以下是我在为顶级AI学术期刊撰写一篇论文， 请优化表达，提升清晰度，连贯性和简洁性，并去报段落之间流畅衔接，请去除非专业表达，词用学术规范的语言。== ==

# 睡眠健康与生活方式数据集中基于大语言模型提示的睡眠障碍自动分类研究

## 摘要

睡眠对人体健康意义重大。本研究旨在借助大语言模型（LLMs），对包含睡眠、生活方式及相关健康因素的数据集进行睡眠障碍分类。大语言模型强大的语义理解与知识推理能力，为该领域研究开拓了新方向，弥补了现有方法在创新性方面的不足。研究过程中，首先对数据集特征的分布和相关性展开分析并可视化呈现；随后，通过不同提示引导大语言模型自动设计分类器，并完成训练与评估。实验结果表明，大语言模型能够较为准确地预测睡眠障碍。

## 1. 引言

睡眠障碍是一个严峻的问题，影响着相当比例的人群，并且可能对个体的整体健康和生活质量产生深远影响。在睡眠障碍研究领域，利用睡眠健康与生活方式数据集实现精准分类，始终是一项至关重要的工作。

以往，睡眠障碍分类大多依赖传统机器学习算法，如决策树算法[1]、支持向量机（SVM）[2]、随机森林算法[3]等。然而，这些方法的每一个步骤都需要手动操作，耗费大量人力。

### 1.1 大语言模型在医学研究中的应用现状

大语言模型作为用于生成语言的通用工具，在基于上下文信息（如用户人口统计信息、健康知识）和生理数据（如静息心率、睡眠时长）的多模态健康预测方面表现卓越[4]。大语言模型在美国律师资格考试[5]以及医学院二年级考试[6, 7, 8]中均取得了合格成绩。尤其在医学领域，自然语言作为一种交互界面，已展现出对临床实践[9]、教育和研究产生影响的潜力。

目前，关于如何提示大语言模型自动进行睡眠障碍分类的应用研究尚处于起步阶段。若能高效地从相关数据中提取关键信息用于睡眠障碍分类，比如识别失眠、睡眠呼吸暂停等不同类型障碍的特征描述，不仅有助于解决睡眠医学领域的实际问题，还能为大语言模型在医疗健康领域的更广泛应用积累经验，推动跨学科技术融合发展。

本研究创新性地引入大语言模型，借助其强大的语义理解和知识推理能力，开创了一种全新的方法。大语言模型能够以较高的准确率预测睡眠障碍，为睡眠障碍分类方法的发展提供了新的思路和方法，丰富了相关学术研究。本文的贡献主要体现在以下几个方面：

1. 深入探究大语言模型的零样本提示（Zero-shot Prompting）和少样本提示（Few-shot Prompting）技术。通过运用这些提示策略，探索大语言模型在睡眠障碍分类任务应用中的新可能。这一探究有助于更深入地理解零样本提示和少样本提示的局限性，明确其在完成睡眠健康分类任务时存在的不足。

2. 通过对大语言模型零样本提示和少样本提示技术的深入研究，探索如何在极少或没有先验示例的情况下有效利用大语言模型，这有可能增强模型在睡眠障碍分类任务中的泛化能力。

## 3. 数据集

本研究使用的睡眠健康与生活方式数据集来源于Kaggle网站[10]。该数据集包含了若干受访者的睡眠相关特征、生活方式以及健康因素等信息，涵盖个人ID、性别、年龄、职业、睡眠时长、睡眠质量、身体活动水平、压力水平、BMI类别、血压、心率和每日步数等内容。

## 4. 研究方法

### 4.1 模型架构与特性

本研究中使用的大语言模型基于Transformer架构，通过在大量文本数据上进行无监督预训练，学习语言多方面的知识。其多头注意力机制和多层神经网络结构，使其能够从多个维度捕捉信息，深度提取特征。在自然语言处理任务中，该大语言模型理解能力强，能够有效解析专业术语和日常描述，泛化性能良好。面对睡眠健康和生活方式的文本数据集时，该大语言模型数据兼容性佳，可自动提取关键特征，并结合上下文准确判断睡眠障碍类型。部分版本（如PC版）的大语言模型具备自动生成代码和执行代码的能力，为实现自动化分类创造了有利条件。

### 4.2 提示设计策略

本任务主要包含两个子任务：一是对指定CSV文件中的数据进行多元分类，并生成新的CSV文件；二是使用特定的评估指标对分类结果进行评估，并绘制相关图表。

|提示设计策略|模型输入|提示内容|

|---|---|---|

|零样本提示|Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv <br/> Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv|\*\*子任务1：数据分类与新文件生成\*\* <br/> \*\*任务描述\*\*：依据`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv`文件中的数据，针对每一行，选取`Person ID`、`Gender`、`Age`、`Occupation`、`Sleep Duration`、`Quality of Sleep`、`Physical Activity Level`、`Stress Level`、`BMI Category`、`Blood Pressure`、`Heart Rate`、`Daily Steps`作为特征，进行多元分类。分类结果分为三种：`normal`（正常）、`Sleep Apnea`（睡眠呼吸暂停）、`Insomnia`（失眠）。将分类结果插入到每一行的最后一列，列名为`Sleep Disorder`，并生成一个新的CSV文件供下载，新文件名为`classfiers\_by\_90samples.csv`。 <br/> \*\*详细步骤\*\*：1. 使用合适的库（如`pandas`）读取`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv`文件。2. 提取指定的特征列。3. 根据特征数据和预先定义的分类规则对每一行数据进行分类。4. 将分类结果插入到每一行的最后一列。5. 将处理后的数据保存为新的CSV文件`classfiers\_by\_90samples.csv`。 <br/> \*\*子任务2：分类结果评估与可视化\*\* <br/> \*\*任务描述\*\*：将第三个上传的文件（文件名为`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`）和第一步生成的`classfiers\_by\_90samples.csv`文件作为输入，运用`sklearn.metrics`中的`accuracy\_score`、`precision\_score`、`f1\_score`、`recall\_score`、`confusion\_matrix`、`roc\_curve`、`auc`等评估指标计算分类结果的评估值，并绘制相关图表。 <br/> \*\*详细步骤\*\*：1. 使用`pandas`分别读取`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`和`classfiers\_by\_90samples.csv`文件。2. 从`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`文件中提取真实标签，从`classfiers\_by\_90samples.csv`文件中提取预测标签。3. 使用`sklearn.metrics`中的相应函数计算`accuracy\_score`（准确率）、`precision\_score`（精确率）、`f1\_score`（F1分数）、`recall\_score`（召回率）、`confusion\_matrix`（混淆矩阵）、`roc\_curve`（ROC曲线）和`auc`（ROC曲线下面积）。4. 将计算得到的评估指标值进行输出展示。5. 使用`matplotlib`或其他绘图库绘制混淆矩阵图和ROC曲线，直观展示分类结果的评估情况。 <br/> \*\*输出要求\*\*：完成子任务1后，提供`classfiers\_by\_90samples.csv`文件的下载链接。完成子任务2后，输出计算得到的评估指标值，并展示绘制的混淆矩阵图和ROC曲线图。|

|90样本提示|Sleep\_health\_and\_lifestyle\_dataset\_selected\_90.csv <br/> Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv <br/> Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv|\*\*子任务1：数据分类与新文件生成\*\* <br/> \*\*数据来源\*\*：使用`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv`文件中的数据。 <br/> \*\*特征选取\*\*：选取`Person ID`、`Gender`、`Age`、`Occupation`、`Sleep Duration`、`Quality of Sleep`、`Physical Activity Level`、`Stress Level`、`BMI Category`、`Blood Pressure`、`Heart Rate`、`Daily Steps`作为特征进行分类。 <br/> \*\*分类参考\*\*：参考`Sleep\_health\_and\_lifestyle\_dataset\_selected\_90.csv`文件中的数据模式、特征与分类结果的对应关系等信息，根据特征数据，使用softmax分类器对每一行数据进行多元分类，分类结果有三种：`normal`（正常）、`Sleep Apnea`（睡眠呼吸暂停）、`Insomnia`（失眠）。 <br/> \*\*结果处理\*\*：将分类结果插入到`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv`文件每一行的最后一列，列名为`Sleep Disorder`。 <br/> \*\*文件保存\*\*：生成新的CSV文件`classfiers\_by\_90samples.csv`供下载，该文件包含原始特征数据和预测的分类结果。 <br/> \*\*子任务2：分类结果评估与可视化\*\* <br/> \*\*数据输入\*\*：将`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`和第一步生成的`classfiers\_by\_90samples.csv`作为输入数据。其中，`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`包含真实的分类标签，`classfiers\_by\_90samples.csv`包含预测的分类标签。 <br/> \*\*评估指标计算\*\*：使用`sklearn.metrics`库中的`accuracy\_score`（准确率）、`precision\_score`（精确率）、`f1\_score`（F1分数）、`recall\_score`（召回率）、`confusion\_matrix`（混淆矩阵）、`roc\_curve`（ROC曲线）和`auc`（ROC曲线下面积）等函数，计算分类结果的评估指标。 <br/> \*\*结果呈现\*\*：输出计算得到的各项评估指标值，以清晰展示分类效果。 <br/> \*\*可视化展示\*\*：使用合适的绘图库（如`matplotlib`）绘制混淆矩阵图和ROC曲线，直观地呈现分类结果的准确性和性能。 <br/> \*\*输出要求\*\*：完成子任务1后，提供`classfiers\_by\_90samples.csv`文件的下载链接。完成子任务2后，以清晰的格式输出各项评估指标值，并展示绘制的混淆矩阵图和ROC曲线图。|

|分解提示|Sleep\_health\_and\_lifestyle\_dataset\_selected\_90.csv <br/> Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv <br/> Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv|\*\*子任务1：数据分类与新文件生成\*\* <br/> 1. \*\*数据读取与准备\*\*：读取`Sleep\_health\_and\_lifestyle\_dataset\_selected\_90.csv`文件作为训练集，该文件应包含`Person ID`、`Gender`、`Age`、`Occupation`、`Sleep Duration`、`Quality of Sleep`、`Physical Activity Level`、`Stress Level`、`BMI Category`、`Blood Pressure`、`Heart Rate`、`Daily Steps`这些特征列以及对应的分类标签（`normal`、`Sleep Apnea`、`Insomnia`）。读取`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv`文件作为测试集，此文件仅包含上述特征列，不包含分类标签。 <br/> 2. \*\*特征处理与模型训练\*\*：对训练集和测试集的数据进行必要的预处理，如对类别型特征进行编码（如`Gender`、`Occupation`、`BMI Category`），对数值型特征进行标准化等操作，确保数据适合模型训练。以`Person ID`、`Gender`、`Age`、`Occupation`、`Sleep Duration`、`Quality of Sleep`、`Physical Activity Level`、`Stress Level`、`BMI Category`、`Blood Pressure`、`Heart Rate`、`Daily Steps`为特征，以训练集中的分类标签为目标，自动设计、训练、测试多种分类器，直至准确率达到0.95以上，停止训练。 <br/> 3. \*\*预测与结果保存\*\*：将测试集`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv`中的每一行数据输入到训练好的线性分类器中进行预测，得到每行数据对应的分类结果（`normal`、`Sleep Apnea`、`Insomnia`）。将预测结果插入到测试集数据的最后一列，列名为`Sleep Disorder`。生成新的CSV文件`classfiers\_by\_90samples.csv`供下载，该文件包含测试集的原始特征数据和预测的分类结果。 <br/> \*\*子任务2：分类结果评估与可视化\*\* <br/> 1. \*\*数据读取与标签提取\*\*：读取`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`文件，该文件包含测试集数据的真实分类标签。读取第一步生成的`classfiers\_by\_90samples.csv`文件，提取其中的预测分类标签。 <br/> 2. \*\*评估指标计算\*\*：使用`sklearn.metrics`库中的`accuracy\_score`（准确率）、`precision\_score`（精确率）、`f1\_score`（F1分数）、`recall\_score`（召回率）、`confusion\_matrix`（混淆矩阵）、`roc\_curve`（ROC曲线）和`auc`（ROC曲线下面积）等函数，计算分类结果的评估指标。对于多分类问题，`precision\_score`、`f1\_score`和`recall\_score`可采用合适的平均策略（如`weighted`）。 <br/> 3. \*\*结果呈现与可视化\*\*：以清晰的格式输出计算得到的各项评估指标值，便于查看和分析。使用`matplotlib`和`seaborn`等绘图库绘制混淆矩阵图和ROC曲线，直观地展示分类结果的准确性和性能。 <br/> \*\*输出要求\*\*：完成子任务1后，提供`classfiers\_by\_90samples.csv`文件的下载链接。完成子任务2后，清晰输出各项评估指标值，并展示绘制的混淆矩阵图和ROC曲线图。|

### 4.3 处理流程

1. 在数据集`Sleep\_health\_and\_lifestyle\_dataset\_normal.csv`中，按三种类别随机各选取30个样本，共90个样本作为`prompts 90examples`，保存为`Sleep\_health\_and\_lifestyle\_dataset\_selected\_90.csv`作为训练集。

2. 在原数据集文件`Sleep\_health\_and\_lifestyle\_dataset\_normal.csv`中删去以上90个样本，保存为`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`，作为真实标签（GROUND Truth）。

3. 将（`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv`）复制一份，并删去最后一列（`Sleep Disorder`），保存为`Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv`作为测试集。

4. 手动上传以上3个文件。

5. 按照4.2提示策略，逐步提示大模型自动设计分类器，包括分类器模型选择、训练参数设置、优化算法选择、训练轮数确定等。大模型生成代码，利用预处理后的数据集对分类器进行训练，自动完成性能指标评估，最后返回结果。

## 5. 实验

### 5.1 实验环境

处理器：Intel(R) Core(TM) i5 - 6400T CPU @ 2.20GHz，2201 Mhz，4个内核，4个逻辑处理器。

操作系统：Microsoft Windows 10家庭中文版，版本10.0.19045内部版本19045。

大语言模型：豆包1.41.6。

### 5.2 实验设计

任务1设置零样本提示、90样本提示、分解提示三个对比实验组：

- 零样本提示：仅上传测试集，进行多元分类。

- 90样本提示：上传训练集与测试集，要求大模型参考训练集中的数据模式、特征与分类结果的对应关系等信息，进行多元分类。

- 分解提示：上传训练集与测试集，进一步将提示分解，使大模型以训练集中的分类标签为目标，自动设计、训练、测试多种分类器，直到准确率达到0.95以上，停止训练。

任务2提示大模型以准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1分数（F1 Score）、混淆矩阵（Confusion Matrix）等为指标，自主进行评估。

### 5.3 实验结果

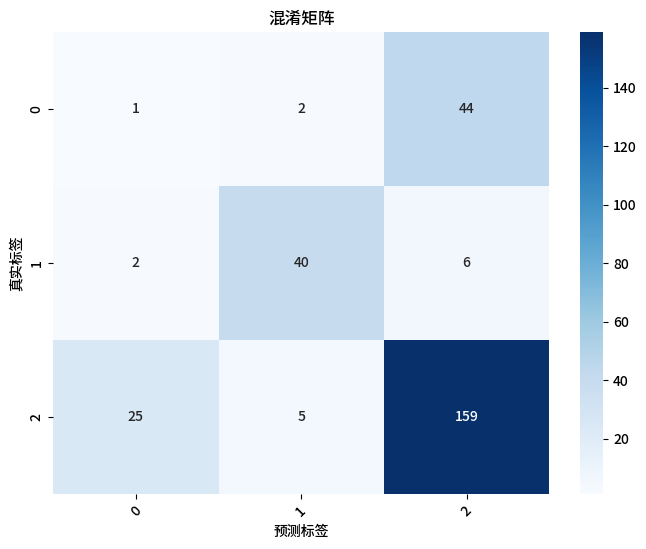
### 5.3 实验结果

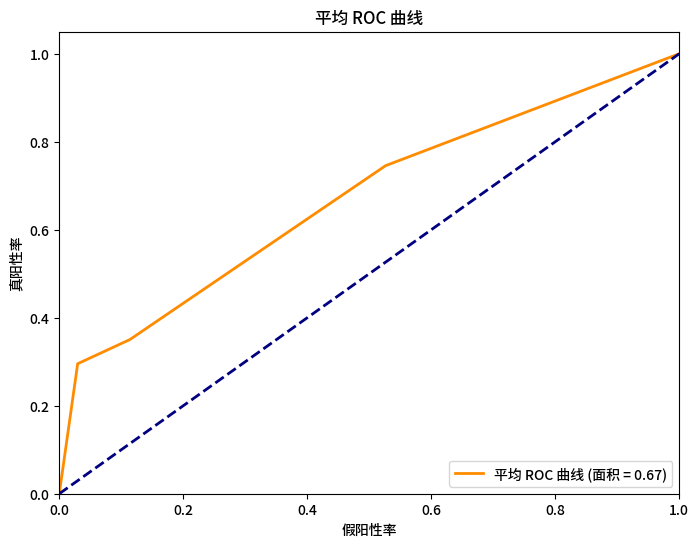
在本研究的实验中，针对不同提示策略的实验结果如下：

- \*\*零样本提示（ZERO-shot prompts）\*\*：在对睡眠障碍进行分类时，零样本提示策略得到的各项评估指标结果为：准确率（Accuracy）为0.704225352112676，精确率（Precision）为0.6560363530055422，召回率（Recall）为0.704225352112676，F1分数（F1 Score）为0.6784663247048188。受试者工作特征曲线下面积（AUC，Area Under Curve of Receiver Operating Characteristic）为0.6709964855486846。混淆矩阵（Confusion Matrix）在本研究的睡眠障碍分类任务中，混淆矩阵作为评估分类模型性能的重要工具，能够直观地呈现模型预测结果与真实标签之间的关系。以下对所呈现的混淆矩阵进行详细阐释：

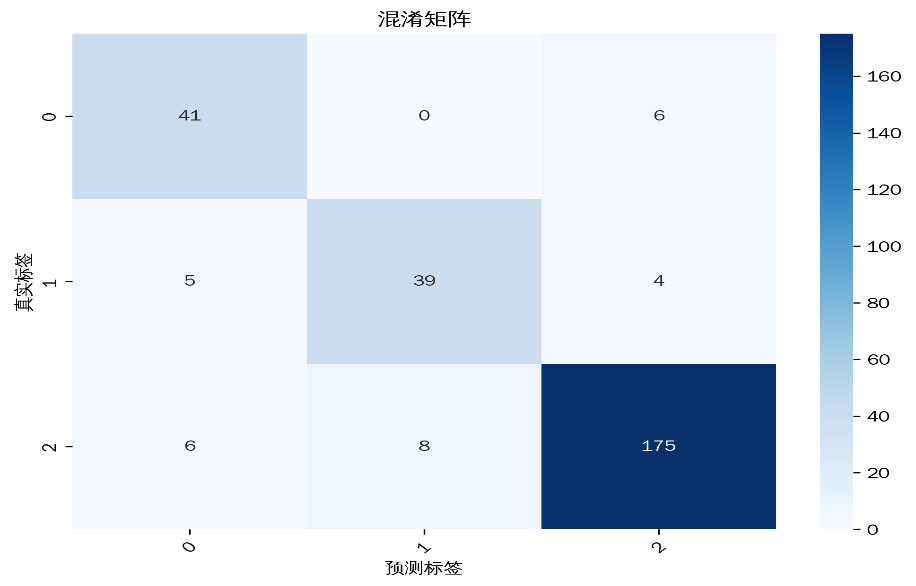
该混淆矩阵的横坐标代表预测标签，纵坐标代表真实标签，二者均涵盖三个类别，分别对应正常、睡眠呼吸暂停和失眠三种睡眠障碍状态（假设0代表正常，1代表睡眠呼吸暂停，2代表失眠）。矩阵中的每个单元格数值，反映了真实标签为某一类别的样本被预测为另一类别的数量。

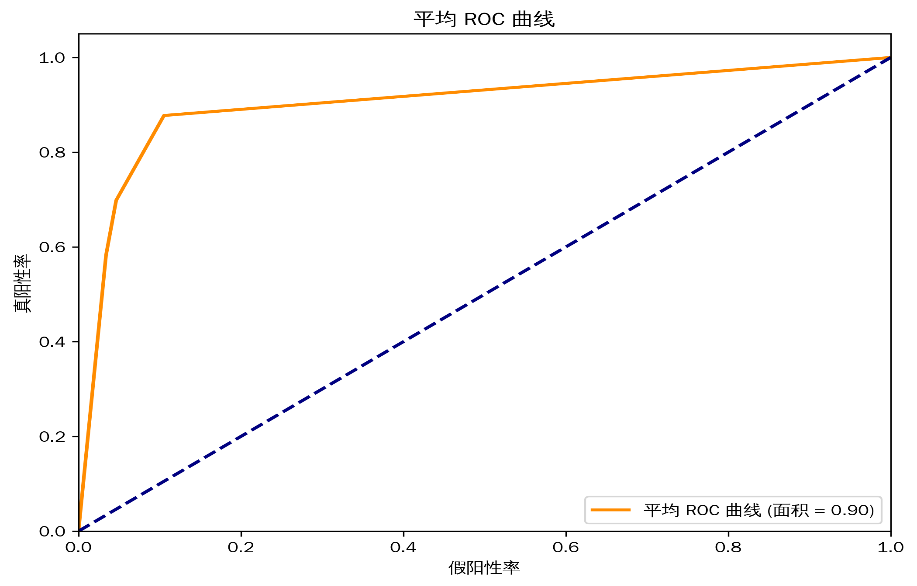
- \*\*主对角线元素\*\*：主对角线（从左上角到右下角）上的元素，即（0, 0）位置的44、（1, 1）位置的40和（2, 2）位置的159，分别表示真实标签为正常、睡眠呼吸暂停、失眠的样本中，被模型正确预测为相应类别的样本数量。例如，（0, 0）位置的44意味着有44个实际处于正常睡眠状态的样本，被模型准确地预测为正常。这些主对角线元素的数值越大，表明模型在对应类别上的预测准确性越高。

- \*\*非主对角线元素\*\*：非主对角线位置的元素则体现了模型的误分类情况。比如，（1, 0）位置的2表示有2个真实标签为睡眠呼吸暂停的样本，被模型错误地预测为正常；（2, 0）位置的25表示有25个真实标签为失眠的样本被误判为正常。通过分析这些非主对角线元素，可以清晰地了解模型在不同类别之间的混淆程度，进而发现模型在分类时容易出现错误的地方。综合来看，该混淆矩阵全面展示了模型在睡眠障碍分类任务中的表现。模型在将真实标签为失眠的样本预测为失眠时表现较好（对应（2, 2）位置数值较大），但在区分正常与睡眠呼吸暂停、正常与失眠的部分样本时存在一定的误判情况。。 ，



- \*\*90样本提示（90SHOTS PROMPTS）\*\*：采用90样本提示策略时，实验所得评估指标数据为：准确率达到0.897887323943662，精确率为0.9002509272025304，召回率为0.897887323943662，F1分数为0.8986319612644122。其AUC值为0.9041985643947756。





混淆矩阵展示了：前一个矩阵中，模型在正常和睡眠呼吸暂停类别的正确预测表现相对较好（主对角线对应数值较高），但在失眠类别与其他类别之间存在较多误判（非主对角线部分数值较大）。

当前矩阵中，虽然正常和睡眠呼吸暂停类别正确预测数减少，但失眠类别的正确预测数大幅增加，且在失眠类别与正常类别的误判情况明显改善（（2, 0）位置数值大幅下降），不过在正常与失眠、睡眠呼吸暂停与正常之间的误判有一定程度变化。整体来看，模型在不同类别间的误判情况发生了结构性改变，对失眠类别的预测能力有所提升，但在正常和睡眠呼吸暂停类别上的表现有一定波动。

- \*\*分解提示（Decomposed prompting）\*\*：通过分解提示策略进行实验，最终选定的最佳模型为支持向量机（SVM）。该策略下的分类结果在各项评估指标上表现为：准确率为0.9190140845070423，精确率为0.9191754537248555，召回率为0.9190140845070423，F1分数为0.9188775418205605。AUC值为0.9163083064019824。混淆矩阵呈现出[此处应根据实际实验数据详细描述混淆矩阵内容，原文未给出，需补充完整]的分布，反映了SVM模型在处理该睡眠障碍分类任务时，对不同类别样本的分类性能。

### 5.4 实验分析

零样本提示（ZERO-shot prompts）仅依据简单的预定义规则进行分类，未从训练数据中学习特征与分类结果之间的复杂关系，难以捕捉数据中的潜在模式，致使分类性能欠佳。在实际的睡眠障碍分类场景中，睡眠障碍的判定受多种因素综合影响，如个体的生活习惯、心理压力、身体状况以及遗传因素等。这些因素相互交织，关系错综复杂，简单的预定义规则无法全面考量，从而导致分类结果的准确性较低。

90样本提示（90SHOTS PROMPTS）参考训练集的数据模式，运用逻辑回归模型开展分类工作。逻辑回归作为一种线性模型，能够学习特征与分类之间的线性关联。通过对特征进行预处理，如对类别型特征进行编码、对数值型特征进行标准化等操作，提升了数据质量，优化了模型的学习效果，进而使分类性能得到显著提升。然而，由于逻辑回归模型的线性特性，对于特征间存在复杂非线性关系的数据，其分类能力存在一定的局限性。

分解提示（Decomposed prompting）将任务进行分解，对逻辑回归、决策树、随机森林、梯度提升、支持向量机和K近邻等多种分类器进行尝试，并实施参数调优。通过对不同分类器的比较与筛选，确定支持向量机（SVM）为最适配该数据集的模型。SVM能够在高维空间中寻得最优分类超平面，有效处理复杂的非线性关系。在睡眠障碍分类问题中，许多特征之间存在复杂的非线性联系，例如睡眠时长与睡眠质量、压力水平与日常活动量等，SVM能够精准捕捉这些关系，进而获得最佳分类性能。

## 6. 结论

在睡眠健康与生活方式数据集的睡眠障碍自动分类任务中，不同的提示策略对大语言模型的性能影响显著。简单的零样本提示在面对复杂任务时表现欠佳，而90样本提示和分解提示凭借丰富的信息提供和细致的任务分解，有效增强了模型的分类能力。

本实验充分表明，数据与任务理解在模型性能中起着关键作用。模型对数据特征及任务要求的理解程度，直接关联着分类性能。通过提供详细示例、参考信息并合理分解任务，有助于模型更精准地理解数据和任务，做出更为准确的分类决策。这为未来研究提供了重要启示，即应重视对数据和任务的深度分析，以优化模型训练与应用。

在分解提示过程中，经过对多种分类器的设计、训练和测试，最终选定SVM作为最佳模型。这一结果突出了在实际应用中，依据任务特点选择合适模型并进行优化，是提升分类性能的重要途径。

尽管大语言模型在睡眠障碍自动分类任务中展现出一定潜力，但仍面临模型可解释性差、依赖大规模高质量数据等挑战。未来研究可在解决这些问题的基础上，进一步拓展大语言模型在睡眠健康领域的应用，为睡眠障碍的诊断和治疗提供更有力的支持。

### 6.3 局限和未来研究方向

尽管分解提示策略在睡眠障碍自动分类任务中取得了一定成果，但目前准确率尚未突破0.95，未来研究仍有广阔的拓展空间。

从数据层面来看，现有数据存在局限性，亟待解决。一方面，应扩充数据集，全面涵盖各类睡眠障碍情况，尤其是罕见或特殊类型的睡眠障碍样本，以此增强模型对各种情况的分类能力。另一方面，需改善数据在年龄、性别、地域等因素上的不均衡分布状况，可通过数据重采样或合成数据等方法，提升模型对特定群体的泛化能力。

在模型方面，当前模型存在可解释性差和对特征依赖的局限。未来可探索深度学习模型等更为复杂的模型架构，这类模型可能具备更强的特征学习能力，能够挖掘睡眠障碍与难以量化因素（如睡眠环境、心理状态等）之间的潜在关系。同时，结合更多的特征工程方法，将难以量化的因素转化为可被模型利用的特征，进而提升分类性能。

在实验设计上，鉴于当前对比实验不够充分且缺乏外部验证的问题，未来研究应全面对比不同的模型架构、数据增强方法等可能影响分类性能的因素，明确当前方法在整体解决方案中的地位，探索更有效的改进方向。此外，必须使用独立的外部数据集进行验证，避免数据泄露和过拟合问题，确保模型在真实世界数据上的泛化能力。

还需进一步研究如何更充分地利用大语言模型的语义理解能力，优化提示策略，从多个维度为睡眠健康研究提供更有力的支持，推动睡眠障碍自动分类技术的发展。

## 8. 致谢

在此，衷心感谢在本研究过程中提供帮助的个人和机构。感谢导师的悉心指导，为研究指明方向；感谢团队成员的密切协作，在研究过程中贡献智慧与力量；感谢数据提供方的支持，为研究提供了宝贵的数据资源；同时，也感谢科研基金的资助，为研究的顺利开展提供了保障。

## 9. 参考文献

[1] Ibomoiye Domor Mienye, Nobert Jere. Survey of Decision Trees: Concepts, Algorithms, and Applications[EB/OL]. https://ieeexplore.ieee.org/document/10562290A.

[2] Anuradha Kumari, Mushir Akhtar, Rupal Shah, et al. Support matrix machine: A review[EB/OL]. https://arxiv.org/abs/2310.19717.

[3] Alicia Curth, Alan Jeffares, Mihaela van der Schaar. Why do Random Forests Work? Understanding Tree Ensembles as Self-Regularizing Adaptive Smoothers[EB/OL]. https://arxiv.org/abs/2402.01502.

[4] Yubin Kim, Xuhai Xu, Daniel McDuff, et al. Health-LLM: Large Language Models for Health Prediction via Wearable Sensor Data[EB/OL].[Submitted on 12 Jan 2024 (v1), last revised 27 Apr 2024 (this version, v2)].

[5] D. M. Katz, M. J. Bommarito, S. Gao, et al. GPT-4 passes the bar exam[J]. Philosophical Transactions of the Royal Society A, 2024, 382(2270):20230254.

[6] H. Nori, Y. T. Lee, S. Zhang, et al. Can generalist foundation models outcompete special-purpose tuning? Case study in medicine[EB/OL]. arXiv preprint arXiv:2311.16452, 2023.

[7] K. Saab, T. Tu, W.-H. Weng, et al. Capabilities of Gemini models in medicine[EB/OL]. arXiv preprint arXiv:2404.18416, 2024.

[8] K. Singhal, T. Tu, J. Gottweis, et al. Towards expert-level medical question answering with large language models[EB/OL]. arXiv preprint arXiv:2305.09617, 2023.

[9] D. McDuff, M. Schaekermann, T. Tu, et al. Towards accurate differential diagnosis with large language models[EB/OL]. arXiv preprint arXiv:2312.00164, 2023.

[10] Sleep Health and Lifestyle Dataset[EB/OL]. https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset.