似度最低的样本。通过增加样本与其负样本在嵌入空间中的距离, 模型能够更好地捕捉到不同类别样本间的差异, 使不同类别的样本在嵌入空间中具有更好的分离度, 进而提升模型的分类性能。

1.3.3 对比损失

加权交叉熵损失会使模型倾向于将同一类别的所有样本具有相似的特征嵌入, 即在嵌入空间中属于同一个簇, 而对比损失则可能使模型将同一类别的样本映射为多个簇。为了避免两者的优化目标出现冲突, 更好地发挥各自的优势, 本文考虑通过数据增强获得两组增强特征来计算对比损失, 从而使模型间接地对样本间的相互关系进行学习。首先, 通过随机掩码策略获得两组增强特征:

$$\begin{cases} \tilde{\mathbf{X}} = \text{mask}(\mathbf{X}, r_1), \\ \hat{\mathbf{X}} = \text{mask}(\mathbf{X}, r_2), \end{cases} \quad (12)$$

式中, $\text{mask}(\mathbf{X}, r)$ 表示将矩阵 \mathbf{X} 每一行中 $r\%$ 的元素值转变为 0。接下来, 分别将两组增强特征输入模型, 获得模型第一层输出的特征嵌入:

$$\begin{cases} \tilde{\mathbf{H}} = \mathbf{W}^{(1)} \tilde{\mathbf{X}} + \mathbf{b}^{(1)}, \\ \hat{\mathbf{H}} = \mathbf{W}^{(1)} \hat{\mathbf{X}} + \mathbf{b}^{(1)}, \end{cases} \quad (13)$$

式中, $\tilde{\mathbf{H}}$ 和 $\hat{\mathbf{H}}$ 分别表示预测模型对 $\tilde{\mathbf{X}}$ 和 $\hat{\mathbf{X}}$ 提取的特征嵌入。在获得增强特征的特征嵌入后, 使用余弦相似度来衡量样本在嵌入空间中的距离, 并将样本对在嵌入空间中的距离表示如下:

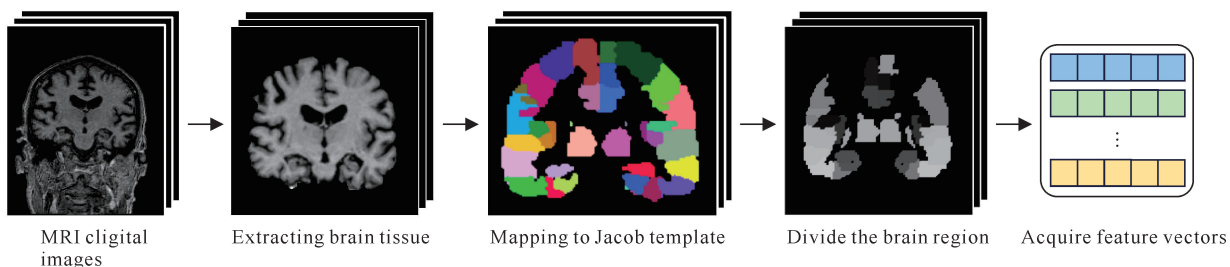


图2 从MRI数字图像提取特征向量的预处理流程

Fig. 2 Preprocessing pipeline of extracting feature vectors from MRI digital images

共处理 805 位受试者的 MRI 数字图像, 其中包括 186 名 AD 患者, 393 名轻度认知障碍(MCI)患者

$$d(h_i, h_j) = \frac{\tilde{\mathbf{h}}_i \cdot \tilde{\mathbf{h}}_j}{|\tilde{\mathbf{h}}_i| |\tilde{\mathbf{h}}_j|} + \frac{\tilde{\mathbf{h}}_i \cdot \hat{\mathbf{h}}_j}{|\tilde{\mathbf{h}}_i| |\hat{\mathbf{h}}_j|} + \frac{\hat{\mathbf{h}}_i \cdot \hat{\mathbf{h}}_j}{|\hat{\mathbf{h}}_i| |\hat{\mathbf{h}}_j|}, \quad (14)$$

式中, $\tilde{\mathbf{h}}_i$ 和 $\hat{\mathbf{h}}_i$ 分别表示将 $\tilde{\mathbf{H}}$ 和 $\hat{\mathbf{H}}$ 作为输入时模型对样本 v_i 输出的特征嵌入。最终, 将所使用的对比损失表达如下:

$$\mathcal{L}_{\text{con}} = \sum_{i \in T} [\sum_{v_j \in \text{Pos}_i} d(h_i, h_j) - \sum_{v_j \in \text{Neg}_i} d(h_i, h_j)]. \quad (15)$$

通过最小化 \mathcal{L}_{con} , 每个样本与正样本在嵌入空间中会互相靠近, 与负样本在嵌入空间中会互相远离, 从而学习到样本间的相互关系, 使得不同类别样本的特征嵌入具有更好的区分度。此外, 由于本文使用增强特征来计算对比损失, 因此能够使模型获得更强的鲁棒性, 提升模型的泛化能力。

1.4 总损失函数

将加权交叉熵损失 \mathcal{L}_{wce} 和对比损失 \mathcal{L}_{con} 进行组合, 得到本文所提出方法的总损失函数:

$$\mathcal{L} = \mathcal{L}_{\text{wce}} + \alpha \mathcal{L}_{\text{con}}, \quad (16)$$

式中, α 表示平衡 \mathcal{L}_{wce} 和 \mathcal{L}_{con} 的加权系数。

2 实验与结果分析

2.1 数据集

从 ADNI 数据库下载原始 MRI 数字图像, 并遵循文献[17]中描述的方式进行预处理, 具体步骤包括提取脑组织、修正运动和时间、配准、过滤和平滑。在完成对 MRI 数字图像的预处理后, 将其映射到 Jacob 模板^[18]中, 从而将大脑划分为 93 个脑区。最后, 提取每个脑区的灰质体积作为特征, 这样每个样本(患者)就可以用一个 93 维的特征向量来表示。数据预处理的具体流程见图 2。

和 226 名正常人(对照组, NC)。此外, 393 名 MCI 患者又分为 226 名 MCI 非转化者(MCI_n)和 167 名

MCI 转化者(MCIc)。最后,将这些样本分成 4 个二分类数据集(即 AD-NC、NC-MCI、AD-MCI 和 MCIIn-MCIc)进行实验。4 个数据集的具体信息见表 1。

表 1 数据集信息

Table 1 Dataset information

数据集 Dataset	样本 Samples	特征维度 Characteristic dimensions
AD-NC	186 : 226	93
NC-MCI	226 : 393	93
AD-MCI	186 : 393	93
MCIIn-MCIc	226 : 167	93

2.2 对比算法

本文使用的对比算法包括 1 种传统机器学习算法(L1SVM)和 5 种深度学习算法(GCN、GAT、Transformer、SUGRL、DGLCN)。

L1SVM^[5]在 SVM 的基础上引入 L1 正则化来过滤掉对分类决策贡献较小的特征。

GCN^[9]是一种基于频谱理论的图卷积网络,通过图结构来对节点特征进行邻域聚合的方式学习节点间的相互关系。

GAT(Graph Attention Network)^[19]通过注意力机制来计算节点与其邻居之间的重要性权重,并利用这些权重对节点的邻域信息进行加权聚合。

Transformer^[20]是一种基于自注意力机制的深度学习模型,通过计算每个特征向量中不同位置之间的相互依赖关系,为每个位置分配注意力权重。

SUGRL(Simple Unsupervised Graph Representation Learning)^[21]使用结构信息和领域信息来为节点选择正样本,并通过上界损失限制节点与正样本在特征空间中的距离。

DGLCN^[11]为样本构建目标图和特征图,并通过双图来学习样本的目标相关性和特征相关性。

2.3 实验设置

在一台 4 张 NVIDIA GeForce RTX 2080 Ti (每张 11 GB 内存)的服务器上使用 PyTorch Geometric (PyG)框架进行实验。参考 2.2 节中对比算法的相关文献[5][9][11][19-21]来为其设置最优参数,使其达到最佳结果。由于本实验使用的数据集没有提供预先建立的图结构,因此对于所有需要使用图结构的算法,本文使用 K 最近邻算法为它们构建图结构,K 值统一设置为 10。对于所有对比算法,设置最

大训练轮次为 10^4 轮,预测模型的隐藏层神经元数量为 32,优化器为 Adam (Adaptive moment estimation)。GCN、GAT 以及本文提出方法的学习率均为 10^{-2} 。SUGRL 的学习率为 2.5×10^{-3} ,权重衰减为 5×10^{-5} 。Transformer 的学习率为 2×10^{-4} ,权重衰减为 10^{-3} ,隐藏层神经元数量为 256,注意力头数量和编码器的层数均为 4。DGLCN 的学习率为 5×10^{-3} ,特征提取器为两层 GCN,神经元数量分别为 128、64。

通过 4 种常用指标对所有方法进行评估,包括准确率(ACC)、敏感度(SEN)、特异度(SPE)和 ROC 曲线下面积(AUC)。对于这些评价指标,更高的数值都意味着更好的性能。为了减少实验结果的随机性,采用 5 倍交叉验证,并使用不同的随机数种子在 4 个数据集上对所有方法独立进行了 5 次交叉验证实验,最后取 5 个结果的平均值作为最终结果。

2.4 结果分析

本文提出的方法与 6 种对比算法的平均值和相应的标准差见表 2。

深度学习算法在这 4 个数据集上的表现明显优于传统机器学习算法。例如,与传统机器学习算法 L1SVM 相比,深度学习算法中整体表现最差的 GAT 在所有数据集的 4 个评价指标上平均提高了 1.1%,原因是深度学习算法具有比传统机器学习算法更强大的特征提取能力,能够更好地理解受试者的 MRI 数据,从而更准确地对神经系统疾病进行诊断。

从整体上看,不使用图卷积神经网络算法的表现优于使用图卷积神经网络算法。例如,与使用图卷积神经网络算法相比,不使用图卷积神经网络算法在所有数据集的 4 个评价指标上平均提高了 2.06%,原因是神经系统疾病数据集中样本特征差异较小,难以构建高质量的图结构,导致图卷积神经网络难以充分发挥其特征学习能力。

在所有深度学习算法中,本文提出的方法在整体上表现出最好的性能,然后依次是 DGLCN、Transformer、SUGRL、GCN、GAT。例如,与 DGLCN 相比,本文提出的方法在所有数据集的 4 个评价指标上平均提高了 1.4%。原因是:①通过样本重加权策略有效降低了类不平衡问题对模型的影响,有效提升了模型的泛化能力。②本文提出的方法通过监督对比损失使模型能够更好地学习到样本间的相互关系,提升了模型提取的特征嵌入质量,进而提高了模型的性能。

表 2 分类实验结果[平均值(%) and 标准差]

Table 2 Results of the classification experiment [average value (%) and standard deviation]

数据集 Dataset	指标 Index	L1SVM	GCN	GAT	Transformer	SUGRL	DGLCN	Proposed
AD-NC	ACC	76.3 (2.1)	81.5 (1.0)	81.3 (1.3)	84.0 (1.4)	81.6 (0.4)	86.9 (1.6)	85.8 (0.7)
	SEN	77.5 (1.2)	84.7 (0.7)	85.1 (2.1)	<u>88.1 (2.7)</u>	84.9 (1.5)	85.9 (1.6)	88.8 (1.1)
	SPE	75.1 (2.4)	77.5 (1.4)	76.2 (2.0)	78.9 (1.2)	81.1 (1.1)	87.2 (1.7)	82.1 (1.1)
	AUC	77.6 (1.8)	88.3 (0.6)	87.0 (1.2)	78.4 (3.0)	88.1 (0.5)	<u>90.6 (2.5)</u>	91.3 (0.6)
AD-MCI	ACC	65.3 (1.8)	66.1 (1.4)	65.1 (0.8)	67.7 (0.6)	66.5 (0.4)	<u>68.7 (1.5)</u>	69.9 (1.2)
	SEN	66.4 (1.6)	<u>73.8 (2.3)</u>	73.5 (1.7)	66.0 (1.1)	66.8 (0.9)	68.5 (1.6)	78.0 (2.2)
	SPE	64.2 (3.4)	51.0 (6.7)	47.3 (3.6)	70.5 (2.3)	<u>72.0 (1.0)</u>	74.8 (2.4)	53.0 (2.4)
	AUC	61.4 (2.2)	69.2 (2.1)	67.3 (0.6)	70.9 (0.8)	<u>71.6 (0.6)</u>	65.5 (2.8)	71.8 (1.1)
NC-MCI	ACC	61.2 (1.4)	64.2 (1.6)	61.8 (1.6)	<u>67.7 (0.6)</u>	63.5 (1.3)	67.5 (2.4)	69.1 (1.0)
	SEN	56.8 (1.6)	64.6 (1.2)	64.9 (3.2)	<u>66.0 (1.1)</u>	65.6 (2.1)	61.7 (1.5)	67.2 (1.1)
	SPE	63.5 (2.7)	63.9 (3.4)	56.5 (2.9)	<u>70.5 (2.4)</u>	68.2 (1.7)	67.7 (1.3)	72.4 (1.2)
	AUC	63.5 (2.0)	70.1 (1.3)	66.3 (1.7)	<u>70.9 (0.8)</u>	68.4 (0.5)	69.2 (2.8)	74.8 (0.3)
MCIc-MCI _n	ACC	62.4 (2.8)	65.5 (2.2)	62.0 (2.3)	64.7 (3.6)	65.0 (0.9)	<u>67.7 (2.4)</u>	68.8 (0.7)
	SEN	66.4 (3.1)	<u>70.1 (3.2)</u>	67.9 (3.3)	67.9 (4.6)	67.2 (1.2)	64.0 (2.5)	72.0 (2.4)
	SPE	62.4 (2.3)	58.9 (0.5)	53.6 (2.9)	59.3 (6.9)	<u>66.6 (1.7)</u>	69.9 (3.1)	64.4 (1.8)
	AUC	64.4 (1.6)	<u>69.8 (2.2)</u>	65.8 (1.7)	66.0 (3.8)	68.9 (1.1)	63.7 (3.0)	72.3 (0.8)

Note: data in bold and underlined are the best and second best results, respectively.

2.5 消融实验

本文提出的方法包括两个重要部分,即样本重加权(a)和监督对比学习(b)。为了探究本文提出的方法中各个部分的有效性,在所有数据集上对不同部分组合的性能进行消融实验,结果见表3。使用样本重加权策略与不使用时相比,在所有数据集的4个评价指标上平均提高了3.2%,这是因为样本重加权策略通过重新为不同类别的样本分配权重,使模型能够更

加公平地对每个类别进行学习,获得更好的泛化性能。使用监督对比学习与不使用时相比,在所有数据集的4个评价指标上平均提高了6.8%,原因是监督对比学习能够使模型学习到样本间的相互联系,提升模型的特征提取能力,从而进一步提升模型分类性能。因此,同时考虑样本重加权策略和监督对比学习是合理且有效的。

表 3 消融实验一结果[平均值(%) and 标准差]

Table 3 Results of the ablation experiment 1 [average value (%) and standard deviation]

数据集 Dataset	a	b	准确率 ACC	敏感度 SEN	特异度 SPE	曲线下面积 AUC
AD-NC	✓		85.2 (0.8)	88.6 (1.4)	81.2 (0.4)	91.1 (0.6)
		✓	83.0 (0.8)	85.1 (1.5)	80.4 (0.7)	89.0 (0.7)
	✓	✓	85.8 (0.7)	88.8 (1.1)	82.1 (1.1)	91.3 (0.6)
AD-MCI	✓		70.2 (1.0)	83.0 (1.8)	43.5 (1.2)	71.6 (1.4)
		✓	68.8 (1.2)	76.7 (1.6)	51.9 (2.5)	68.7 (1.2)
	✓	✓	69.9 (1.2)	78.0 (2.2)	53.0 (2.4)	71.8 (1.1)
NC-MCI	✓		67.7 (0.8)	80.4 (0.8)	46.5 (2.7)	74.5 (0.3)
		✓	65.9 (0.7)	70.9 (1.6)	57.6 (1.8)	72.0 (0.6)
	✓	✓	69.1 (1.0)	67.2 (1.1)	72.4 (1.2)	74.8 (0.3)

续表

Continued table

数据集 Dataset	a	b	准确率 ACC	敏感度 SEN	特异度 SPE	曲线下面积 AUC
MCIIn-MCIc	✓		67.5 (1.3)	72.7 (1.1)	60.7 (3.2)	72.0 (1.0)
		✓	63.4 (1.3)	67.3 (1.7)	58.3 (3.1)	68.2 (1.7)
	✓	✓	68.8 (0.7)	72.0 (2.4)	64.4 (1.8)	72.3 (0.8)

Note, data in bold are the best results.

为进一步探究本文提出的方法在计算监督对比损失时不使用原始特征做法的合理性,分别使用以下4种组合计算监督对比损失进行消融实验:(I) \mathbf{H} , $\hat{\mathbf{H}}$, \mathbf{H} ; (II) \mathbf{H} , $\hat{\mathbf{H}}$; (III) \mathbf{H} , $\hat{\mathbf{H}}$; (IV) $\hat{\mathbf{H}}$, \mathbf{H} , 实验结果如表4所示。相比于使用原始特征(III)时最好的情况,不使用原始特征(IV)时在所有数据集上的分类准确率平均提高了1.8%,原因是当使用原始特征计算监督

对比损失时,监督对比损失与监督交叉熵损失的优化目标可能会发生冲突,即监督对比损失允许模型将同一类别的样本在嵌入空间中映射为多个簇,而加权交叉熵损失则希望模型将同一类别的样本在嵌入空间中映射为一个簇。因此,在计算对比损失时不考虑原始特征的做法是合理的。

Table 4 Results of the ablation experiment 2 [average value (%) and standard deviation]

数据集 Dataset	设置 Setting	准确率 ACC	敏感度 SEN	特异度 SPE	曲线下面积 AUC
AD-NC	I	84.7 (0.8)	87.8 (1.6)	80.8 (1.1)	91.0 (0.7)
	II	84.7 (1.0)	87.7 (1.6)	81.0 (0.7)	90.5 (0.4)
	III	85.3 (1.2)	88.5 (1.1)	81.3 (1.5)	91.1 (0.6)
	IV	85.8 (0.7)	88.8 (1.1)	82.1 (1.1)	91.3 (0.6)
AD-MCI	I	69.1 (0.9)	79.8 (3.1)	47.1 (4.4)	70.8 (0.8)
	II	68.8 (1.3)	76.6 (2.3)	52.2 (2.0)	71.2 (1.1)
	III	69.4 (0.8)	78.1 (1.7)	51.3 (2.6)	71.1 (0.8)
	IV	69.9 (1.2)	78.0 (2.2)	53.0 (2.4)	71.8 (1.1)
NC-MCI	I	67.5 (1.0)	63.1 (1.4)	75.2 (1.5)	74.3 (0.3)
	II	67.3 (0.4)	63.6 (1.4)	73.7 (2.1)	74.0 (0.6)
	III	68.0 (0.8)	64.9 (0.7)	73.6 (2.1)	74.3 (0.3)
	IV	69.1 (1.0)	67.2 (1.1)	72.4 (1.2)	74.8 (0.3)
MCIIn-MCIc	I	66.6 (2.0)	70.3 (1.0)	61.9 (4.3)	71.9 (0.6)
	II	67.1 (0.8)	70.2 (1.4)	63.0 (1.4)	72.1 (0.8)
	III	66.6 (1.8)	69.9 (0.8)	62.2 (3.6)	71.8 (1.2)
	IV	68.8 (0.7)	72.0 (2.4)	64.4 (1.8)	72.3 (0.8)

Note, data in bold are the best results.

2.6 参数敏感度实验

本文提出的方法包括2个关键超参数,即用于确定选择正负样本数量的 k 和用于调整监督对比损失在总损失中权重的 α 。下面分别对这两个参数进行参数敏感度实验。

在 $\{10, 20, \dots, 50\}$ 范围内改变 k 的取值,结果见图3。显然,本文提出的方法对 k 的取值并不敏感,

原因是本文提出的监督对比学习在选择正负样本时考虑了类别因素, k 的增大不会导致将非同类样本选为正样本或将同类样本选为负样本,因此依然能够使模型合理地学习样本间的相互关系。

在 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 范围内改变 α 的取值,实验结果如图4所示。本文提出的方法对 α 的取值较为敏感。具体来说,本文提出的方法在 α 的取值为

10^1 时达到最佳性能。原因是当 α 的取值过小时, 监督对比损失在总损失中的占比过小, 模型无法学习到样本间的相互关系导致分类效果不佳。而当 α 的取值过大时, 加权交叉熵损失在总损失中的占比过小, 模型难以从标签中学习对对不同类别的区分能力, 同样会导致分类效果不佳。因此, 选择合适的 α 对本文提出的方法至关重要。

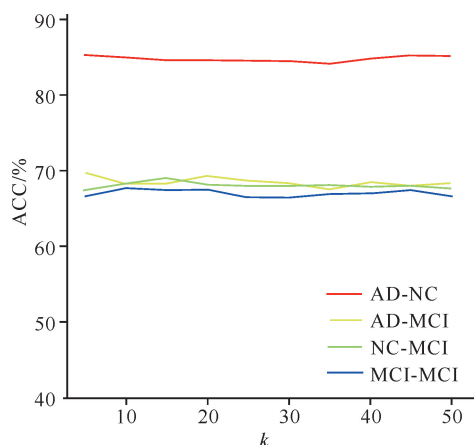


图3 参数敏感度实验结果(k)

Fig. 3 Results of the parameter sensitivity test (k)

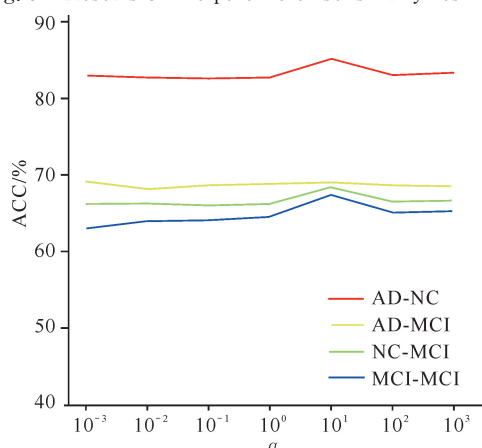


图4 参数敏感度实验结果(α)

Fig. 4 Results of the parameter sensitivity test (α)

3 结论

本文提出了一种新的基于深度神经网络的阿尔茨海默病早期诊断算法。首先在交叉熵函数中引入样本重加权策略, 以降低疾病诊断数据集中的类不平衡问题对模型造成的负面影响; 其次设计监督对比学习使模型学习到样本间的相互关系, 从而提高模型的特征提取能力。实验结果表明, 与现有最先进的方法相比, 本文提出的方法具有更强的分类能力。

参考文献

[1] 丁秀梅, 杜晓歌, 郭艳芳. 老年痴呆症患者精神行为症状

现状调查及影响因素分析[J]. 临床医学, 2023, 43(12): 53-55.

[2] 袁天漫, 徐菊玲, 钱秀群, 等. 老年痴呆病人认知功能受损的研究进展[J]. 全科护理, 2021, 19(15): 2050-2053.

[3] 刘建英, 于颢荔, 何明鸣. 早期康复训练联合健康教育对老年痴呆患者认知功能、生活质量和肢体功能的影响[J]. 中国老年学杂志, 2022, 42(7): 1703-1705.

[4] SVETNIK V, LIAW A, TONG C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling [J]. Journal of Chemical Information and Computer Sciences, 2003, 43(6): 1947-1958.

[5] ZHU J, ROSSET S, TIBSHIRANI R, et al. 1-norm support vector machines [J]. Advances in Neural Information Processing Systems, 2003: 15379424.

[6] 洪月华, 徐霜, 梁家荣. 一种基于粗糙集和遗传神经网络的数据分类器模型[J]. 广西科学, 2013, 20(2): 128-131, 136.

[7] 吴建生. 旱涝灾害的遗传-神经网络集成预测方法研究[J]. 广西科学, 2006, 13(3): 203-206, 211.

[8] COVER T, HART P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.

[9] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [Z/OL]. (2017-12-22) [2024-06-02]. <https://arxiv.org/abs/1609.02907v4>.

[10] ZHU Y H, MA J B, YUAN C G, et al. Interpretable learning based dynamic graph convolutional networks for Alzheimer's disease analysis [J]. Information Fusion, 2022, 77: 53-61.

[11] XIAO T S, ZENG L, SHI X S, et al. Dual-graph learning convolutional networks for Interpretable Alzheimer's disease diagnosis [C]//Medical Image Computing and Computer Assisted Intervention: MICCAI 2022. Cham, Switzerland: Springer Nature, 2022: 406-415.

[12] LI H X, SHI X S, ZHU X F, et al. FSNet: dual interpretable graph convolutional network for Alzheimer's disease analysis [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(1): 15-25.

[13] ZHANG Y, HE X, CHAN Y H, et al. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans [J]. Computers in Biology and Medicine, 2023, 164: 107328.

[14] QU Z S, YAO T, LIU X H, et al. A graph convolution-

- al network based on univariate neurodegeneration biomarker for Alzheimer's disease diagnosis [J]. IEEE Journal of Translational Engineering in Health and Medicine, 2023, 11: 405-416.
- [15] 文竹, 袁立宁, 黄伟, 等. 基于图多层感知机的节点分类算法[J]. 广西科学, 2023, 30(5): 942-950.
- [16] 李华, 陈宁江. 基于 PSO 的 WFCM 算法研究及其在医保欺诈行为发现中的应用[J]. 广西科学院学报, 2017, 33(1): 32-39.
- [17] SHEN H T, ZHU X, ZHANG Z, et al. Heterogeneous data fusion for predicting mild cognitive impairment conversion [J]. Information Fusion, 2021, 66: 54-63.
- [18] KABANI N J, MACDONALD D J, HOLMES C J, et al. 3D anatomical atlas of the human brain [J]. NeuroImage, 1998, 7(4): S717.
- [19] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [Z/OL]. (2018-12-04) [2024-06-02]. <https://arxiv.org/abs/1710.10903v3>.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [Z/OL]. (2017-01-12) [2024-06-02]. <https://arxiv.org/abs/1706.03762>.
- [21] MO Y J, PENG L, XU J, et al. Simple unsupervised graph representation learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI, 2022: 7797-7805.

An Algorithm for Early Diagnosis of Alzheimer's Disease Based on Deep Neural Network

LIU Xingyi

(Guangxi Vocational & Technical Institute of Industry, Nanning, Guangxi, 530001, China)

Abstract: Alzheimer's Disease (AD) is currently incurable and its deterioration can only be mitigated by early intervention. Therefore, early diagnosis of Alzheimer's disease by neuroimaging has been widely used in clinical practice. A deep learning network model is built based on the Magnetic Resonance Imaging (MRI) data for the early diagnosis of Alzheimer's disease. First, a sample reweighting strategy is introduced to mitigate the impact of the class imbalance problem prevalent in disease diagnosis datasets on the model. Second, supervised contrastive learning is designed to improve the feature extraction capability of the model. The experimental results on four neurodegenerative disease diagnosis datasets show that the method proposed in this paper achieves better performance than existing methods.

Key words: Alzheimer's disease; neurodegenerative disease; disease diagnosis; deep learning; contrastive learning; multilayer perceptron

责任编辑: 陆媛峰



微信公众号投稿更便捷

联系电话: 0771-2503923

邮箱: gxkx@gxas.cn

投稿系统网址: <http://gxkx.ijournal.cn/gxkx/ch>