A generalized model for laryngeal physiological activity classification based on fine-tuned pre-trained LLMs

摘要：

LLM4SC

喉部生理活动监测是喉部疾病诊疗的重要手段，在吞咽功能评估、发生功能研究等方面有着广泛的应用。在此，我们提出了一种基于高效参数微调的大语言模型（PEFT LLM）的新型框架LLM4SC(LLM for Signal Classification)，该框架适用于多种喉部生理活动分类任务。LLM4SG通过卷积对齐输入信号数据与模型嵌入，通过选择性和重参数化PEFT方法微调大语言模型，进行信号特征提取与建模，并加入双向LSTM层捕捉信号中的时间依赖关系。重要的是，LLMs本身拥有强大的特征提取能力和上下文建模能力，同时层冻结与LoRA微调策略相结合可在少量数据中取得较好结果。最后，我们选取三种轻量级LLM（GPT2, Llama3-1B, Qwen2.5B）进行实验，结果表明，LLM4SG的分类准确率显著优于机器学习和深度学习模型，准确率平均提升6.3%；同时在few-shot和zero-shot任务中表现出色，平均提升8.8%，7.8%。

1 引言

喉部是人体呼吸、发声和吞咽的核心部位，其生理活动的准确监测对于诊断相关疾病（如喉炎引发的声带功能评估等）以及开发新型治疗手段（如吞咽障碍患者的肌肉协调性分析等）具有重要意义[1,2]。传统的喉部监测手段主要依赖于大型医疗设备的周期性检查和主观评估量表[3-7]，近年来智能可穿戴设备（如加速度传感器和表面肌电传感器[8,9]）能够非侵入地捕捉喉部肌肉运动的细微变化，更加便捷的采集到大量监测数据。随着监测数据量的增长，传统分析方法在特征提取与模式识别层面体现出一定局限，对此人工智能算法通过深度学习框架对大量数据进行学习，较好的完成信号分类任务，为喉部疾病的早期诊断和康复过程提供帮助[10,11]。

喉部生理动作分类任务中，支持向量机（SVM）、K近邻（KNN）等传统的机器学习算法通过提取手工设计的特征来对喉部动作信号进行分类[12]。Miyagi等人（2020）提出了一种基于支持向量机的喉部动作信号分类方法，通过提取时频域特征，区分正常吞咽和异常吞咽[13]。类似地，Liu等人（2020）提出一种通过线性判别分析模型的语音识别模型，对小波变换降噪处理后的信号构建特征向量，在语音指令中表现出90%以上准确率[14]。这些方法依赖手工设计特征进行分类，虽在特定任务有不错表现，但存在一些局限，如处理复杂时序信号能力欠佳[15]。相比之下，深度学习凭借其强大的特征提取能力显示出优势。卷积神经网络（CNN）拥有高效提取局部特征的能力，在喉部动作分类任务中展现出显著潜力[16-18]。Yang等人（2023）则提出了一种基于AlexNet（特征提取）、ReliefF（特征选择）和SVM（分类）的三步集成模型，显著提升了语音识别的准确率[19]。Song等人（2025）则利用WaveNet、ResNet50和EfficientNet三种深度神经网络，实现了咳嗽、说话、吞咽和清喉等任务的分类，准确率高达95.96%[20]。此外，循环神经网络（RNN）凭借隐藏状态传递机制，使得其在动态特征提取任务中展现出一定优势[21]。Hou等人（2024）提出了一种基于碳纳米管的触觉口腔垫，结合LSTM网络，能够有效捕捉喉部动作信号中的时间依赖性，进一步提升了分类性能[22]。然而，目前的算法在多模态数据融合和特征提取方面需依赖大量数据的训练，尤其在数据标注稀缺的医学场景中有一定的不足[23]。

预训练模型强大的迁移学习能力和自监督学习范式，可以在大量数据中学习潜在特征，从而具有更强的泛化能力[25]。例如，Xue等人(2021)提出通过自监督对比学习提取EEG信号的潜在特征，提升睡眠分期任务准确率至88.16%[24]；类似的，Sadiq等人提出基于预训练 CNN 的自动化框架，用于脑机接口（BCI）系统中运动和心理意象 EEG 信号识别，分类准确率达到 99.52%。当前，基于Transformer框架的预训练模型逐渐成为新的研究热点，这类模型的核心在于多头注意力机制[25,26]，通过并行计算输入序列各位置间的关联权重，实现了对长距离依赖关系的动态捕获；同时轻量化设计有效解决了数据量有限的不足。模型对时序特征的非线性建模能力，与脑电图（EEG）、心电图（ECG）等生理信号的高维度、非平稳特性相适应。以EEGPT为例，其提出的双自监督预训练框架结合了时空表示对齐和掩码信号重建任务，通过剪枝注意力头、量化低秩投影进行轻量化设计，在运动想象、睡眠分期等任务中超越基线模型[27]。同样基于Transformer框架的大语言模型（BERT[28]、GPT[29]等）通过大规模预训练具备跨模态任务的处理能力，参数高效微调（PEFT） 通过仅优化少量参数适应下游任务，PEFT方法包括选择性PEFT（如固定层冻结[30]，动态层冻结[31]），附加式PEFT[32]，提示式PEFT，重参数化PEFT（如LoRA[34]），混合式PEFT等[35]，有效应对数据稀缺的情况。

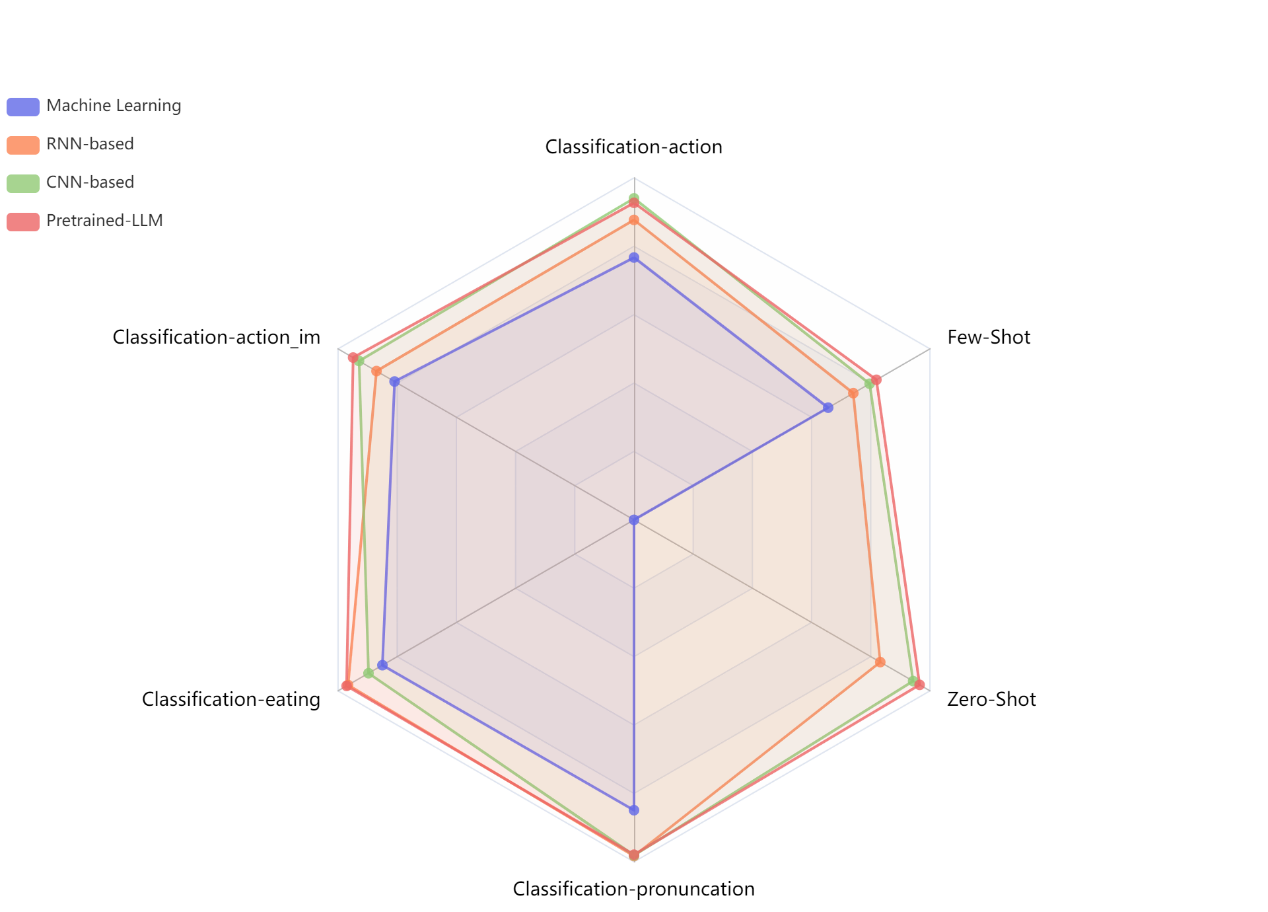
针对喉部信号数据集稀缺情况下的分类任务，我们提出了Fine-tune Pretrained LLM For Signal Classification(LLM4SG)模型架构，通过卷积嵌入层与位置编码实现跨模态特征映射，选取轻量级的 Llama3-1B、Qwen-2.5B和GPT2等预训练模型，使用层冻结与低秩适应的混合微调策略展开实验验证。研究结果表明，微调后的大模型在小体量喉部动作数据集上优势明显，对比机器学习模型提升6%~24%，深度学习模型提高1%~4%，并且在Few-Shot和Zero-Shot任务中表现优异。

我们的创新点主要体现在以下几个方面：

1. 我们引入LLS4SG，这是一种适用多种LLMs的模型架构，并加入微调机制。该架构是专门设计的，能够在小数据集的情况下，捕捉数据中的规律。

2. 我们对微调机制进行设计，选择层冻结与LoRA结合的微调方式，同时在架构中加入双向LSTM层，显著提高模型分类准确率；

3. 我们进行了全面的实验，对比多种轻量级预训练模型在小数据集上的表现，通过Few-Shot和Zero-Shot实验，展现了预训练大语言模型在喉部信号分类任务上的通用性。



2 准备工作

1. 问题表述

本文将讨论吞咽信号的分类任务，吞咽信号的采集来源为多种传感器。对于连续的吞咽信号，，其中C表示变量维度，T表示信号长度。通常，信号会进行降采样成固定的长度N。信号的标签为，根据不同分类任务取值范围不同。我们的任务是对信号依据标签进行分类。

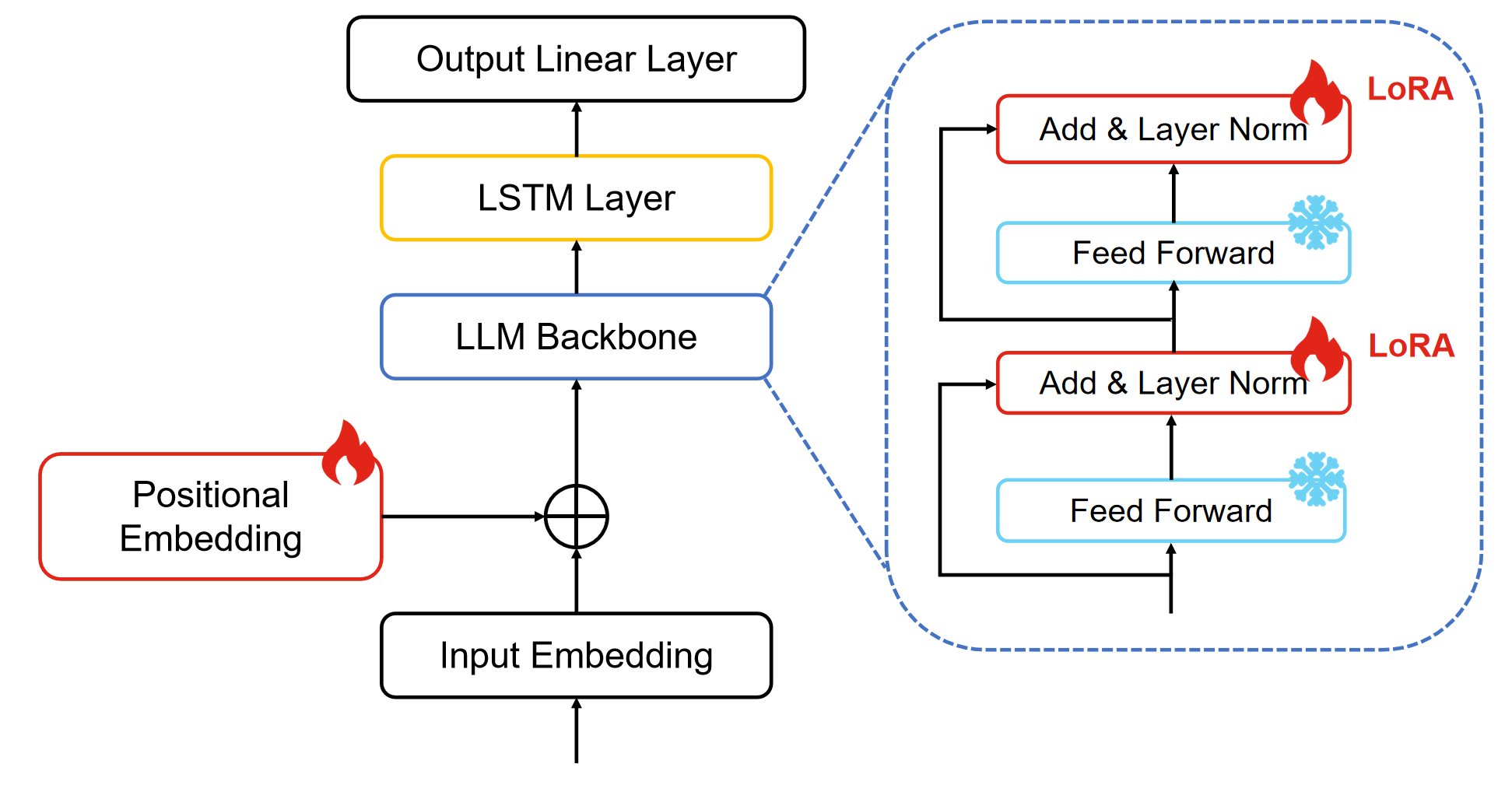
B．用于信号分类的LLM Backbone

在我们的模型框架中，对于LLM的使用，保留了原有预训练模型中的位置嵌入层和子注意力模块。在一般微调策略中，考虑到自注意力模块和前馈层是预训练语言模型所积累的大量信息库，选择冻结自注意力模块，同时我们有选择性地对Add层和Norm层进行微调，防止过度拟合小样本数据，保持模型的泛化能力。在进行试验后，我们选择加入LoRA微调策略，达到更好的微调效果。在实证测试中，我们使用GPT2, Llama3-1B, Qwen-2.5B作为基础模型。不过值得注意的是，我们的架构非常灵活，可以将其他大型语言模型（LLM）作为替代。

3 提出的使用预训练大语言模型

A. 模型结构

我们提出了一种利用GPT-2[36] 、LLaMA[37] 和Qwen[38]等预训练的大型语言模型(LLM)进行吞咽信号分类的新方法。我们的模型利用了这些LLM强大的表征学习能力，针对固定长度的一维时间序列信号分类这一特定任务对其进行了微调。其核心创新在于通过嵌入层、位置编码和分类头的组合，使这些LLMs适应吞咽信号时间序列数据的处理。



模型架构如图所示，由以下关键部分组成：

· 嵌入层。输入的时间序列信号首先经过卷积嵌入层，将一维信号投射到与预训练LLM的输入维度相匹配的高维空间。该嵌入层由两个一维卷积层组成，采用 ReLU 激活和剔除进行正则化。卷积层的输出可以表示为：

· 位置编码(可选)。由于LLM是针对顺序文本设计的，因此我们加入了位置编码为模型提供有关时间序列数据时间顺序的信息。位置编码公式如下：

其中是时间序列中的位置索引，是维度索引，是模型的嵌入维度

· 预训练的LLM backbone。我们使用预训练的LLM模型(GPT-2、Llama-3.1、Qwen-2.5)作为架构的backbone，我们通过微调这些模型，使其适应吞咽信号分类任务。为了降低计算复杂度，我们选择合适的层数，保留了其捕捉数据中长距离依赖关系的能力。

· LoRA微调。为了有效微调预训练的LLM模型，我们采用了低层自适应(Low-Rank Adaptation，LoRA)技术。LoRA将低秩矩阵引入注意力层，使模型适应新任务，在保持模型性能的同时大大减少了可训练参数的数量。假设原始权重矩阵为W，LoRA 将其分解为两个低秩矩阵A和B：

其中是预训练模型的原始权重矩阵，和是低秩矩阵，，是低秩适应矩阵。

· LSTM层。将时间序列数据通过LLM backbone后，我们使用双向LSTM层进一步捕捉数据中的时间依赖性。LSTM 层输出一系列隐藏状态，然后用于分类。LSTM的更新公式可以表示为：

其中是遗忘门，是输入门，是候选记忆单元，是当前记忆单元，是输出门，是隐藏状态，是sigmoid激活函数，是双曲正切激活函数，是权重矩阵，是偏置项。

· 分类头。最终分类由线性层完成，该层获取 LSTM 在第一个时间步的隐藏状态，并将其映射到分类任务类别的数量。分类头的输出可以表示为：

其中是分类层的权重，是偏置项，是激活函数，用于将输出转换为概率分布。

B. 前向过程

给定输入时间序列信号 x 的形状为 (B, L, D)，B为batch size, L为序列长度，D为序列为维度，模型的前向过程可概述如下：

· 嵌入。输入信号经过嵌入层，得到形状为(B, L, )的高维表示，其中 是LLM backbone的嵌入维数。

· 位置编码。位置编码添加到中，加入序列位置信息，得到。

· LLM backbone。嵌入并加入位置编码的序列将通过预训练的 LLM 模型，输出隐藏状态层表示，其形状为(B, L, )。

· LSTM 层。隐藏状态层表示通过双向LSTM层，该层输出形状为(B, L, )的隐藏状态序列(假设每个方向的隐藏大小为)。

· 分类。取第一个时间步长的隐藏状态通过线性层，得到最终的分类结果。

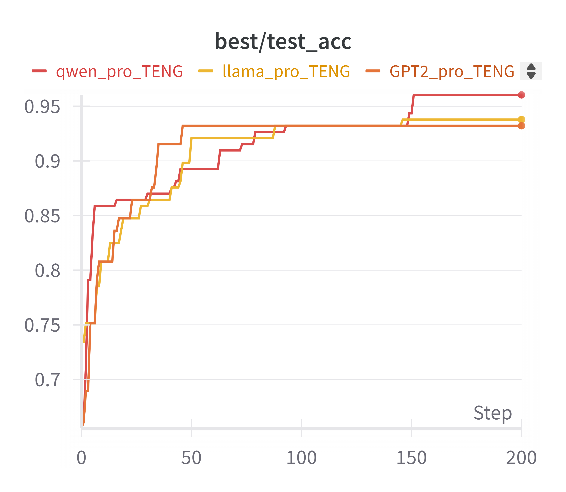
C. 损失公式

我们的模型使用交叉熵损失进行训练，交叉熵损失用于衡量预测的类别概率与地面实况标签之间的差异。损失的计算方法如下：

其中，是样本数，是类别数，是样本和类别的基本真实标签，是样本和类别的预测概率。

D. 训练

我们使用带有学习率调度器的 AdamW 优化器进行训练，初始学习率为0.01，权重衰减为1e-4，批量大小设置为16。在训练过程中，模型在吞咽信号数据集上进行微调，LoRA指定的参数和分类层是可训练的主要部分。训练过程包括在多个历时中最小化交叉熵损失，并在验证集上评估模型的性能，以防止过拟合。训练过程如图所示。



4 Experiment

A. 数据集

我们在多个吞咽信号数据集上评估了微调预训练LLM模型的性能，包括摩擦电动作、电阻抗动作、饮水、吃东西等数据集。这些数据集涵盖了不同的吞咽相关动作，每个数据集的时间序列信号经过预处理，确保长度和格式的一致性。数据集的具体信息如下：

·摩擦电动作数据集：包含低头、咳嗽、干咽、咀嚼、打嗝动作的时间序列信号，信号通过摩擦电传感器采集。

·电阻抗动作数据集：通过电阻抗传感器采集的低头、咳嗽、干咽、咀嚼、打嗝动作信号，信号反映了吞咽过程中喉部的阻抗变化。

·饮水数据集：记录了3种饮水量的吞咽信号。

·吃东西数据集：包含吃固体、液体等3种类型食品的过程中的吞咽信号。

1. Baseline和实验设置。

我们将微调预训练LLM模型与几种传统机器学习模型进行了比较，包括决策树、KNN、逻辑回归、随机森林和 SVM，以及与深度学习模型进行比较，包括LSTM、GRU、CNN(三层CNN架构)。

* 决策树（Decision Tree）[39]：一种树形结构的分类模型，通过递归地选择最优特征进行数据分割
* K近邻（KNN, K-Nearest Neighbors）[40]：基于距离进行分类，通过计算新样本与训练样本的距离，将其分类为最近K个邻居中的多数类别
* 逻辑回归（Logistic Regression）[41]：线性分类，通过Sigmoid函数将线性组合的特征映射为概率值
* 随机森林（Random Forest）[42]：一种基于决策树的集成分类模型，通过构建多棵树并综合其预测结果
* 支持向量机（SVM, Support Vector Machine）[43]：一种基于间隔最大化的分类模型，通过寻找最优超平面将不同类别的样本分开
* 长短期记忆网络（LSTM, Long Short-Term Memory）[44]：一种适用于时间序列数据的分类模型，通过记忆单元和门控机制捕捉长期依赖关系
* 门控循环单元（GRU, Gated Recurrent Unit）[45]：一种简化版的LSTM分类模型，通过减少门控单元数量，在保持时序建模能力的同时提高计算效率
* 卷积神经网络（CNN, Convolutional Neural Network）[46]：一种适用于图像和信号数据的分类模型，通过卷积层提取局部特征，池化层降低维度

所有模型都使用相同的实验设置进行训练和评估，序列长度和批量大小固定不变。我们使用准确率作为评估分类性能的主要指标。

1. 实验

我们提出了基于微调预训练LLM模型的三个变体，每个变体都使用不同的微调预训练LLM骨干：GPT-2、Llama3-1B 和 Qwen-2.5B。所有变体共享相同的架构，包括卷积嵌入层、位置编码、LoRA 微调、双向 LSTM 和分类头。这些模型在吞咽信号数据集上进行了微调，并将结果与选取的baseline进行了比较。

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | actions | | | actions-im | |  | Pronunciation | |
|  |  | dataset-1 | dataset-2 | dataset-3 | dataset-1 | dataset-2 | eating | TENG | Im |
| Machine Learning | Decision Tree | 0.787 | 0.766 | 0.778 | 0.875 | 0.800 | 0.966 | 0.859 | 0.867 |
| KNN | 0.833 | 0.757 | 0.855 | 0.929 | 0.800 | 0.871 | 0.859 | 0.875 |
| Logistic | 0.713 | 0.701 | 0.667 | 0.732 | 0.691 | 0.707 | 0.785 | 0.836 |
| Random  Forest | 0.880 | 0.850 | 0.872 | 0.893 | 0.836 | 0.948 | 0.876 | 0.898 |
| SVM | 0.676 | 0.710 | 0.667 | 0.750 | 0.782 | 0.759 | 0.808 | 0.828 |
| Deep Learning | LSTM | 0.713 | 0.925 | 0.906 | 0.821 | 0.927 | **0.974** | 0.972 | 0.984 |
| GRU | 0.880 | 0.916 | 0.923 | 0.804 | 0.927 | 0.957 | **0.977** | **1.000** |
| CNN | **0.926** | **0.944** | **0.949** | 0.875 | 0.982 | 0.897 | **0.977** | 0.984 |
| Pretrained-LLM | GPT2 | 0.907 | 0.925 | 0.915 | **0.946** | 0.945 | 0.966 | 0.955 | 0.977 |
| Llama3 | 0.889 | **0.944** | **0.949** | 0.857 | **1.000** | **0.974** | 0.972 | **1.000** |
| Qwen | 0.917 | **0.944** | **0.949** | **0.946** | **1.000** | **0.974** | 0.972 | **1.000** |

表1中列出了基于微调预训练LLM的模型和基线模型在不同数据集上的分类准确率。总体来说，微调预训练LLM的模型在大多数情况下都优于传统机器学习和深度学习基线模型：

·Qwen-2.5B：在8个数据集中，Qwen-2.5B在6个数据集中取得了最高的准确率，2个数据集中排名第二。如在其中一个电阻抗传感器采集的吞咽动作数据集中，Qwen-2.5B达到了100%的准确率，显著优于其他模型。

·Llama-3.1B：在多个数据集中表现优异，尤其是在发音数据集中，Llama-3.1B在摩擦电传感器和电阻抗传感器数据集上分别取得了97.2%和100%的准确率。

·GPT-2：虽然在某些数据集上表现略逊于Qwen和Llama，但在摩擦电传感器采集的吞咽动作数据集中，GPT-2达到了94.6%的准确率，表现突出。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ACC ↑ | | | | |
|  | actions | actions-im | eating | pronunciation | **AVG** |
| Machine Learning | 15.91 | 14.02 | 12.11 | 13.02 | 13.8 |
| RNN | 4.94 | 7.92 | 0.58 | -0.39 | 3.3 |
| CNN | -1.31 | 2.05 | 7.43 | -0.12 | 2.0 |

表2展示了在不同类型数据集中，相较于Machine Learning, RNN, CNN model，LLM4SG框架下平均准确率提升。

与传统的机器学习模型相比，基于微调预训练LLM的模型准确率有了显著提高，大幅度领先于大部分机器学习模型，准确率平均提高13.8%。对于深度学习模型，在RNN中，准确率平均提高3.3%，相较于LLM，LSTM和GRU在一些数据集上表现突出，但通用性较差，并不能在大多数数据集中表现出较好的结果；CNN在深度学习模型中表现最好，体现了卷积神经网络的信息捕捉能力，在一些数据集上也取得了最优结果，LLM准确率平均提高2.0%，但同样通用性较弱。同时，三种轻量级LLM的实验结果也体现了LLM4SG框架的泛化性。

D. Few-Shot实验

大语言模型在Few-Shot和Zero-Shot任务中表现出卓越的能力，为了评估预训练大语言模型在广泛的信号分类任务中的能力，我们进行了实验。

与传统的实验设置类似，每个信号数据集被分成三部分：训练数据、验证数据和测试数据。对于少次学习，只使用一定比例（10%）的训练数据。

10%少量学习的结果如表所示。我们选取GPT2模型进行实验，与RNN-based模型LSTM, GRU和CNN相比，GPT2的性能最好。传统上，基于CNN的模型被认为在训练时数据效率更高，适用于少量学习的方法。与基于卷积的三层CNN和基于RNN的LSTM, GRU相比GPT2的平均准确率分别提升了11.5%和39%。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | actions | | |  | pronunciation |
|  | dataset-1 | dataset-2 | dataset-3 | eating | TENG |
| LSTM | 0.676 | 0.748 | 0.729 | 0.724 | 0.785 |
| GRU | 0.657 | 0.748 | 0.766 | 0.776 | 0.797 |
| CNN | 0.704 | **0.841** | **0.832** | 0.759 | 0.842 |
| GPT2 | **0.815** | 0.813 | 0.804 | **0.802** | **0.859** |

E. Zero-Shot实验

这项任务用于评估我们提出的算法的跨数据集适应能力，即当一个模型在数据集A（没有来自B的任何训练数据）上进行训练时，它在数据集B上的表现如何。

我们选取两个吞咽动作和两个发音数据集进行实验，对于每一种数据集，在其中一个数据集上进行训练，在另一个数据集上进行测试，结果如表所示。我们选取了基于RNN的LSTM和三层CNN模型进行比较，结果显示三种预训练大语言模型表现普遍好于以上两种模型，尤其是GPT2模型。与LSTM和CNN模型相比，预训练大语言模型的平均准确率分别提升了13.3%和2%。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | action | | pronunciation | |
|  | dataset-2 | dataset-3 | TENG | im |
| LSTM | 0.860 | 0.795 | 0.876 | 0.797 |
| CNN | 0.897 | 0.932 | 0.966 | 0.977 |
| GPT2 | **0.963** | **0.966** | **0.994** | **0.984** |
| Llama3 | 0.944 | 0.940 | 0.989 | 0.938 |
| Qwen | **0.963** | 0.940 | 0.983 | 0.977 |

F. 消融实验

LoRA微调的影响。我们进行了针对LoRA的消融实验，评估LoRA微调对模型性能的影响。我们比较了使用LoRA微调的模型和未使用LoRA微调的模型的性能。结果如表 2所示，LoRA 微调显著提高了模型的准确性，同时减少了可训练参数的数量。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | actions-dataset2 | | | actions-dataset3 | | |
| Lora | None\_Lora | ACC↑ | Lora | None\_Lora | ACC↑ |
| GPT2 | 0.925 | 0.888 | 4.2% | 0.897 | 0.915 | - 2.0% |
| Llama3 | 0.944 | 0.936 | 0.9% | 0.949 | 0.932 | 1.8% |
| Qwen | 0.944 | 0.925 | 2.1% | 0.949 | 0.923 | 2.8% |

从表格中结果可以看出，在大多数情况下，使用LoRA微调的模型表现优于未使用LoRA微调的模型，我们选取actions-dataset2和actions-dataset3数据集进行实验，实验表明使用LoRA后准确率提升了0.9%到4.2%不等。尽管在某些数据集上未使用LoRA微调的模型表现略好，但整体来看，LoRA微调在减少计算成本的同时，依然能够保持较高的分类准确率。

双向 LSTM 的影响。我们还评估了双向 LSTM 层对模型性能的影响。我们比较了有 LSTM 层和无 LSTM 层模型的性能。结果如表3所示，表明双向LSTM层对于捕捉吞咽信号中的时间依赖性至关重要，从而提高了分类准确性。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | actions-dataset2 | | | actions-dataset3 | | |
|  | Lora | None\_LSTM | ACC↑ | Lora | None\_LSTM | ACC↑ |
| GPT2 | 0.925 | 0.794 | 16.5% | 0.897 | 0.795 | 12.8% |
| Llama3 | 0.944 | 0.748 | 26.2% | 0.949 | 0.803 | 18.2% |
| Qwen | 0.944 | 0.804 | 17.4% | 0.949 | 0.803 | 18.2% |

从表格中结果可以看出，双向LSTM层对模型性能的提升非常显著。在两个数据集的实验中，包含双向LSTM层的模型准确率均大幅高于未使用双向LSTM层的模型，准确率提升了12.8%到26.2%不等。这表明，双向LSTM层在捕捉吞咽信号中的时间依赖性方面起到了关键作用，尤其是在处理复杂的时序信号时，能够有效提升分类性能。

5 Conclusion

本文聚焦于喉部生理活动分类，提出了一种基于微调预训练大语言模型（LLMs）的通用模型。在模型构建上，设计 LLM-Backbone，利用卷积嵌入层将预训练 LLMs 应用于一维时间序列的吞咽信号数据处理，使其能够有效提取信号特征。同时，结合低秩适应（LoRA）微调技术，在减少可训练参数数量、降低计算成本的情况下，显著提高了模型训练效率；加入双向 LSTM 层，成功捕捉了吞咽信号中的时间依赖关系，进一步提升了模型对复杂时序信号的处理能力。实验结果充分验证了该架构的优越性，Zero-Shot和Few-Shot实验中也体现了模型良好的跨数据集适应能力。

Reference

[1] Clavé P, Shaker R. Dysphagia: current reality and scope of the problem. *Nature Reviews Gastroenterology & Hepatology* **12**, 259-270 (2015).

[2] Waters AM, Patterson J, Bhat P, Phillips AW. Investigating dysphagia in adults. *Bmj*, (2022).

[3] Cubero L*, et al.* Automated dysphagia characterization in head and neck cancer patients using videofluoroscopic swallowing studies. *Computers in Biology and Medicine* **187**, (2025).

[4] Badenduck LA*, et al.* Fiber-optic endoscopic evaluation of swallowing to assess swallowing outcomes as a function of head position in a normal population. *Journal of Otolaryngology-Head & Neck Surgery* **43**, (2014).

[5] Ryu JS, Park D, Kang JY. Application and Interpretation of High-resolution Manometry for Pharyngeal Dysphagia. *Journal of Neurogastroenterology and Motility* **21**, 283-287 (2015).

[6] Trapl M*, et al.* Dysphagia Bedside Screening for Acute-Stroke Patients. *Stroke* **38**, 2948-2952 (2007).

[7] Chen PC, Chuang CH, Leong CP, Guo SE, Hsin YJ. Systematic review and meta‐analysis of the diagnostic accuracy of the water swallow test for screening aspiration in stroke patients. *Journal of Advanced Nursing* **72**, 2575-2586 (2016).

[8] Kim MK*, et al.* Flexible submental sensor patch with remote monitoring controls for management of oropharyngeal swallowing disorders. *Science Advances* **5**, (2019).

[9] Zhang D*, et al.* Stretchable and durable HD-sEMG electrodes for accurate recognition of swallowing activities on complex epidermal surfaces. *Microsystems & Nanoengineering* **9**, (2023).

[10] Wong F, de la Fuente-Nunez C, Collins JJ. Leveraging artificial intelligence in the fight against infectious diseases. *Science* **381**, 164-170 (2023).

[11] Shelmerdine SC, Martin H, Shirodkar K, Shamshuddin S, Weir-McCall JR. Can artificial intelligence pass the Fellowship of the Royal College of Radiologists examination? Multi-reader diagnostic accuracy study. *Bmj*, (2022).

[12] Taunk K, De S, Verma S, Swetapadma A. A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. In: *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*) (2019).

[13] Miyagi S, Sugiyama S, Kozawa K, Moritani S, Sakamoto SI, Sakai O. Classifying Dysphagic Swallowing Sounds with Support Vector Machines. Healthcare (Basel). 2020 Apr 21;8(2):103. doi: 10.3390/healthcare8020103. PMID: 32326267; PMCID: PMC7349358.  
[14] Liu H, Dong W, Li Y, et al. An epidermal sEMG tattoo-like patch as a new human–machine interface for patients with loss of voice[J]. Microsystems & nanoengineering, 2020, 6(1): 16.

[15] Yi X, Ma Y, Li Y, Xu H, Ma J. 人工智能促进复杂环境中的信息融合感知。*创新*, (2025).  
[16] Sleeman WC, Kapoor R, Ghosh P. Multimodal Classification: Current Landscape, Taxonomy and Future Directions. *ACM Comput Surv* **55**, Article 150 (2022).

[17] Hou B, Yang D, Ren X, Yi L, Liu X. A tactile oral pad based on carbon nanotubes for multimodal haptic interaction. *Nature Electronics* **7**, 777-787 (2024).

[18] Yoshua B. *Learning Deep Architectures for AI*. now (2009).

[19] Yang Q, Jin W, Zhang Q, et al. Mixed-modality speech recognition and interaction using a wearable artificial throat[J]. Nature Machine Intelligence, 2023, 5(2): 169-180.

[20] Song, Y., Yun, I., Giovanoli, S. et al. Multimodal deep ensemble classification system with wearable vibration sensor for detecting throat-related events. npj Digit. Med. 8, 14 (2025).

[21] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks[C]//International conference on machine learning. Pmlr, 2013: 1310-1318.

[22] Hou, B., Yang, D., Ren, X. et al. A tactile oral pad based on carbon nanotubes for multimodal haptic interaction. Nat Electron **7**, 777–787 (2024). <https://doi.org/10.1038/s41928-024-01234-9.>

[23] Marcus GF. Deep Learning: A Critical Appraisal. *ArXiv* **abs/1801.00631**, (2018).

[24] Xue Jiang et al., "Self-supervised Contrastive Learning for EEG-based Sleep Staging," arXiv:2109.07839v1, 2021

[25] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.

[26] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[27] Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, Haifeng Li. EEGPT: Pretrained Transformer for Universal and Reliable Representation of EEG Signals. NeurIPS Conference (2024)

[28] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019: 4171-4186.

[29] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018.

[30] Lee, Jaejun, Raphael Tang, and Jimmy Lin. "What would elsa do? freezing layers during transformer fine-tuning." *arXiv preprint arXiv:1911.03090* (2019).

[31] Liu, Yuhan, Saurabh Agarwal, and Shivaram Venkataraman. "Autofreeze: Automatically freezing model blocks to accelerate fine-tuning." *arXiv preprint arXiv:2102.01386* (2021).

[32] Houlsby, Neil, et al. "Parameter-efficient transfer learning for NLP." *International conference on machine learning*. PMLR, 2019.

[33] Li, Xiang Lisa, and Percy Liang. "Prefix-tuning: Optimizing continuous prompts for generation." *arXiv preprint arXiv:2101.00190* (2021).

[34] Dettmers T, Pagnoni A, Holtzman A, et al. Qlora: Efficient finetuning of quantized llms[J]. Advances in neural information processing systems, 2023, 36: 10088-10115.

[35] Zhang, Dan, et al. "Parameter-Efficient Fine-Tuning for Foundation Models." *arXiv preprint arXiv:2501.13787* (2025).

[36] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8): 9.

[37] Touvron H, Lavril T, Izacard G, et al. Llama: Open and efficient foundation language models[J]. arXiv preprint arXiv:2302.13971, 2023.

[38] Ahmed I, Islam S, Datta P P, et al. Qwen 2.5: A Comprehensive Review of the Leading Resource-Efficient LLM with potentioal to Surpass All Competitors[J].

[39] Quinlan J R. Induction of decision trees[J]. Machine learning, 1986, 1: 81-106.

[40] Li R, Li J, Zhou M, et al. Learning-Based Distributed Spatio-Temporal $ k $ Nearest Neighbors Join[J]. IEEE Transactions on Big Data, 2024.

[41] Kleinbaum D G, Dietz K, Gail M, et al. Logistic regression[M]. New York: Springer-Verlag, 2002.

[42] Breiman L. Random forests[J]. Machine learning, 2001, 45: 5-32.

[43] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers[C]//Proceedings of the fifth annual workshop on Computational learning theory. 1992: 144-152.

[44] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[45] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.

[46] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural inform

cessing systems, 2012, 25.