arXiv:2305.01666v4 [q-bio.NC] 2 Aug 2023

BrainNPT：用于脑功能网络分类的 Transformer 网络预训练

Jinlong Hu[1*] **YangminHuang**[1] **NanWang**[2] **Shoubin Dong**[1] 广东省通信与计算机网络重点实验室，计算机科学与工程学院，华南理工大学，中国广州[2] 计算机科学与技术学院，华东师范大学，中国上海

摘要

过去几年，深度学习方法在脑影像分析领域取得了迅速进展，但通常受限于标注数据不足。在无标签数据上训练的预训练模型已在许多领域（包括自然语言处理和计算机视觉）展现出在特征学习方面的显著提升。然而，这项技术在脑网络分析中的应用尚未得到充分探索。本文聚焦于利用 Transformer 网络进行预训练的方法，以利用现有的无标注数据完成脑功能网络分类。首先，我们提出了一种名为 BrainNPT 的基于Transformer的神经网络，用于脑功能网络分类。所提出的方法将 <cls> 标记用作 Transformer 模型的分类嵌入向量，从而有效捕获脑网络表示。其次，我们提出了一个 BrainNPT 模型的预训练框架，以利用无标注脑网络数据学习脑网络的结构信息。分类实验结果表明，未经预训练的 BrainNPT 模型达到了与最先进模型相当的最佳性能，而经过预训练的 BrainNPT 模型则显著优于最先进模型。与未经预训练的模型相比，预训练后的 BrainNPT 模型准确率提高了 8.75%。此外，我们还比较了不同的预训练策略，分析了模型参数的影响，并对训练后的模型进行了阐释。

关键词 脑功能网络 · Transformer · 预训练 · 分类

1 引言

过去几年，深度学习方法在脑影像分析领域取得了快速进展 [1, 2]。为了表征各种脑部疾病中的脑功能连接及其变化，越来越多的图深度学习模型被开发用于脑功能网络分析，例如采用专门设计的卷积层的图神经网络（GNNs） [3], 图注意力网络 [4],结合信息最大化正则化的图同构网络（GIN） [5], GroupINN [6], IBGNN [7],和 FBNetGEN [8]。

近年来，Transformer [9] 在自然语言处理（NLP）领域展现出了卓越的性能，例如 BERT [10] 和 GPT [11], , 它们凭借多头注意力机制以及捕获长程依赖的能力取得了出色的表现。一些研究也尝试将 Transformer 结构应用于图数据领域。它们通常针对特定图设计位置嵌入策略，将位置信息融入节点嵌入，以进行自注意力计算。例如，Graph-Bert [12] 提出了三种基于子图结构的位置编码策略，包括 Weisfeiler-Lehman 绝对角色嵌入、基于亲密度的相对位置嵌入，以及基于跳数的相对距离嵌入。GROVER [13] 利用图神经网络（GNNs）代替多层感知机（MLP），获取用于自注意力计算的查询、键和值的表征向量。GraphiT [14] 将图神经网络（GNNs）用作相对位置嵌入的核函数，以进一步计算自注意力分数。Graphormer [15]提出了

*通讯作者。电子邮箱：jlhu@scut.edu.cn

三种专为分子图结构设计的位置编码策略。结构感知 Transformer (SAT) [16] 将子图结构用作位置嵌入。GraphTrans [17] 利用图神经网络引入位置和局部结构信息。GraphGPS [18] 采用消息传递图神经网络 (MPNNs) 与 Transformer 的交替组合结构，以增强位置信息提取。近期，BrainNetTF [19] 将Transformer结构与脑网络邻接矩阵的属性相结合，在脑网络分类任务中取得了最先进的性能。

然而，深度学习方法通常受到标注数据有限的制约。由于采集静息态功能磁共振成像 (rs-fMRI) 数据通常耗时较长，且需要借助专业设备开展繁琐的实验，而某些脑部疾病较为罕见，与之相关的静息态功能磁共振成像数据规模则更加有限。因此，基于静息态功能磁共振成像数据构建的脑功能网络数据通常较为稀缺。此外，与传统卷积方法相比，数据不足的情况下，Transformer 可能会导致较差的性能 [20, 21]。预训练是一种解决标注数据有限问题的有效技术。例如，一种基于图神经网络的预训练方法已被应用于利用脑成像数据对脑网络进行分类，且无需相关任务标签 [22]。此外，基于 Transformer 的预训练方法已在自然语言处理 [10, 11] 和计算机视觉 (CV) [23, 24],领域取得了巨大成功，但这一预训练技术尚未应用于脑功能网络分析。

本文重点研究了利用 Transformer 网络进行预训练的方法，以充分利用现有的无标签数据进行脑功能网络分类。首先，我们提出了一种新型的基于Transformer的神经网络，命名为 BrainNPT，用于脑功能网络分类，其中设计了一个可学习的 <cls> token，作为分类嵌入向量，以捕获脑网络表示。其次，我们提出了一种利用大规模无标签脑功能网络数据对 BrainNPT 模型进行预训练的框架，并提出了替换感兴趣脑区 (ROI) 预测作为预训练策略。我们使用来自自闭症脑成像数据交换项目 (ABIDE) II [25]和 REST-meta-MDD [26] 的两个公开数据集对 BrainNPT 模型进行了评估。实验结果表明，未经预训练的基础 BrainNPT 模型取得了与当前最先进模型相当的最佳性能，而经过预训练的 BrainNPT 模型则优于当前最先进的模型。此外，我们进一步比较了不同的预训练策略，分析了模型参数的影响，并对微调模型进行了可解释性分析。

主要贡献总结如下：

(i) 我们提出了一种带有 <cls> 标记的基于Transformer的神经网络，用于脑功能网络分类。该模型在概念上简单而有效，能够对脑网络进行分类，并且可以按照 Transformer 风格轻松进行预训练和深度堆叠。此外，所提出的模型可以通过逐层相关性传播 (LRP) 进行解释，这是深度神经网络最突出的解释技术之一。

(ii) 我们提出了一种具有有效预训练策略的预训练框架，用于检测脑感兴趣区是否被替换。这些预训练策略使模型能够从无标注脑网络数据中捕获脑功能网络的内在结构。与未经预训练的基础模型相比，预训练模型的准确率提高了 8.75%。

(iii) 我们通过对静息态功能磁共振成像数据随时间应用滑动窗口，对脑功能网络数据进行增强，并使用大量此类脑快照网络数据进行预训练。鉴于基于Transformer的模型具有强大的能力，即使是未标注的脑功能网络数据也十分稀缺。实验结果表明，使用滑动窗口进行数据增强对模型的预训练非常有帮助。

2 方法

2.1 定义

本文将脑功能网络视为图，其中每个感兴趣区域都视为图中的一个节点，而感兴趣区域之间的功能连接值则视为节点之间的边权重。

基于预处理后的静息态功能磁共振成像数据，依据脑图谱为每位受试者获取 N 个感兴趣区域；通过对每个感兴趣区域内所有体素的值求平均，计算该感兴趣区域的时间序列；对各感兴趣区域的时间序列进行皮尔逊相关运算，得到功能连接；并为每位受试者构建形状为 $N \times N$ 的功能连接矩阵。将功能连接矩阵视为脑功能网络 $X \in \mathbb{R}^{N \times N}$，并对该脑网络进行费希尔变换以满足正态性。此外，我们将某个感兴趣区域与所有其他感兴趣区域之间的功能连接值组成的向量，视为该感兴趣区域的特征向量。

2.2 BrainNPT 模型的架构

我们提出了一种基于 Transformer 的神经网络 BrainNPT。受自然语言处理中的 <cls> token 启发，该网络将 <cls> 分类嵌入向量用作虚拟感兴趣区域，以收集丰富的脑网络结构模式进行分类。

我们提出的 BrainNPT 架构如图1所示。BrainNPT 基于 Transformer结构 [9], 构建 <，其中包括三部分：分类嵌入向量 <cls>、Transformer 模块和多层感知机（MLP）模块。对于输入形状为 $N \times N$ 的脑网络，我们首先将其与可学习的分类嵌入向量 <cls> 拼接，该向量长度为 N，而 $(N+1) \times N$ 为实际输入形状。随后，我们将嵌入向量 <cls> 作为脑功能网络的表征向量，并将其输入 Transformer块。最后，将脑功能网络的表征向量输入多层感知机（MLP）模块，以生成分类概率。

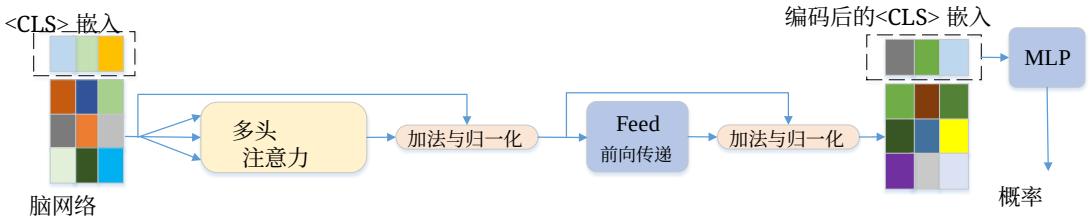


图 1：BrainNPT 的架构。该架构包括分类嵌入向量 <cls>、Transformer 模块和 MLP块。Transformer 模块可以堆叠成多个层。

为了利用输入序列的位置信息，Transformer 方法通常会采用源自自然语言处理或计算机视觉领域的 Transformer 方法所使用的位置嵌入策略，例如正弦余弦编码策略 [10]。对于脑功能网络，由于功能连接矩阵具有对称性和稠密性，连接值表示感兴趣区域之间的连接信息，因此对稠密的脑网络使用特征向量或其他位置编码策略，成本高昂且效果不佳 [19]。此外，将边权重结合到自注意力中可能会损害自注意力的有效性和性能[27]。因此，在我们提出的方法中，直接将功能连接矩阵视为带有位置信息嵌入的输入，以便进一步执行自注意力运算。

2.2.1 分类嵌入向量 <cls>

受 GraphTrans [17] 以及自然语言处理中的 <cls> 标记启发，我们为 BrainNPT 中的 Transformer 使用了一个可学习的分类嵌入向量 <cls>，它有别于图神经网络方法中传统的读出层，例如最大池化、平均池化和基于注意力的读出层 [28, 29]。具体而言，我们将 <cls>分类嵌入向量与脑功能网络拼接，然后将其作为自注意力模块的输入，并将其设为可学习参数。由于我们没有位置嵌入模块， <cls> 分类嵌入向量在训练初始阶段与模型参数一起随机初始化，因此没有实际含义。在后续训练过程中，模型会隐式地学习将脑功能网络的全局信息汇集到 <cls>嵌入向量中。

在 BrainNPT 中， <cls> 分类嵌入向量可以视为脑功能网络中的一个虚拟感兴趣区域。在执行自注意力运算时，该虚拟感兴趣区域会与其他感兴趣区域交互，并从中获取信息。同时，它有助于自注意力模块学习感兴趣区域之间更多的长程交互依赖。

2.2.2 Transformer 模块

Transformer 模块包含 Transformer 的两个核心模块：自注意力和前馈神经网络（FFN）。考虑到脑网络中感兴趣区域（ROI）位置固定，类似于自然语言处理中对句子中每个位置执行的归一化操作 [30], ，我们在这两个子模块中分别执行层归一化操作，以对输入脑网络中每个 ROI 的连接向量进行归一化，并使用残差连接确保训练的稳定性。此外，我们在 Transformer 模块中采用 GELU [31] 作为激活函数，以增强模型的表达能力。

为了在 Transformer 模块中学习脑网络各感兴趣区域（ROI）之间的相互作用，我们为 BrainNPT 采用了与 BERT [10], 相同的双向编码策略，且不使用额外的掩码。

Transformer 模块的自注意力和前馈神经网络子模块定义如下。对于自注意力，我们采用可学习加权矩阵将脑功能网络投影为三个表示矩阵（查询、键和值），并利用查询和键获得每个感兴趣区域对应各感兴趣区域的权重。最后，结合值和感兴趣区域的权重得到更新后的嵌入，如公式（1）所示：

$$Attention\left(x\right)=softmax\left(\frac{QxK^{T}x}{\sqrt{d_{k}}}\right)Vx \tag{1}$$

其中， Q、 K、 V分别表示查询、键和值， d_k表示连接向量的维度， $softmax$ （·）表示 Softmax 函数。

随后，采用多头自注意力机制，并对每个头生成的表示矩阵采用拼接策略，以汇总得到最终表示，作为自注意力子模块的输出，如公式（2）所示：

$$mh_Attention\left(x\right)=Concatenate\left(Attention\left(x\right)\right) \tag{2}$$

其中， $Concatenate$（·）表示拼接操作。

最后，将自注意力子模块的输出输入前馈神经网络，以进一步增强非线性表达能力，如公式（3）所示：

$$FFN\left(x\right)=max\left(0,xW_{1}+b_{1}\right)W_{2}+b_{2} \tag{3}$$

其中， W_1 和 W_2 是用于线性投影的权重矩阵， b_1 和 b_2 表示偏置。分类嵌入向量 <cls> 获取全局表征信息后，将其用作脑网络的表征向量进行分类。

2.2.3 多层感知机

获得的脑功能网络表征向量被送入多层感知机进行分类，并以 softmax 作为最后一层来生成预测概率。多层感知机中使用丢弃层以帮助减少过拟合。

2.3 BrainNPT 的预训练框架

我们提出了一个 BrainNPT 预训练框架，包括预训练数据集的构建和所设计的预训练策略。在针对下游分类任务进行微调之前，BrainNPT 首先通过预训练任务进行训练，以初始化参数并学习脑网络的结构信息。BrainNPT 的预训练和微调框架如图 2 所示。Transformer 模块和 <cls>分类嵌入向量的设置在预训练模型和微调模型之间保持一致，而预训练模型与微调模型之间的差异主要在于多层感知机部分。

为了利用无标签数据进行预训练学习，我们首先提出了一种构建脑功能网络预训练数据集的方法。随后，我们提出了 BrainNPT 的预训练策略，使其能够在微调下游分类任务之前获取脑功能网络的一般结构信息。

2.3.1 预训练数据集的构建

在本研究中，我们采用滑动窗口处理方法构建预训练数据集，以扩大预训练数据集的规模。基于无标签且经过预处理的静息态功能磁共振成像数据，我们从每个受试者中获得了多个功能连接矩阵，窗口长度为 W，步数为 S，并根据与下游有标签数据相同的脑图谱提取脑网络的时间序列。由于每个时间窗的连接矩阵（即快照网络）仍然反映了整个时间段内脑功能网络的部分特征，因此可以将其视为用于预训练任务的脑网络数据。通过滑动窗口操作，预训练数据的规模得到了有效扩展。具体而言，对于时间序列长度为 T 的受试者，通过滑动窗口可以构建 $((T-W)\ /S\ +1)$ 个脑网络，以进行进一步的预训练。

2.3.2 替换ROI预测策略

下游分类任务要求模型能够理解感兴趣区域之间的相互作用，并捕捉脑网络的结构信息，例如大脑默认模式网络（DMN）中的子图模式 [32]。这就要求模型学习功能连接矩阵的特征，例如对称性。为了使模型能够获取脑网络中感兴趣区域之间的相互作用，我们提出了一种替换感兴趣区域预测（RRP）方法。

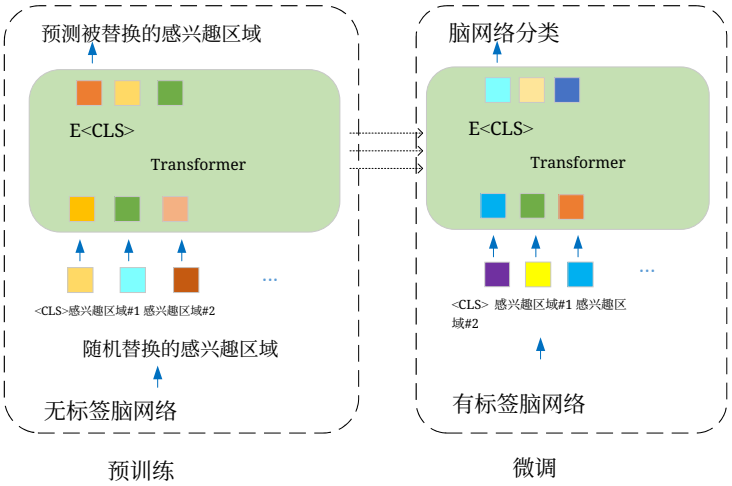


图 2：BrainNPT 的预训练与微调框架。预训练部分在左侧模块中采用随机替换 ROI 策略，微调部分在右侧模块中利用预训练得到的参数执行下游任务。

策略如图 3 所示。该策略将一对脑网络输入 RRP 模型，以预测它们被替换的概率。具体而言，我们将脑网络 A 和 B 作为输入。以 50% 的概率，我们直接将 A 作为 RRP 模型中 Transformer 部分的输入。以剩余的 50% 概率，我们将 A 中 50% 的 ROI 替换为 B中对应 ROI 的功能连接向量。为了增加学习难度，实验中为 A 和 B 选择了最相似的功能连接矩阵。我们选择同一受试者相邻时间窗的功能连接矩阵作为输入A和 B。

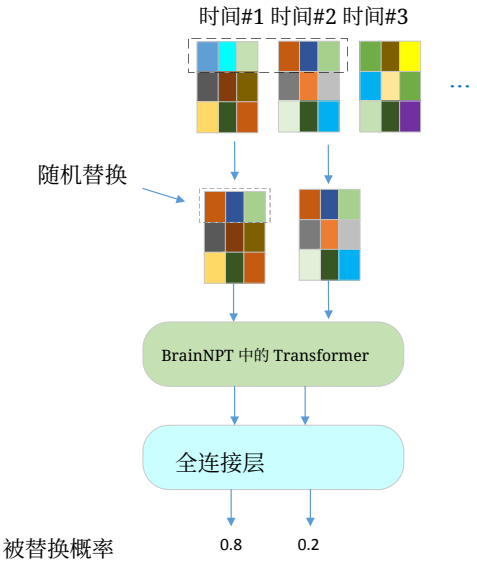


图 3：采用 RRP 进行预训练的框架。

在 RRP 模型的预训练阶段，对于二分类任务，我们采用了与下游微调任务相同的设置。将 <cls> 分类嵌入向量拼接到输入中，并将其传入 RRP 模型的 Transformer 部分。这些设置有助于模型隐式地学习在 <cls> 分类嵌入向量中汇集全局信息。最后，我们将 Transformer 部分输出的 <cls> 分类嵌入向量作为脑网络的表征向量。使用一个输出为全连接层的...

形状为 2，并使用 Softmax 函数预测脑网络是否已被替换。针对 RRP 模型，采用二元交叉熵损失函数进行监督学习。

2.4 模型解释

带有 <cls> 嵌入向量的 Transformer 模型，例如自然语言处理领域的 BERT-base [10] 模型和计算机视觉领域的 ViT-base [23] 模型，已被证明可以通过自归因分数 [33]和逐层相关性传播 (LRP) [34]进行解释。在本研究中，我们针对 BrainNPT 模型调整了一种基于 LRP 的解释方法，以探索脑网络中哪些感兴趣区域对分类具有关键影响。

BrainNPT 包含 <cls> 分类嵌入向量、Transformer 模块和全连接层，并能够使用 LRP 计算各感兴趣区域对分类结果的相关性分数。基于用于解释 BrainNPT 的 LRP，我们可以通过深度泰勒分解方法 [35]获得输入样本的局部相关性。对于训练好的 BrainNPT 模型，可以基于 LRP 计算各感兴趣区域在 Transformer 的自注意力、前馈神经网络、GELU 激活函数和多层感知机中的相关性分数。最后，通过链式法则，可以估算输入脑网络中各感兴趣区域的相关性分数。具体而言，对于 Transformer 模块的解释输出 C，考虑到 BrainNPT 仅使用 <cls> 分类嵌入向量作为表征向量，我们将相应的解释输出 $C_{\langle cls \rangle} \in \mathbb{R}^s$ 视为分类的解释结果，其中包含各感兴趣区域对分类结果的相关性分数。

3 结果与讨论

3.1 BrainNPT模型的分类性能

3.1.1 实验数据集与预处理

我们使用了两个公开静息态功能磁共振成像数据集，即 ABIDE II 数据集 [25] 和 REST-meta-MDD [26], ，开展 BrainNPT 的对比分类实验。我们获取了 ABIDE II [36], 的预处理数据，并通过剔除部分质量较差的样本，选取了 540 个样本，其中包括 288 名自闭症患者和 252 名健康对照组受试者。对于 REST-meta-MDD 数据集，我们筛除了部分图像质量较差的受试者，最终保留了 2027 名受试者，其中包括 1041 名抑郁症患者和 986 名健康对照组受试者。两个数据集的数据均使用 DPARSF [37]进行了预处理。对于每名受试者，我们根据 Craddock200图谱 [36]将全脑划分为 200 个脑区/感兴趣区域。参照第 3 节定义的构建流程，我们为两个数据集中的每名受试者构建了一个大小为 200×200 的功能连接矩阵。

3.1.2 与当前最先进模型的比较

我们将未经预训练的基础 BrainNPT 模型与三类模型进行比较，包括典型的图神经网络模型、专为脑功能网络设计的图神经网络模型，以及专为图数据设计的基于 Transformer 的模型。三类模型共包含十二种：（i）对于典型的图神经网络方法，我们选择了 GAT [38], GCN [39],GraphSAGE [40], GIN [41] 和 DiffPool [42], ，这些方法旨在捕捉图结构信息，并用于来自不同领域的图数据。具体而言，对于 GAT、GCN 和 GraphSAGE，我们选择全局注意力层 [29] 作为读出层。此外，对于典型的图神经网络方法，我们通过边二值化构建脑网络，并将连接值的阈值设为 0.5。（ii）对于专为脑网络设计的图神经网络方法，我们选择 BrainGNN [3] BrainnetCNN [43] 和 FbNet [8] 作为比较方法。这些方法针对脑网络提出，旨在利用脑网络的特征，例如时变特性以及感兴趣区域之间的关系。（iii）对于专为图数据设计的基于 Transformer 的方法，我们将所提出的方法与适用于包括脑网络在内的图数据的基于 Transformer 的方法进行比较，包括 BrainNetTF [19], SAN [44], GraphGPS [18], 和 GraphTrans [17]。GraphTrans 针对图数据设计了特定的位置嵌入策略。

对于我们提出的 BrainNPT，我们采用了五层 Transformer 和三个全连接层。Transformer 的丢弃率设为 0.1，前馈神经网络的维度设为 800，自注意力头数设为 5，激活函数采用 GELU；全连接层中的丢弃层丢弃率设为 0.5。

此外，训练期间选择 AdamW [45] 作为优化器，并选择线性递增预热和平方根倒数衰减策略作为学习率 [46] 调度器。学习率设为 1，训练采用二元交叉熵损失，共进行 500 个训练轮次。在实验中，我们随机划分数据集，其中 80% 用于训练，10% 用于验证，10% 用于测试。性能基于测试集上 5 次随机运行结果的平均值及标准差进行评估。

在实验中，我们选择准确率和曲线下面积作为分类性能的指标。我们将未进行预训练的 BrainNPT 模型与 ABIDE II 数据集和 REST-meta-MDD 数据集上的 12 个模型进行了比较，分类结果如表 1 所示。

表 1：ABIDE II 和 REST-meta-MDD 上的分类性能（%，均值±标准差 ）.

Type	方法	ABIDE II		REST-meta-MDD	
		准确率	AUC	准确率	AUC
典型图神经网络	GAT	57.41±6.62	59.21±7.68	64.53±2.51	67.95±4.03
	GCN	61.11±7.67	66.92±7.79	64.73±2.16	**69.11±2.97**
	GraphSAGE	58.33±4.67	63.69±4.83	62.23±2.17	66.00±2.05
	GIN	62.00±6.89	60.73±7.60	56.09±3.40	55.48±4.10
	DiffPool	59.26±6.88	59.90±7.23	64.05±2.68	68.87±4.61
用于脑网络	BrainGNN	59.26±6.76	66.19±12.52	59.80±2.59	63.79±5.32
	BrainNetCNN	60.69±3.23	64.75±2.88	60.99±1.44	66.16±1.29
	FbNet	58.78±4.01	62.90±5.00	63.05±2.16	65.52±0.90
图 Transformer	SAN	**62.03±10.95**	61.14±14.64	61.25±2.33	63.71±2.83
	GraphGPS	59.88±7.48	57.21±7.78	**65.68±2.22**	**69.73±1.97**
	GraphTrans	61.11±8.25	62.60±9.82	63.25±2.54	69.01±2.94
	BrainNetTF	61.21±5.12	**67.60±5.69**	62.76±1.31	68.27±1.82
Ours	BrainNPT	**62.50±7.14**	**67.88±6.80**	**65.42±3.21**	68.46±3.17

每行中准确率和曲线下面积最高的两项以粗体标出。

从表 1 可以看出，我们提出的 BrainNPT 模型在 13 个模型中性能最佳：在 ABIDE II 数据集上取得了最高的准确率和曲线下面积，在 REST-meta-MDD 数据集上的准确率排名前二。与其他模型相比，基于Transformer的模型表现出较强的竞争力，其中 BrainNetTF 在 ABIDE II 数据集上的曲线下面积排名前二，而 GraphGPS 在 REST-meta-MDD 数据集上取得了最高的准确率和曲线下面积。这表明 Transformer 在脑网络分类方面的强大能力。总之，未经预训练的 BrainNPT 模型在两个数据集上的表现最佳，基于Transformer的模型也展现出卓越的脑网络分类能力。

3.1.3 BrainNPT 模型分析

读出层分析。对于图分类，读出层通常会聚合节点特征，以获得整图表示。我们比较了 BrainNPT 模型中采用两种类型读出层的 <cls> 分类嵌入向量：正交聚类读出层 [19] 和平均池化读出层 [47] 。BrainNPT 在 ABIDE II 数据集上使用不同读出层的结果如表 2 所示。

表 2：BrainNPT 中不同的读出层（均值±标准差）。

读出	准确率（%）	曲线下面积（%）
正交聚类	61.21±5.12	67.60±5.69
平均池化	58.80±4.68	57.36±4.31
Ours	**62.50±7.14**	**67.88±6.80**

从表 2 可以看出，我们提出的 <cls> vector 在准确率和曲线下面积方面取得了最佳性能。与我们提出的方法相比，正交聚类取得了相近但略低的性能。这可能是由于正交聚类的复杂性以及数据集规模较小所致。

Transformer 模块层数的影响。此外，我们还在 ABIDE II 上对 BrainNPT 中 Transformer 模块的层数进行了影响实验，如表 3 所示。

表 3：Diffe BrainNPT 中不同层的 Transformer 模块 （均值±标准差）。

层	准确率 (%)	曲线下面积 (%)
2	58.33±8.50	57.41±6.61
3	57.29±2.50	57.31±4.17
4	60.42±5.52	67.01±9.04
5	62.50±7.14	**67.88±6.80**
6	**62.50±6.24**	64.35±5.05

由 5 层 Transformer 模块构成的模型在准确率和曲线下面积方面表现最佳。随着层数从 2 层增加到 5 层，我们模型的性能持续提升；但 6 层模型的性能并未较 5 层模型显著提高。数据集规模可能会限制更深层 Transformer 模块模型的表达能力和性能。

3.2 BrainNPT模型的预训练

3.2.1 实验数据集与预处理 ng

对于预训练实验， 除 ABIDE II 和 REST-meta-MDD 数据集外，我们还使用了另外两个大规模 e 公开数据集，包括人类 连接组项目（HCP）数据集 [48]和 ABIDE I数据集 [49]。

对于 ABIDE I数据集，我们使用了从图像质量较高的 974 名受试者中筛选出的静息态功能磁共振成像数据 [36], ，其中包括 467 名自闭症患者和 507 名健康对照组。ABIDE I 的静息态功能磁共振成像数据已通过 DPARSF [37]进行预处理。对于 HCP 数据集，我们使用了 S1200 数据集，剔除了采集图像中含有伪噪声的部分受试者，最终保留了 1088 名受试者。HCP 数据集的数据通过 GRETNA [50] 进行预处理，包括头动校正、正则化和平滑。对于每名受试者，使用 Craddock200图谱定义了 200 个脑区/感兴趣区域。

参照第 4.2 节定义的滑动窗口过程，时间窗长度固定为 30，滑动窗口的步长分别根据扫描参数设定。随后，我们对每个时间窗内感兴趣区域的时间序列进行皮尔逊相关系数运算，以构建功能连接矩阵，并对该矩阵进行费希尔变换。ABIDE I、REST-meta-MDD 和 HCP数据集的滑动窗口设置如表 4 所示。

表 4：预训练脑网络的滑动窗口设置。

数据集	受试者	时间窗长度	步数	最小时间窗数	最大时间窗数
ABIDE I	974	30	1	48	286
REST-meta-MDD	2027	30	1	90	240
HCP	1088	30	3	295	387

我们从 ABIDE I数据集获取了 112,879 个脑网络，从 HCP 数据集获取了 421,129 个脑网络，并从 REST-meta-MDD 数据集获取了 504,558 个脑网络用于预训练，将每个时间窗的每个连接矩阵视为一个脑网络。在这些大规模预训练数据中，我们随机选取了 99% 的数据用于训练，1% 用于评估。

3.2.2 与预训练模型的比较

我们将 BrainNPT 与 BrainNetTF [19], 进行了比较，后者在多个脑功能网络数据集上取得了最先进的性能。BrainNetTF 是一种基于 Transformer 的神经网络，与我们的 BrainNPT 不同，其读出层也有所不同。我们使用相同数量的层（6 层 Transformer），对 BrainNPT 和 BrainNetTF模型进行了预训练实验。此外，除读出层外，BrainNPT 和 BrainNetTF 的超参数均设置相同。两个模型均采用了 RRP 预训练策略，以 50% 的概率替换数据中 50% 感兴趣区域的连接向量。

此外，选择 AdamW [45] 作为优化器，并选择线性递增预热和平方根倒数衰减策略作为训练过程中学习率 [46] 的调度器。学习率设置为 1，采用二元交叉熵损失进行训练，训练轮次为 500。在实验中，我们还随机划分了 80% 的数据集用于训练，10% 用于验证，10% 用于测试。此外，性能通过测试集上 5 次随机运行的平均值及其标准差进行评估。

为评估预训练和微调的有效性，我们采用了两种流程：(i) 在 ABIDE I+ REST-meta-MDD + HCP 上进行预训练，并在 ABIDE II 上进行微调；(ii) 在 ABIDE I + HCP 上进行预训练，并在 REST-meta-MDD 上进行微调。此外，我们采用准确率和曲线下面积（AUC）作为下游分类实验的评价指标。

我们分别使用两种流程，在下游分类任务上比较了预训练的 BrainNPT 模型和预训练的 BrainNetTF 模型，结果如表 6 所示。结果表明，BrainNPT

表 5：采用预训练的模型性能（均值±标准差）。

模型	ABIDE II		REST-meta-MDD	
	准确率（%）	曲线下面积（%）	准确率（%）	曲线下面积（%）
BrainNPT	62.50±6.24	64.35±5.05	65.42±3.21	68.46±3.17
BrainNetTF	61.21±5.12	67.60±5.69	62.76±1.31	68.27±1.82
BrainNetTF+预训练	65.16±3.09	69.67±5.57	65.03±2.47	69.60±1.76
BrainNPT+预训练	**71.25±3.42**	**71.54±11.94**	**66.67±3.58**	**70.55±3.48**

采用预训练的模型 ng的表现优于经过预训练的BrainNetTF模型。这两个经过预训练的模型的表现 g

性能优于未经过预训练的模型。经过预训练的 BrainNPT 模型优于未经预训练的模型，准确率显著提高了 8% 以上，曲线下面积提高了 7% 以上。
与表1中其他十二种模型的性能相比，经过预训练的BrainNPT模型在所有指标上均取得了最佳性能，这有力地表明我们的RRP预训练策略有效，能够增强模型对脑网络结构的理解。此外，模型在REST-meta-MDD上的提升不如在ABIDE II 数据集上的提升明显，这可能是由于预训练数据集规模较小所致。这也说明了预训练数据集的规模对于基于Transformer的模型的重要性。与BrianNetTF模型相比，BrainNPT中特别设计的 <cls> 嵌入向量可能通过RRP预训练策略得到了充分预训练，这可能对下游分类任务的性能产生了关键影响。

3.2.3 与预训练策略的比较

为了比较不同的预训练策略，我们提出了一种新的预训练策略，即用于脑功能网络学习的掩码ROI模型（MRM）。受BERT预训练策略启发 [10]，（这是Transformer最广泛使用的预训练策略之一），我们假设随机掩码输入并预测原始掩码值，有助于生成更高层次的全局表示，并帮助模型理解脑网络。对于脑网络，我们随机掩码部分感兴趣区域（ROI）的功能连接向量，并使用由Transformer和全连接层组成的预训练模型来预测原始掩码值。MRM的框架如图4所示。

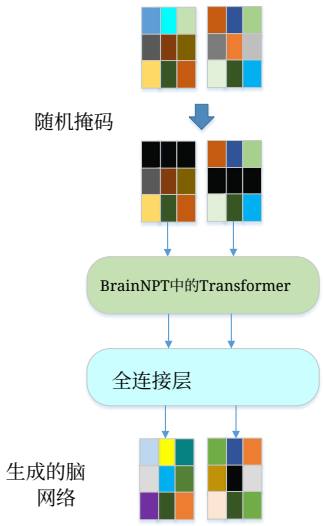


图4：MRM预训练框架。

在实验和分析中，我们随机掩码每个输入脑功能网络中30%的ROI。参照 [51], , 自监督任务是预测整个原始脑网络的值，从而使模型能够学习更多有关脑网络交互的信息。为了尽可能保持预训练数据与下游微调数据的一致性，在实验中进行掩码时，我们没有将所有连接向量都设为0。具体而言，对于被掩码的ROI，我们随机选取其中80%，将其完整的连接向量设为0；另随机选取剩余10%的ROI，将其连接向量设为取值范围与功能连接值相同的正态分布随机值，以进一步增加学习难度，并提高其与原始数据的相似性。最后，剩余10%的ROI保持不变。

在 MRM 的预训练阶段，由于任务是回归预测，因此不采用 <cls> 分类嵌入向量。相反，我们将 Transformer 部分的输出直接输入全连接层，其输出形状与感兴趣区域的功能连接向量长度相同。使用 Dropout层提高稳定性后，将最终输出与原始输入脑网络进行比较，并使用均方误差（MSE）损失函数进行监督训练。我们在 ABIDE I、REST-meta-MDD 和 HCP数据集上对模型进行预训练，并在 ABIDE II 数据集上对模型进行微调，以执行下游分类。结果如表 6 所示。

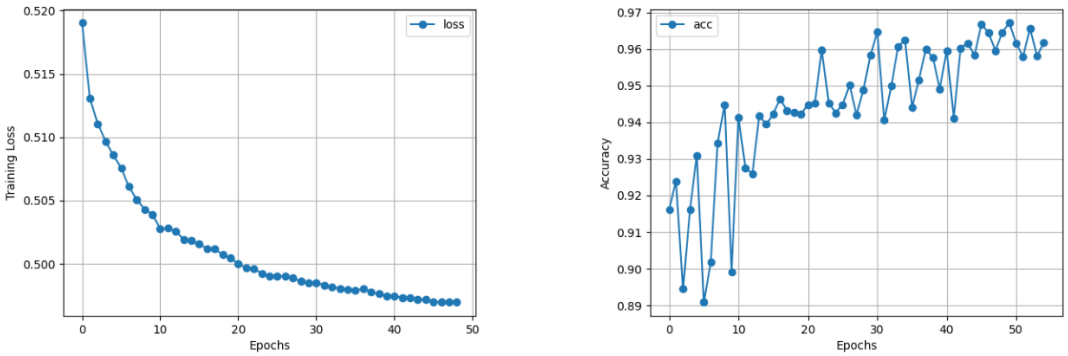
表 6：采用 MRM 和 RRP 的 BrainNPT 的性能（均值±标准差）。

模型	准确率（%）	曲线下面积（%）
BrainNPT	62.50±6.24	64.35±5.05
BrainNPT+MRM	66.67±4.29	**71.73±6.26**
BrainNPT+RRP	**71.25±3.42**	71.54±11.94

结果表明，采用 MRM 和 RRP 预训练的模型，其性能均优于未经过预训练的模型。MRM 的性能不如 RRP 策略，在实验中准确率较低。MRM 的预训练任务是回归，而 RRP 的预训练任务和下游任务都是分类。因此，MRM 的优化损失函数采用均方差，而 RRP 和下游任务的优化损失函数采用交叉熵。不同的优化损失函数可能会影响模型的性能。此外，MRM 模型在预训练时没有 <cls>分类嵌入向量，但在使用小规模数据进行下游微调分类任务时，仍然需要训练 <cls>向量。

3.2.4 预训练中的训练损失与测试准确率分析

在 BrainNPT 模型基于 ABIDE I、REST-meta-MDD 和 HCP 数据集进行预训练的过程中，训练损失和测试准确率如图5所示。结果表明，RRP 的训练损失在约45轮后趋于稳定。随着训练轮次的增加，RRP 的分类准确率稳步提升，最终在96%左右上下波动。这表明，RRP 模型在大规模数据上进行训练后实现了收敛，并有效提升了其预训练任务的性能。



(a) RRP 在训练中的损失　　(b) RRP 在测试中的准确率

图5：RRP 模型的训练和测试评估。训练损失在约45轮后趋于稳定，使用 RRP 预测被替换的感兴趣区域的测试准确率约为 96%。

3.2.5 预训练策略中替换比例的影响

为探究 RRP预训练策略中替换感兴趣区域比例的影响，我们进行了替换比例对比实验，如表7所示。我们使用 ABIDE I 作为预训练数据集进行预训练。选取 95% 的数据作为预训练集，其余 5% 的数据用于评估预训练性能。最后，在 ABIDE II 数据集上采用相同的训练集、验证集和测试集划分设置，以评估其对下游分类任务的影响。我们比较了随机替换 30%、40%、50% 和 60% 感兴趣区域的影响。从表7的结果可以看出，替换比例达到 50% 时，采用 RRP 的模型性能最佳。当替换比例相对较低时，提高替换比例可以提升模型性能；但当替换比例超过 50% 时，模型性能不再提升。这可能是因为持续提高替换比例对模型学习的贡献并不显著。

表7：RRP 中替换比例的影响（均值±标准差）。

比例	准确率（%）	曲线下面积（%）
60%	63.19±5.09	64.46±7.69
50%	**64.58±4.73**	**66.84±6.28**
40%	63.89±8.71	63.77±7.96
30%	63.19±7.67	61.08±7.65

3.2.6 预训练数据规模的影响

公开静息态功能磁共振成像数据集的规模通常有限，仅有数百或数千个样本；而自然语言处理或计算机视觉领域用于 Transformer 预训练的数据集规模可达数百万，因此，在 Transformer 上使用无标签脑功能网络进行预训练仍面临挑战。

为评估预训练数据规模的影响，我们比较了不同规模预训练数据下游分类任务的性能。我们为 RRP 模型设置了三种不同的预训练数据集：（i）ABIDE I 数据集；（ii）ABIDE I 和 HCP 数据集；（iii）全部三个数据集：ABIDE I、HCP 和 REST-meta-MDD。我们采用分类实验中相同比例的训练集、验证集和测试集，在 ABIDE II 数据集上评估下游分类的性能。同时，我们针对 RRP 模型开展了替换比例统一为 50% 的实验。不同规模预训练数据下各预训练策略的性能如图 6 所示。图 6 的结果表明，在替换比例相同的情况下，准确率

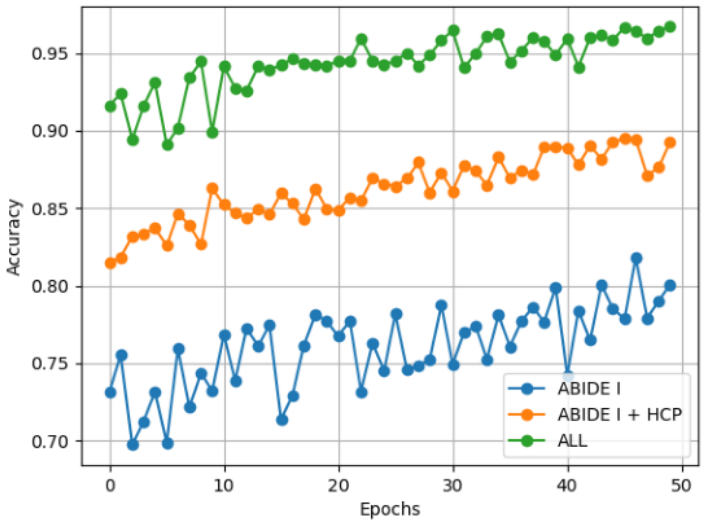


图 6：不同规模的预训练数据集对下游分类准确率的影响。

表8：采用不同规模的预训练数据集在 ABIDE II 上的分类结果（均值±标准差）。

预训练数据集	准确率（%）	曲线下面积（%）
ABIDE I	64.58±4.73	66.84±6.28
ABIDE I +HCP	65.28±3.18	70.04±9.62
ALL	**71.25±3.42**	**71.54±11.94**

随着预训练数据集规模的扩大，RRP替换感兴趣区域预测的准确率也随之提高：仅使用ABIDE I数据集时，准确率约为80%；使用ABIDE I和HCP数据集时，准确率约为90%；使用全部三个数据集时，准确率接近97%。下游分类结果如表8所示，这表明预训练数据越多，模型性能越好。

3.2.7 微调数据集规模的影响

为了评估微调数据集规模的影响，我们在 ABIDE II 数据集上使用不同规模的训练数据对 BrainNPT 进行微调。选择微调性能最佳的模型及 RRP 的预训练参数后，我们使用不同规模的训练集开展分类实验，包括使用数据集的 80%、70%、60% 和 50% 进行训练，并将数据集的 10% 设置为验证集。结果如表 9 所示。

表 9：BrainNPT 在不同规模训练数据集上的性能（均值±标准差）

训练集的比例	已加载预训练参数	准确率（%）	曲线下面积（%）
80%	No	62.50±6.24	64.35±5.05
	Yes	70.83±3.10	71.68±4.34
70%	No	60.42±4.15	64.84±6.39
	Yes	63.19±7.54	63.79±5.95
60%	No	58.85±4.24	62.86±5.78
	Yes	60.42±6.32	62.82±5.93
50%	No	56.94±5.53	58.92±7.53
	Yes	60.94±5.82	63.51±6.79

结果显示，在相同设置下，随着训练集规模缩小，模型性能大幅下降。准确率和曲线下面积最终分别稳定在约 57% 和 60%。这表明训练数据的减少对模型性能产生了巨大影响。采用预训练的 BrainNPT 模型仍保持较高的性能，显著优于未采用预训练的模型。尤其是在仅使用 50% 训练数据时，其分类准确率仍达到 60%，而未采用预训练的 BrainNPT 模型仅为 56.94%。这些结果表明，即使标注数据有限，预训练也能显著提升下游分类任务的性能。

3.3 模型解释

3.3.1 分类模型中的关键感兴趣区域

基于采用RRP预训练的BrainNPT最佳性能模型，开展了解释性实验，以探索与分类最相关的感兴趣区域。借助训练好的BrainNPT模型，对于任意给定的脑网络，我们都可以使用LRP方法获得相应感兴趣区域的相关性分数。具体而言，在获得所有相关性分数后，采用最小-最大标准化将分数缩放至（0,1）范围内。最小-最大标准化的操作如下所示。

$$\bar{C}=\frac{C-\mathrm{C}_{\min}}{\mathrm{C}_{\max}-\mathrm{C}_{\min}} \tag{4}$$

其中，$\mathrm{C}_{\max}$ 和$\mathrm{C}_{\min}$ 分别表示最高和最低相关性分数。

通过对缩放后的相关性分数进行可视化，可以观察每个受试者对应预测类别中最相关的感兴趣区域，如图7和图8所示，其中颜色越深表示相关性越高。

为了获得BrainNPT模型在ABIDE II上预测自闭症谱系障碍时最重要的感兴趣区域，采用基于频率的方法 [52]评估感兴趣区域在模型中的重要性。对于每个受试者，将相关性分数最高的 K 个感兴趣区域加入感兴趣区域子集 F_K。统计所有受试者后，将子集 F_K 中出现频率超过受试者数量60%的感兴趣区域视为预测自闭症受试者时最重要的感兴趣区域。为了获得预测自闭症受试者时排名前3的关键感兴趣区域，我们逐步将 K 的数量提高到84，直到有3个感兴趣区域满足上述最重要感兴趣区域的条件。这3个关键感兴趣区域分别是Craddock200图谱中的第191个、第82个和第2个感兴趣区域。

3.3.2 关键感兴趣区域分析

由于 Craddock200图谱中的感兴趣区域属于聚类感兴趣区域，因此我们根据这些感兴趣区域在 AAL脑图谱 [53], 中对应的感兴趣区域进行了分析，该图谱在脑影像分析中得到了最广泛的应用。排名前三的感兴趣区域涉及 AAL脑图谱中的五个感兴趣区域，分别为左侧颞中回（Temporal_Mid_L）、左侧角回（Angular_L）、左侧枕中回（Occipital_Mid_L）、左侧额上回（Frontal_Sup_L）和左侧额中回（Frontal_Mid_L）。我们的解释结果与此前关于自闭症谱系障碍的研究一致。例如，已有研究表明，角回与阅读、文本理解和注意力高度相关 [54], ，而左侧角回也有】【。

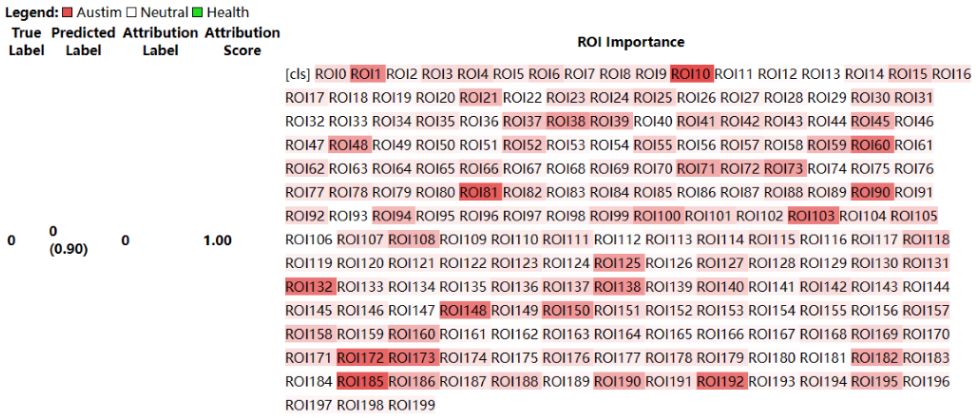


图 7：预测为自闭症的受试者的可视化结果。感兴趣区域 ID 的颜色越深，表示该感兴趣区域与自闭症的相关性越高。

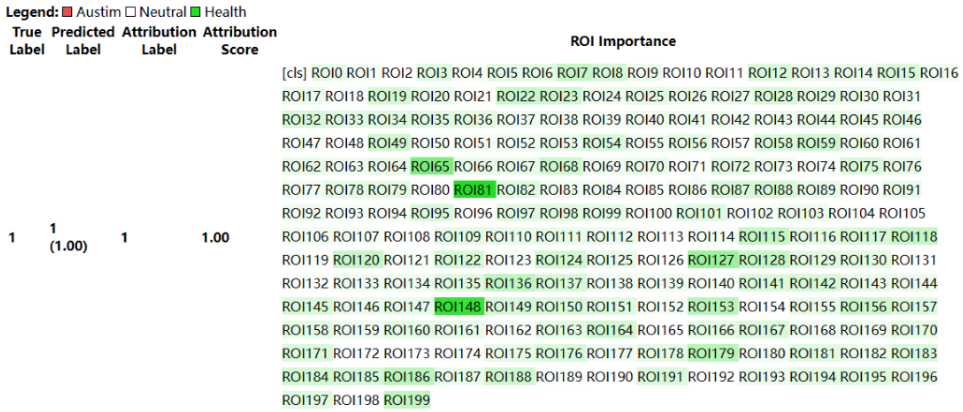


图 8：预测为健康对照的受试者的可视化结果。感兴趣区域 ID 的颜色越深，表示该感兴趣区域与健康状态的相关性越高。

已被证实与社交互动障碍有关 [55]。此外，作为大脑中负责处理社会网络的关键区域，颞中回已被证实与自闭症 [56], 高度相关，近期研究进一步证实，自闭症患者的左侧颞中回存在显著的功能连接变化 [57]。此外，枕中回是负责物体识别的重要脑区，并且经常参与面部识别和情绪识别任务 [58]。研究还表明，在执行面部表情识别任务时，自闭症患者与健康对照组之间存在功能连接差异 [59]。额上回是默认模式网络的重要组成部分 [32] ，并在社会行为中发挥关键作用，研究还发现自闭症患者存在功能连接损伤 [60]。此外，神经认知研究发现，额中回与自闭症的病理生理学有关 [61, 62, 63], ，其皮层体积和连接存在异常，进而导致工作记忆加工缺陷及其他异常。

4 结论

本文提出了一种用于脑功能网络分类的基于Transformer的神经网络和预训练框架，利用预训练从无标注脑网络数据中捕获内在结构信息。本文提出了一种可学习的 <cls> 标记，作为分类嵌入向量，使Transformer能够捕获脑网络表示；并提出了替换感兴趣区域预测策略用于预训练。实验结果表明，经过预训练的BrainNPT模型优于当前最先进的模型。此外，

进一步分析了预训练策略和模型参数的影响，并通过 LRP 对训练好的 BrainNPT 模型进行了阐释。

未来，BrainNPT 可以通过增加 Transformer 的层数和更多无标签脑功能网络数据来改进，并可用作预训练模型，以开展进一步的脑网络分析，例如识别精神障碍的亚型，以及表征健康成年人的脑功能连接及其变异性。

CRediT 作者贡献声明

Jinlong Hu: 概念构思、方法学、调查研究、撰写——初稿、撰写——审阅与编辑、监督、项目管理、经费获取。**Yangmin Huang:** 概念构思、方法学、调查研究、撰写——初稿、撰写——审阅与编辑。**Nan Wang:** 撰写——审阅与编辑。**ShoubinDong:** 概念构思、方法学、资源、经费获取。

利益冲突声明

作者声明不存在任何利益冲突。

致谢与基金资助

本研究得到广东省自然科学基金 [2021A1515011942], 和中山市引进高端科研机构创新基金 [2019AG031]的资助。我们感谢不列颠哥伦比亚大学的李晓晓博士分享意见，帮助改进本研究。

参考文献

[1] Meenakshi Khosla、Keith Jamison、Gia H. Ngo、Amy Kuceyeski 和 Mert R. Sabuncu。静息态功能磁共振成像分析中的机器学习。《磁共振成像》，64：101–121，2019年。[2] Wutao Yin、Longhai Li 和吴方相。基于功能磁共振成像的脑部疾病诊断深度学习。《神经计算》，469：332–345，2022。[3] 李晓晓、周远、Nicha Dvornek、张慕涵、高思远、庄俊堂、Dustin Scheinost、Lawrence H Staib、Pamela Ventola 和 James S Duncan。BrainGNN：用于功能磁共振成像分析的可解释脑图神经网络。《医学图像分析》，74：102233，2021年。[4] Jinlong Hu、Lijie Cao、Tenghui Li、Shoubin Dong 和 Ping Li。GAT-LI：一种用于功能性脑网络分类的基于图注意力网络的学习与解释方法。《*BMC*生物信息学》，22(1)：1–20，2021年。[5] Byung-Hoon Kim 和 Jong Chul Ye。理解用于静息态功能磁共振成像功能连接分析的图同构网络。《神经科学前沿》，第630页，2020年。[6] 严宇君、朱炯、Marlena Duda、Eric Solarz、Chandra Sripada 和 Danai Koutra。GroupINN：用于有限、含噪脑数据分类的分组式可解释神经网络。载于第*25*届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，第772–782页，2019年。[7] Hejie Cui、Wei Dai、Yanqiao Zhu、李晓晓、Lifang He 和 Carl Yang。用于基于脑连接组的脑部疾病分析的可解释图神经网络。载于医学图像计算与计算机辅助干预——*MICCAI 2022*：第*25*届国际会议，新加坡，*2022*年*9*月*18*日至*22*日，论文集，第*VIII*部分，第375–385页。施普林格，2022。[8] Xuan Kan、Hejie Cui、Joshua Lukemire、Ying Guo 和 Carl Yang。FBNetGEN：通过功能性脑网络生成实现任务感知的基于图神经网络的功能磁共振成像分析。载于医学影像与深度学习国际会议，第618–637页。PMLR，2022。[9] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。注意力就是一切。神经信息处理系统进展，30，2017年。[10] Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。BERT：用于语言理解的深度双向变换器预训练。*arXiv*预印本 *arXiv:1810.04805*，2018。[11] Alec Radford、Karthik Narasimhan、Tim Salimans、Ilya Sutskever 等。通过生成式预训练提升语言理解。2018。

[12] Jiawei Zhang、Haopeng Zhang、Congying Xia 和 Li Sun。Graph-Bert：图表示学习只需要注意力。arXiv 预印本 arXiv:2001.05140，2020年。[13] Yu Rong、Yatao Bian、Tingyang Xu、Weiyang Xie、Ying Wei、Wenbing Huang 和 Junzhou Huang。基于大规模分子数据的自监督图 Transformer。神经信息处理系统进展，33：12559–12571，2020。[14] Grégoire Mialon、Dexiong Chen、Margot Selosse 和 Julien Mairal。GraphiT：在 Transformer 中编码图结构。arXiv 预印本 arXiv:2106.05667，2021年。[15] Chengxuan Ying、Tianle Cai、Shengjie Luo、Shuxin Zheng、Guolin Ke、Di He、Yanming Shen 和 Tie-Yan Liu。Transformer 在图表示方面的表现真的很差吗？神经信息处理系统进展，34：28877–28888，2021年。[16] Dexiong Chen、Leslie O'Bray 和 Karsten Borgwardt。用于图表示学习的结构感知 Transformer。载于国际机器学习会议，第 3469–3489 页。PMLR，2022年。[17] Zhanghao Wu、Paras Jain、Matthew Wright、Azalia Mirhoseini、Joseph E Gonzalez 和 Ion Stoica。利用全局注意力为图神经网络表示长程上下文。神经信息处理系统进展，34：13266–13279，2021年。[18] Ladislav Rampášek、Michael Galkin、Vijay Prakash Dwivedi、Anh Tuan Luu、Guy Wolf 和 Dominique Beaini。通用、强大且可扩展的图 Transformer 的设计方案。神经信息处理系统进展，35：14501–14515，2022年。[19] Xuan Kan、Wei Dai、Hejie Cui、Zilong Zhang、Ying Guo 和 Carl Yang。脑网络 Transformer。arXiv 预印本 arXiv:2210.06681，2022年。[20] Ali Hassani、Steven Walton、Nikhil Shah、Abulikemu Abuduweili、Jiachen Li 和 Humphrey Shi。借助紧凑型 Transformer 摆脱大数据范式。arXiv 预印本 arXiv:2104.05704，2021年。[21] Xinlei Chen、Saining Xie 和 Kaiming He。自监督视觉 Transformer 训练的实证研究。载于 IEEE/CVF 国际计算机视觉会议论文集，第 9640–9649 页，2021年。[22] Yi Yang、Hejie Cui 和 Carl Yang。预训练用于脑网络分析的图神经网络。载于 2022 IEEE 国际大数据会议（Big Data），第 4993–4994 页。IEEE，2022年。[23] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly 等。一张图像胜过 16×16 个词：用于大规模图像识别的 Transformer。arXiv 预印本 arXiv:2010.11929，2020。[24] Ze Liu、Yutong Lin、Yue Cao、Han Hu、Yixuan Wei、Zheng Zhang、Stephen Lin 和 Baining Guo。Swin Transformer：利用移位窗口的层次化视觉 Transformer。载于 IEEE/CVF 国际计算机视觉会议论文集，第 10012–10022 页，2021年。[25] Adriana Di Martino、David O'connor、Bosi Chen、Kaat Alaerts、Jeffrey S Anderson、Michal Assaf、Joshua H Balsters、Leslie Baxter、Anita Beggiato、Sylvie Bernaerts 等。利用自闭症脑成像数据交换 II 增进自闭症连接组研究。《科学数据》，4(1)：1–15，2017年。[26] Chao-Gan Yan、Xiao Chen、Le Li、Francisco Xavier Castellanos、Tong-Jian Bai、Qi-Jing Bo、Jun Cao、Guan-Mao Chen、Ning-Xuan Chen、Wei Chen 等。复发性重度抑郁障碍患者的默认模式网络功能连接减弱。美国国家科学院院刊，116(18)：9078–9083，2019年。[27] Hejie Cui、Wei Dai、Yanqiao Zhu、Xuan Kan、Antonio Aodong Chen Gu、Joshua Lukemire、Liang Zhan、Lifang He、Ying Guo 和 Carl Yang。BrainGB：用于图神经网络脑网络分析的基准。IEEE 医学影像汇刊，2022年。[28] Junhyun Lee、Inyeop Lee 和 Jaewoo Kang。自注意力图池化。载于国际机器学习会议，第 3734–3743 页。PMLR，2019年。[29] Yujia Li、Daniel Tarlow、Marc Brockschmidt 和 Richard Zemel。门控图序列神经网络。arXiv 预印本 arXiv:1511.05493，2015年。[30] Sharan Narang、Hyung Won Chung、Yi Tay、William Fedus、Thibault Fevry、Michael Matena、Karishma Malkan、Noah Fiedel、Noam Shazeer、Zhenzhong Lan 等。Transformer 修改能否跨实现方式和应用场景迁移？arXiv 预印本 arXiv:2102.11972，2021年。[31] Dan Hendrycks 和 Kevin Gimpel。高斯误差线性单元（GELU）。arXiv 预印本 arXiv:1606.08415，2016年。[32] Marcus E Raichle。大脑的默认模式网络。《神经科学年度评论》，38：433–447，2015年。

[33] Yaru Hao、Li Dong、Furu Wei 和 Ke Xu。自注意力归因：解释 Transformer 内部的信息交互。载于《AAAI人工智能会议》论文集，第35卷，第12963–12971页，2021年。[34] Hila Chefer、Shir Gur 和 Lior Wolf。超越注意力可视化的 Transformer可解释性。载于《IEEE/CVF计算机视觉与模式识别会议》论文集，第782–791页，2021年。[35] Grégoire Montavon、Sebastian Lapuschkin、Alexander Binder、Wojciech Samek 和 Klaus-Robert Müller。利用深度泰勒分解解释非线性分类决策。《模式识别》，65：211–222，2017年。[36] Cameron Craddock、Yassine Benhajali、Carlton Chu、Francois Chouinard、Alan Evans、András Jakab、Budhachan- dra Singh Khundrakpam、John David Lewis、Qingyang Li、Michael Milham 等。神经科学局预处理计划：开放共享预处理神经影像数据及其衍生数据。《神经信息学前沿》，7：27，2013年。[37] Chaogan Yan 和 Yufeng Zang。DPARSF：用于静息态 fMRI“流水线”数据分析的 MATLAB工具箱。《系统神经科学前沿》，第13页，2010年。[38] Petar Veli˘ckovi´c、Guillem Cucurull、Arantxa Casanova、Adriana Romero、Pietro Lio 和 Yoshua Bengio。图注意力网络。arXiv预印本 arXiv:1710.10903，2017年。[39] Thomas N Kipf 和 Max Welling。利用图卷积网络进行半监督分类。arXiv预印本 arXiv:1609.02907，2016年。[40] Will Hamilton、Zhitao Ying 和 Jure Leskovec。大型图上的归纳式表示学习。《神经信息处理系统进展》，30，2017年。[41] Keyulu Xu、Weihua Hu、Jure Leskovec 和 Stefanie Jegelka。图神经网络的能力有多强？arXiv预印本 arXiv:1810.00826，2018年。[42] Zhitao Ying、Jiaxuan You、Christopher Morris、Xiang Ren、Will Hamilton 和 Jure Leskovec。利用可微池化进行层次化图表示学习。《神经信息处理系统进展》，31，2018年。[43] Jeremy Kawahara、Colin J Brown、Steven P Miller、Brian G Booth、Vann Chau、Ruth E Grunau、Jill G Zwicker 和 Ghassan Hamarneh。BrainNetCNN：用于脑网络的卷积神经网络；迈向神经发育预测。《NeuroImage》，146：1038–1049，2017年。[44] Devin Kreuzer、Dominique Beaini、Will Hamilton、Vincent Létourneau 和 Prudencio Tossou。重新思考具有谱注意力的图Transformer。《神经信息处理系统进展》，34：21618–21629，2021年。[45] Ilya Loshchilov 和 Frank Hutter。解耦权重衰减正则化。arXiv预印本 arXiv:1711.05101，2017年。[46] Colin Raffel、Noam Shazeer、Adam Roberts、Katherine Lee、Sharan Narang、Michael Matena、Yanqi Zhou、Wei Li 和 Peter J Liu。探索统一文本到文本Transformer的迁移学习极限。《机器学习研究杂志》，21(1)：5485–5551，2020年。[47] Byung-Hoon Kim、Jong Chul Ye 和 Jae-Jin Kim。利用时空注意力学习脑连接组的动态图表示。《神经信息处理系统进展》，34：4314–4327，2021年。[48] David C Van Essen、Stephen M Smith、Deanna M Barch、Timothy EJ Behrens、Essa Yacoub、Kamil Ugurbil、Wu-Minn HCP联盟等。Wu-Minn人类连接组计划：概述。《NeuroImage》，80：62–79，2013年。[49] Adriana Di Martino、Chao-Gan Yan、Qingyang Li、Erin Denio、Francisco X Castellanos、Kaat Alaerts、Jeffrey S Anderson、Michal Assaf、Susan Y Bookheimer、Mirella Dapretto 等。自闭症脑成像数据交换项目：迈向对自闭症内在脑结构的大规模评估。《分子精神病学》，19(6)：659–667，2014年。[50] Jinhui Wang、Xindi Wang、Mingrui Xia、Xuhong Liao、Alan Evans 和 Yong He。GRETNA：用于影像连接组学的图论网络分析工具箱。《人类神经科学前沿》，9：386，2015年。[51] Kevin Clark、Minh-Thang Luong、Quoc V Le 和 Christopher D Manning。ELECTRA：将文本编码器预训练为判别器，而非生成器。arXiv预印本 arXiv:2003.10555，2020年。[52] Jinlong Hu、Lijie Cao、Tenghui Li、Bin Liao、Shoubin Dong 和 Ping Li。静息态功能连接分析中的可解释学习方法：以自闭症谱系障碍为例。《医学中的计算与数学方法》，2020，2020年。[53] Nathalie Tzourio-Mazoyer、Brigitte Landeau、Dimitri Papathanassiou、Fabrice Crivello、Octave Etard、Nicolas Delcroix、Bernard Mazoyer 和 Marc Joliot。利用 MNI单被试脑磁共振影像的宏观解剖分区，对 SPM 中的激活进行自动解剖标注。《NeuroImage》，15(1)：273–289，2002年。

[54] Mohamed L Seghier。角回：多种功能与多个分区。*The Neuroscientist*，19(1)：43–61，2013年。[55] MC Lai、MV Lombardo 和 S Baron-Cohen。自闭症。《柳叶刀》（伦敦版）383：896–910，2014年。[56] Jocelyne Bachevalier。内侧颞叶结构与自闭症：临床和实验研究发现综述。*Neuropsychologia*，32(6)：627–648，1994年。[57] Jinping Xu、Chao Wang、Ziyun Xu、Tian Li、Fangfang Chen、Kai Chen、Jingjing Gao、Jiaojian Wang 和 Qingmao Hu。自闭症谱系障碍儿童和成人颞中回亚区的特定功能连接模式。*Autism Research*，13(3)：410–422，2020年。[58] Marie T Banich 和 Rebecca J Compton。认知神经科学。剑桥大学出版社，2018年。[59] Kevin A Pelphrey、James P Morris、Gregory McCarthy 和 Kevin S LaBar。自闭症患者对面部情绪和身份动态变化的知觉。*Social Cognitive and Affective Neuroscience*，2(2)：140–149，2007年。[60] Christopher S Monk、Scott J Peltier、Jillian Lee Wiggins、Shih-Jen Weng、Melisa Carrasco、Susan Risi 和 Catherine Lord。自闭症谱系障碍患者内在功能连接异常。*NeuroImage*，47(2)：764–772，2009年。[61] Basilis Zikopoulos 和 Helen Barbas。自闭症中兴奋性和抑制性皮层回路的神经连接改变。人类神经科学前沿，7：609，2013年。[62] Evelien M Barendse、Marc PH Hendriks、Jacobus FA Jansen、Walter H Backes、Paul AM Hofman、Geert Thoonen、Roy PC Kessels 和 Albert P Aldenkamp。自闭症谱系障碍高功能青少年的工作记忆缺陷：神经心理学和神经影像学相关因素。*Journal of Neurodevelopmental Disorders*，5(1)：1–11，2013年。[63] Ruth A Carper 和 Eric Courchesne。早期自闭症额叶皮层局部增大。*Biological Psychiatry*，57(2)：126–133，2005年。