逐行对照翻译以论文内容到英文,输出要求：逐段对照,一段中文，一段英文文====

基于大语言模型提示的睡眠健康与生活方式数据集中睡眠障碍自动分类研究Research on the Automatic Classification of Sleep Disorders in Sleep Health and Lifestyle Datasets Based on Large Language Model Prompting

**摘要Abstract**

睡眠障碍是一个紧迫的全球公共问题，影响着全球约 30% 的人口。Sleep disorders are an urgent global public issue, affecting approximately 30% of the global population.本研究利用LLMs 强大的语义理解与知识推理能力，对涵盖睡眠、生活方式及相关健康因素的数据集进行睡眠障碍分类。This study utilizes the powerful semantic understanding and knowledge reasoning capabilities of LLMs to classify sleep disorders in a dataset covering sleep, lifestyle, and related health factors.通过三种新颖的提示范式来引导LLMs 自动完成分类器的设计，训练与评估。Three novel prompting paradigms are employed to guide LLMs to automatically complete the design, training, and evaluation of classifiers.实验结果表明，使用分解提示自动找到的支持向量机分类器，分类准确率高达 91.9%（F1 - 分数：0.919），在准确率上显著优于传统的零样本提示和多样本提示方法。The experimental results show that the support vector machine classifier automatically found using decomposed prompting achieves a classification accuracy as high as 91.9% (F1 - score: 0.919), significantly outperforming traditional zero - shot prompting and few - shot prompting methods in terms of accuracy.这项研究独特地将大语言模型的语义理解知识推理能力与自动化机器学习无缝地融合，为健康信息学中的睡眠障碍分类提供了一种新范式。This study uniquely integrates the semantic understanding and knowledge reasoning capabilities of large language models with automated machine learning seamlessly, providing a new paradigm for the classification of sleep disorders in health informatics.

**关键词**：睡眠障碍分类；大语言模型；提示工程；健康信息学Keywords: Sleep Disorder Classification; Large Language Model; Prompt Engineering; Health Informatics

1. 引言Introduction

1.1 睡眠障碍问题的严峻性The Gravity of Sleep Disorder Issues

睡眠障碍是一个重大问题，影响着很大一部分人群，并可能对整体健康和生活质量产生重大影响。Sleep disorders are a significant problem, affecting a large portion of the population and potentially having a substantial impact on overall health and quality of life.根据世界卫生组织（WHO）的统计，全球约有10%的人口受到睡眠障碍的困扰，且这一比例在城市化进程加快的背景下呈上升趋势。According to statistics from the World Health Organization (WHO), approximately 10% of the global population is troubled by sleep disorders, and this proportion is on the rise against the backdrop of accelerated urbanization.睡眠障碍不仅会导致日间疲劳、注意力下降，还可能引发心血管疾病、糖尿病等慢性疾病。Sleep disorders can not only lead to daytime fatigue and decreased attention but also trigger chronic diseases such as cardiovascular diseases and diabetes因此，准确识别和分类睡眠障碍对于改善公众健康具有重要意义。Therefore, accurately identifying and classifying sleep disorders is of great significance for improving public health.

1.2 传统睡眠障碍分类方法的困境Dilemmas of Traditional Sleep Disorder Classification Methods

在过去，睡眠障碍分类主要依赖于传统机器学习算法，如决策树算法[1]、支持向量机（SVM）[2]、随机森林算法[3]等。In the past, the classification of sleep disorders mainly relied on traditional machine - learning algorithms, such as decision tree algorithms [1], support vector machines (SVM) [2], random forest algorithms [3], etc.这些算法在睡眠障碍研究中发挥了一定作用。These algorithms have played a certain role in sleep disorder research.研究人员利用决策树算法对睡眠监测数据进行分析，通过一系列的条件判断和分支决策，试图识别出不同睡眠障碍的特征模式 。Researchers used decision tree algorithms to analyze sleep monitoring data, attempting to identify the characteristic patterns of different sleep disorders through a series of conditional judgments and branch decisions.支持向量机则通过寻找一个最优的分类超平面，将正常睡眠数据和睡眠障碍数据区分开来。Support vector machines distinguish normal sleep data from sleep disorder data by finding an optimal classification hyperplane.随机森林算法通过构建多个决策树并综合它们的预测结果，提高了分类的准确性和稳定性 。The random forest algorithm improves the accuracy and stability of classification by constructing multiple decision trees and synthesizing their prediction results

然而，传统机器学习算法在睡眠障碍分类中存在诸多困境。The random forest algorithm improves the accuracy and stability of classification by constructing multiple decision trees and synthesizing their prediction results这些算法的每一个步骤都需要大量的手动操作。Every step of these algorithms requires a large amount of manual operation.在数据预处理阶段，研究人员需要手动处理数据缺失值、异常值，对数据进行标准化、归一化等操作，以确保数据的质量和可用性。In the data pre - processing stage, researchers need to manually handle missing values and outliers in the data, and perform operations such as standardization and normalization to ensure the quality and usability of the data.在模型训练和调优过程中，研究人员需要手动选择合适的算法、设置模型参数，并通过反复试验来优化模型性能 In the model training and tuning process, researchers need to manually select appropriate algorithms, set model parameters, and optimize the model performance through repeated trials.。

这种手动操作的方式不仅耗费大量的人力和时间成本，还容易受到人为因素的影响，导致结果的准确性和可靠性存在一定的局限性。This manual operation method not only consumes a large amount of human and time costs but is also easily affected by human factors, resulting in certain limitations in the accuracy and reliability of the results.而且，传统机器学习算法的应用高度依赖专家经验Moreover, the application of traditional machine - learning algorithms highly depends on expert experience.。专家需要根据自己的专业知识和经验，选择合适的算法、确定特征工程的方法以及调整模型参数Experts need to select appropriate algorithms, determine the methods of feature engineering, and adjust model parameters based on their professional knowledge and experience.。对于复杂的睡眠障碍分类问题，不同专家的经验和判断可能存在差异，导致分类结果的不一致性 For complex sleep disorder classification problems, the experiences and judgments of different experts may vary, leading to inconsistencies in classification results.。

传统机器学习算法在睡眠障碍分类中存在的这些问题，限制了睡眠障碍研究的发展和应用，需一种新的技术和方法来突破这些困境These problems of traditional machine - learning algorithms in sleep disorder classification limit the development and application of sleep disorder research, and a new technology and method are needed to break through these dilemmas.。

1.3 大语言模型带来的新契机New Opportunities Brought by Large Language Models

大语言模型（LLMs）作为自然语言处理领域的前沿技术，近年来取得了显著的进展。Large language models (LLMs), as a cutting - edge technology in the field of natural language processing, have made remarkable progress in recent years.它基于 Transformer[14] 架构，通过在海量文本数据上进行无监督预训练，学习到了丰富的语言知识Based on the Transformer [14] architecture, they have learned rich language knowledge through unsupervised pre - training on massive text data.。它在众多领域展现出了巨大的潜力，为解决复杂问题提供了新的思路和方法 They have demonstrated great potential in numerous fields, providing new ideas and methods for solving complex problems.。

在医学研究领域[4]、[6]、[7]、[8]、[9]，大语言模型的应用为睡眠障碍分类带来了新的契机In the field of medical research [4], [6], [7], [8], [9], the application of large language models has brought new opportunities for sleep disorder classification.。大语言模型能够理解和处理自然语言，这使得它可以直接对睡眠健康与生活方式数据集中的文本信息进行分析和解读Large language models can understand and process natural language, enabling them to directly analyze and interpret the text information in sleep health and lifestyle datasets.。它可以从患者的睡眠日志、生活习惯描述等文本数据中提取关键信息They can extract key information from text data such as patients' sleep logs and descriptions of living habits.。

大语言模型还具备强大的知识推理能力[5]，它可以结合已有的医学知识和睡眠障碍的诊断标准，对提取到的信息进行综合分析和判断，从而实现对睡眠障碍的自动分类Large language models also possess powerful knowledge reasoning capabilities [5]. They can comprehensively analyze and judge the extracted information by combining existing medical knowledge and the diagnostic criteria of sleep disorders, thus achieving the automatic classification of sleep disorders.。在面对复杂的睡眠障碍症状时，大语言模型能够通过推理和判断，准确地识别出不同类型的睡眠障碍，如失眠、睡眠呼吸暂停等When faced with complex sleep disorder symptoms, large language models can accurately identify different types of sleep disorders, such as insomnia and sleep apnea, through reasoning and judgment.。

而且，大语言模型还可以通过对大量数据的学习，发现睡眠障碍与其他因素之间的潜在关系，为睡眠障碍的诊断和治疗提供新的依据Moreover, large language models can also discover the potential relationships between sleep disorders and other factors through learning from a large amount of data, providing new bases for the diagnosis and treatment of sleep disorders. 。

1.4 研究贡献Research Contributions

在睡眠障碍分类研究领域，传统方法高度依赖手动操作和专家经验，这不仅效率低下，还难以应对日益复杂的睡眠健康数据，存在明显局限性In the field of sleep disorder classification research, traditional methods are highly dependent on manual operations and expert experience, which is not only inefficient but also difficult to cope with the increasingly complex sleep health data, with obvious limitations。

而提示策略（Prompting）作为大语言模型应用中的关键技术，对引导模型生成预期输出起着决定性作用The prompting strategy (Prompting), as a key technology in the application of large language models, plays a decisive role in guiding the model to generate the expected output.。不同的提示策略在各类任务中的性能表现参差不齐，特别是如何巧妙运用提示策略，使大语言模型仅依据文本信息就能够准确从睡眠健康与生活方式数据中提取关键信息，实现自动睡眠障碍分类，这一难题亟待攻克The performance of different prompting strategies in various tasks varies. In particular, how to skillfully use the prompting strategy to enable the large language model to accurately extract key information from sleep health and lifestyle data based on text information alone and achieve automatic sleep disorder classification is a difficult problem that urgently needs to be solved.。本研究在该背景下，具有多方面重要且独特的贡献Under this background, this study makes important and unique contributions in many aspects：

* **创新模型应用Innovative Model Application**：开创性地将大语言模型引入自动睡眠障碍分类领域Pioneeringly introduce large language models into the field of automatic sleep disorder classification.。传统的睡眠障碍分类方法面对海量、复杂且多变的睡眠健康与生活方式数据时，精度不高、适应性差等问题愈发凸显Traditional sleep disorder classification methods have increasingly prominent problems such as low accuracy and poor adaptability when facing massive, complex, and changeable sleep health and lifestyle data.。而大语言模型凭借其强大的语言理解和生成能力，突破了传统方法的重重局限Large language models, with their powerful language understanding and generation capabilities, have broken through the numerous limitations of traditional methods。本研究构建了一套全新的、基于大语言模型的自动睡眠障碍分类方法，这一方法的核心优势在于，只要提供文本信息，就能让模型自动进行机器学习（auto ML），实现对睡眠障碍的高精度预测，为睡眠障碍的准确诊断开辟了一条崭新且高效的途径This study constructs a new set of automatic sleep disorder classification methods based on large language models. The core advantage of this method is that as long as text information is provided, the model can automatically perform machine learning (auto ML) to achieve high - precision prediction of sleep disorders, opening up a new and efficient way for the accurate diagnosis of sleep disorders。

1. **深度挖掘提示策略助力自动分类Deep Exploration of Prompting Strategies to Facilitate Automatic Classification：**本研究对零样本提示（Zero-shot Prompting）和少样本提示（Few-shot Prompting），分解提示（Decomposed Prompting）技术在大语言模型中的应用展开了深度探索This study deeply explores the application of zero - shot prompting, few - shot prompting, and decomposed prompting techniques in large language models.。本研究系统地挖掘了这些提示策略在自动睡眠障碍分类任务中的全新潜力，清晰明确了它们在基于文本的睡眠健康分类中的优势与不足This study systematically explores the new potential of these prompting strategies in the automatic sleep disorder classification task, clearly clarifying their advantages and disadvantages in text - based sleep health classification.。这些深入分析为后续提示策略的优化提供了明确的方向，有助于持续提升大语言模型在仅依据文本信息进行自动睡眠障碍分类时的性能These in - depth analyses provide a clear direction for the subsequent optimization of prompting strategies and help to continuously improve the performance of large language models in automatic sleep disorder classification based on text information alone.。

* **引领跨学科融合与自动化创新Leading Cross - disciplinary Integration and Automation Innovation：** 本研究通过提示驱动大语言模型，实现了仅凭借文本信息就能自动进行机器学习分类的创新应用，这打破了传统机器学习算法的束缚，真正实现了睡眠障碍分类的自动化和智能化。This study, through prompt - driven large language models, has achieved the innovative application of automatic machine - learning classification based on text information alone, breaking the shackles of traditional machine - learning algorithms and truly realizing the automation and intelligence of sleep disorder classification.这种创新应用不仅有效解决了睡眠医学领域的实际问题，更重要的是，它为大语言模型在医疗健康领域的更广泛应用积累了宝贵经验，极大地推动了跨学科技术的深度融合与发展，为睡眠医学和人工智能的交叉研究开辟了一条充满无限可能的新路径This innovative application not only effectively solves practical problems in the field of sleep medicine but, more importantly, accumulates valuable experience for the wider application of large language models in the field of healthcare, greatly promoting the deep integration and development of cross - disciplinary technologies and opening up a new path full of infinite possibilities for the cross - research of sleep medicine and artificial intelligence.。
* **拓展自动化应用场景Expanding Automation Application Scenarios:：** 在实际应用方面，大语言模型展现出了强大的自动化拓展能力。in practical applications, large language models have demonstrated powerful automation expansion capabilities.它可与多种技术相结合，进一步拓展其在睡眠障碍分类及相关领域（传感器技术、可穿戴设备等）的自动化应用价值They can be combined with a variety of technologies to further expand their automation application value in sleep disorder classification and related fields (such as sensor technology and wearable devices).。

本研究的成果不仅有助于显著提高睡眠障碍的诊断和治疗水平，改善患者的生活质量，而且随着技术的不断发展和完善，大语言模型在睡眠医学领域基于文本信息的自动分类应用前景也将更加广阔，为推动整个睡眠医学行业的自动化、智能化发展注入了强大动力The results of this study not only help to significantly improve the diagnosis and treatment level of sleep disorders and improve the quality of life of patients but also, with the continuous development and improvement of technology, the application prospects of large language models in the automatic classification of sleep medicine based on text information will be broader, injecting strong impetus into the automation and intelligence development of the entire sleep medicine industry.。

## 睡眠健康与生活方式数据集Sleep Health and Lifestyle Dataset

### 3.1 数据集来源、构成Source and Composition of the Dataset

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] The sleep health and lifestyle dataset used in this study is sourced from the Kaggle website [10].。本次使用的睡眠健康与生活方式数据集包含 374 行 13 列数据。具体如下The sleep health and lifestyle dataset used this time contains 374 rows and 13 columns of data. Details are as follows：

1. **个人基本信息Personal basic information:**Person ID（个人编号）作为每个受访者的唯一标识符，有助于在数据处理和分析过程中准确识别和跟踪个体数据。、Gender（性别）信息可以用于研究不同性别在睡眠障碍发生率和睡眠模式上的差异。Age（年龄）是影响睡眠的重要因素之一，随着年龄的增长，睡眠质量往往会下降，睡眠障碍的发生率也会增加 。Occupation（职业）则反映了工作性质、工作时间和工作压力等因素对睡眠的潜在影响Person ID, as the unique identifier for each respondent, helps to accurately identify and track individual data during the data processing and analysis. Gender information can be used to study the differences in the incidence of sleep disorders and sleep patterns between different genders. Age is one of the important factors affecting sleep. As age increases, sleep quality tends to decline, and the incidence of sleep disorders also increases. Occupation reflects the potential impact of work nature, working hours, and work pressure on sleep.。
2. **睡眠相关特征Sleep - related characteristics**：Sleep Duration（睡眠时长）直接反映了个体的睡眠时间，充足的睡眠时长对于维持身体健康和正常的生理功能至关重要。Quality of Sleep（睡眠质量评分）则是一个主观评价指标，通过量表（1 - 10）来衡量，它反映了睡眠的深度、连续性和恢复效果等方面Sleep Duration directly reflects an individual's sleep time. Sufficient sleep duration is crucial for maintaining physical health and normal physiological functions. Quality of Sleep is a subjective evaluation index, measured on a scale from 1 - 10, which reflects aspects such as the depth, continuity, and recovery effect of sleep. 。
3. **生活方式因素Lifestyle factors**：Physical Activity Level（身体活动水平）反映了个体的日常运动量。Stress Level（压力水平）是影响睡眠的重要因素之一 。BMI Category（BMI 类别）是衡量个体体重状况是否健康的指标，与睡眠障碍密切相关。Daily Steps（每日步数）则是一种简单直观的衡量身体活动水平的指标，可以了解个体的日常活动量，进而分析其对睡眠的影响Physical Activity Level reflects an individual's daily exercise volume. Stress Level is one of the important factors affecting sleep. BMI Category is an indicator to measure whether an individual's weight status is healthy and is closely related to sleep disorders. Daily Steps is a simple and intuitive indicator to measure the physical activity level, through which the daily activity volume of an individual can be understood, and then its impact on sleep can be analyzed. 。
4. **健康指标Health indicators**：Blood Pressure（血压）对于维持身体各器官的正常功能至关重要。Heart Rate（心率）反映了心脏的功能状态Blood Pressure is essential for maintaining the normal functions of various organs in the body. Heart Rate reflects the functional state of the heart。

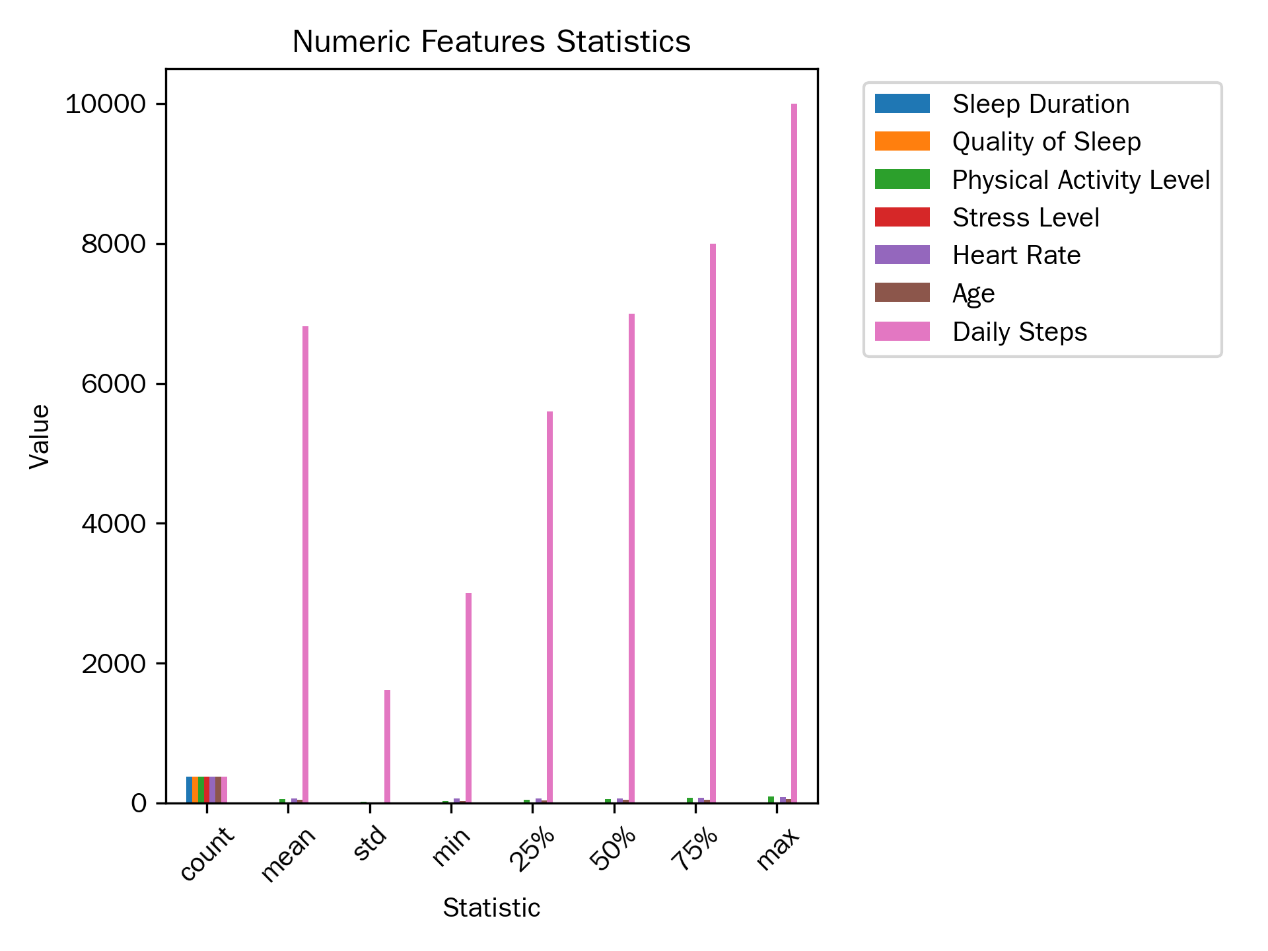
这些变量相互关联，共同反映了受访者的睡眠健康状况和生活方式特点。通过对这些变量的深入分析，可以揭示Sleep Disorder（睡眠障碍情况）与生活方式之间的潜在关系，为睡眠障碍的诊断、治疗和预防提供科学依据 These variables are interrelated and jointly reflect the sleep health status and lifestyle characteristics of the respondents. Through in - depth analysis of these variables, the potential relationship between Sleep Disorder and lifestyle can be revealed, providing a scientific basis for the diagnosis, treatment, and prevention of sleep disorders.。

### 3.2 数据特征分析与可视化Data Feature Analysis and Visualization

**一、数值型特征的统计信息（图 1）Statistical information of numerical features (Figure 1)**

从这些数值型特征的统计数据中可以推测The following inferences can be made from the statistical data of these numerical features：

* **睡眠时长Sleep Duration**：平均睡眠时长约为 7.13 小时，标准差较小，说明整体分布相对集中，大部分人的睡眠时长在 6.4 - 7.8 小时之间，符合成年人正常睡眠时长范围。但仍有部分个体可能存在睡眠时长不足或过长的情况，可能与生活习惯、工作压力等因素有关The average sleep duration is approximately 7.13 hours, with a relatively small standard deviation, indicating that the overall distribution is relatively concentrated. Most people's sleep duration is between 6.4 - 7.8 hours, which is within the normal sleep duration range for adults. However, there may still be some individuals with insufficient or excessive sleep duration, which may be related to factors such as living habits and work pressure.。
* **睡眠质量评分Quality of Sleep score:**平均评分为 7.31 分，标准差 1.20 分，个体之间存在一定差异。25% - 75% 分位数显示大部分人的评分在 6 - 8 分之间，整体睡眠质量处于中等偏上水平。不过，仍有相当一部分受访者睡眠质量较差，评分低于 5 分，这可能受到多种因素的综合影响，如心理压力、生活环境等The average score is 7.31 points, with a standard deviation of 1.20 points, indicating some differences among individuals. The 25% - 75% quantiles show that most people's scores are between 6 - 8 points, and the overall sleep quality is above - average. However, a considerable number of respondents have poor sleep quality, with scores below 5 points, which may be comprehensively affected by various factors such as psychological stress and living environment.。
* **身体活动水平Physical Activity Level**：平均水平为 59.17，标准差 20.83 相对较大，说明不同个体之间的身体活动水平差异明显。这与个人运动习惯、职业特点等因素密切相关，例如从事体力劳动的职业人群和久坐办公室的人群身体活动水平可能有很大差异The average level is 59.17, with a relatively large standard deviation of 20.83, indicating significant differences in physical activity levels among different individuals. This is closely related to factors such as personal exercise habits and occupational characteristics. For example, there may be a large difference in physical activity levels between people in physically - laborious occupations and those who sit in the office for long hours.。
* **压力水平Stress Level**：平均压力水平为 5.39，标准差 1.77，压力水平的分布有一定离散性。不同个体面临的压力程度不同，可能与工作性质、生活事件等因素有关。较高的压力水平可能会对睡眠质量等健康指标产生负面影响The average stress level is 5.39, with a standard deviation of 1.77, indicating a certain degree of dispersion in the distribution of stress levels. Different individuals face different levels of stress, which may be related to factors such as work nature and life events. Higher stress levels may have a negative impact on health indicators such as sleep quality。
* **心率Heart Rate**：平均心率 70.17 次 / 分钟，标准差 4.14，心率的波动范围相对较小，大部分人心率在 68 - 72 次 / 分钟之间。这反映出该数据集中心率总体较为稳定，但仍有部分个体的心率可能偏离正常范围，可能与身体健康状况、运动情况等有关The average heart rate is 70.17 beats per minute, with a standard deviation of 4.14. The fluctuation range of the heart rate is relatively small, and most people's heart rates are between 68 - 72 beats per minute. This reflects that the heart rate in this dataset is generally stable, but the heart rates of some individuals may deviate from the normal range, which may be related to physical health conditions, exercise, etc。
* **年龄Age**：平均年龄为 40.06 岁，标准差 13.34，说明年龄分布有一定的离散性。不同年龄段的人群在睡眠相关特征上可能存在差异，例如年龄较大者可能睡眠时长较短或睡眠质量较差The average age is 40.06 years old, with a standard deviation of 13.34, indicating a certain degree of dispersion in the age distribution. People of different age groups may have differences in sleep - related characteristics. For example, older people may have shorter sleep duration or poorer sleep quality.。
* **每日步数Daily Steps:**平均每日步数为 7070.26 步，标准差 3344.52，步数的差异反映了不同个体的运动习惯和活动量不同。步数可能与身体活动水平、睡眠质量等存在关联，步数较多者可能身体活动水平较高，睡眠质量也较好The average number of daily steps is 7070.26 steps, with a standard deviation of 3344.52. The differences in the number of steps reflect the different exercise habits and activity levels of different individuals. The number of steps may be related to physical activity levels, sleep quality, etc. Those with more steps may have higher physical activity levels and better sleep quality.。



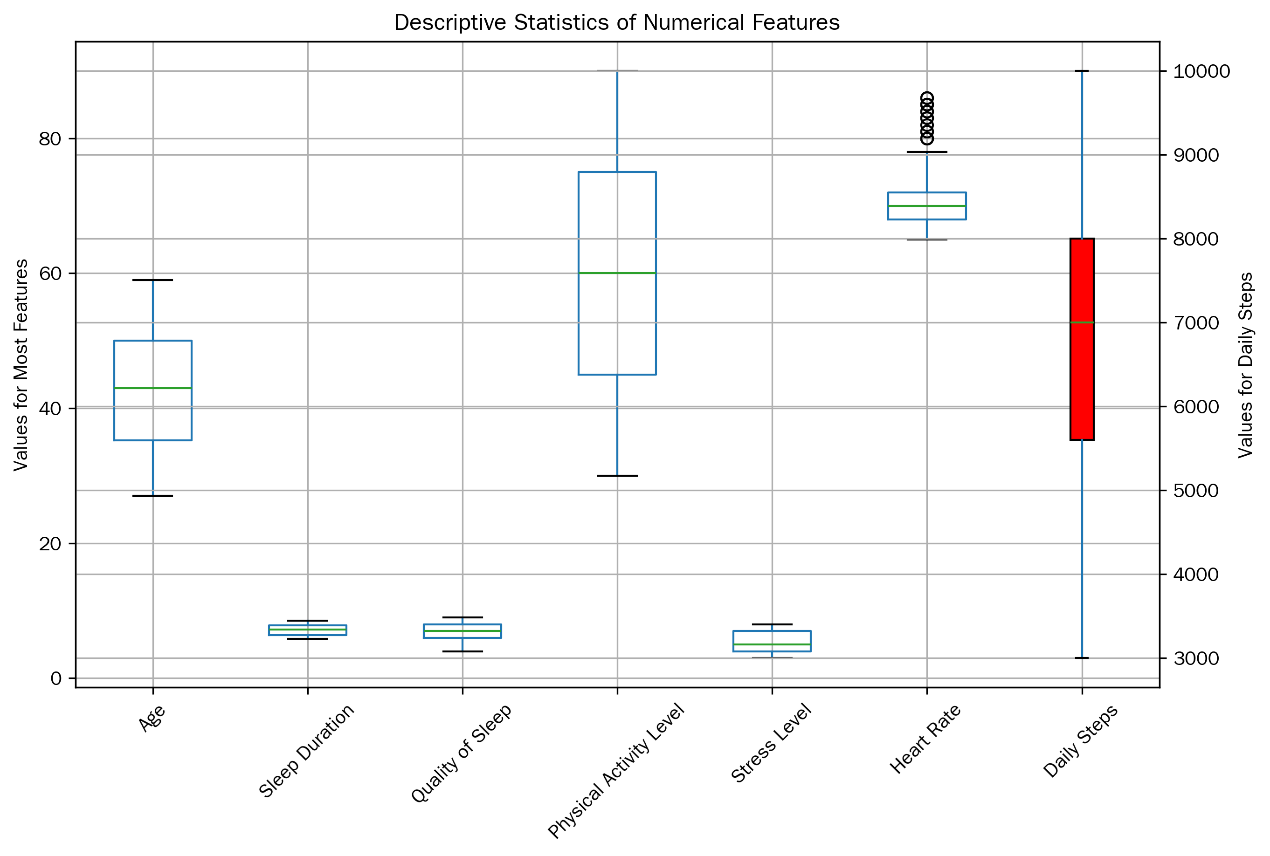


图 1：数值型特征统计图表Figure 1: Statistical chart of numerical features

1. **分类型特征的分布信息（图 2）Distribution information of categorical features (Figure 2)**

从分类型特征的分布中我们可以看出From the distribution of categorical features, we can observe the following：

* **性别Gender**：男性有 212 人，女性有 162 人，男性数量略多于女性。不同性别的人群在睡眠相关特征上可能存在差异，例如女性可能因生理因素在睡眠质量、压力水平等方面表现不同There are 212 males and 162 females. The number of males is slightly more than that of females. People of different genders may have differences in sleep - related characteristics. For example, due to physiological factors, females may perform differently in terms of sleep quality and stress level.。
* **职业Occupation**：各个职业的人数相同，均为 46 人，反映出样本在职业方面具有一定的均衡性。不同职业可能由于工作压力、工作时间等因素，在睡眠时长、质量、压力水平等方面表现出不同的特征。例如，护士可能由于工作的轮班性质，睡眠时长和质量受到影响，压力水平较高The number of people in each occupation is the same, all 46, reflecting a certain balance in the sample in terms of occupation. Different occupations may show different characteristics in terms of sleep duration, quality, and stress level due to factors such as work pressure and working hours. For example, nurses may have their sleep duration and quality affected and a higher stress level due to the shift - work nature of their jobs。
* **BMI 类别Category**：正常体重的受访者有 156 人，超重的有 121 人，肥胖的有 97 人，正常体重和超重的受访者占比较高。不同 BMI 类别的人群可能在睡眠障碍的发生率上存在差异，肥胖人群可能更容易出现睡眠呼吸暂停等问题There are 156 respondents with normal weight, 121 who are overweight, and 97 who are obese. The proportion of respondents with normal weight and overweight is relatively high. People in different BMI categories may have different incidences of sleep disorders. Obese people may be more prone to problems such as sleep apnea.。
* **睡眠障碍情况Sleep Disorder status**：睡眠正常的受访者有 236 人，有睡眠呼吸暂停的有 85 人，失眠的有 53 人。了解不同睡眠障碍类型的分布，有助于针对性地研究睡眠障碍的成因和预防措施。如睡眠呼吸暂停患者可能与肥胖、年龄等因素相关There are 236 respondents with normal sleep, 85 with sleep apnea, and 53 with insomnia. Understanding the distribution of different types of sleep disorders helps to conduct targeted research on the causes and preventive measures of sleep disorders. For example, patients with sleep apnea may be related to factors such as obesity and age。

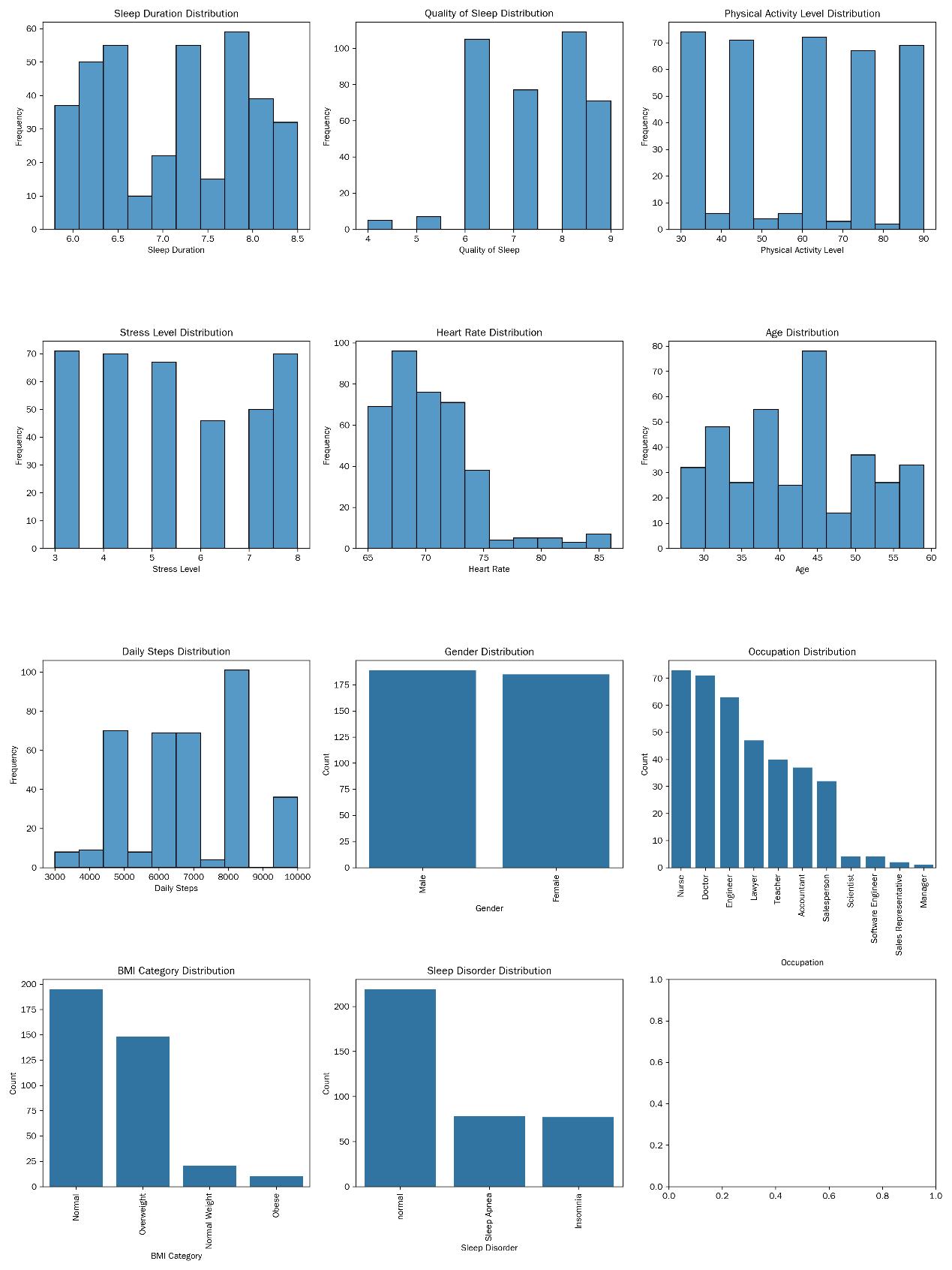


图 2：分类型特征分布图表Figure 2: Distribution chart of categorical features

**三、特征间相关性分析Correlation analysis among features**

计算了各特征之间的皮尔逊相关系数（对于分类型特征，采用合适的方法分析其与其他特征的关系，如卡方检验分析Gender与Sleep Disorder的关系等），通过绘制热力图展示相关系数矩阵（图 3）The Pearson correlation coefficients among various features were calculated (for categorical features, appropriate methods were used to analyze their relationships with other features, such as using the chi - square test to analyze the relationship between Gender and Sleep Disorder, etc.). The correlation coefficient matrix was presented by drawing a heatmap (Figure 3)。

从相关系数矩阵中可以得出以下结论The following conclusions can be drawn from the correlation coefficient matrix：

1. **睡眠时长与睡眠质量Sleep Duration and Quality of Sleep**：二者的相关系数为 0.883213，呈现出较强的正相关关系。这符合一般认知，即睡眠时长越长，身体和大脑能得到更充分的恢复，从而提高睡眠质量。这一结果提示在改善睡眠质量的措施中，可以考虑通过调整睡眠时长来实现The correlation coefficient between the two is 0.883213, showing a strong positive correlation. This is in line with the general understanding that the longer the sleep duration, the more fully the body and brain can recover, thus improving sleep quality. This result suggests that in measures to improve sleep quality, adjusting sleep duration can be considered.。
2. **身体活动水平与其他特征Physical Activity Level and other features**：身体活动水平与睡眠时长、睡眠质量的相关系数分别为 0.212360 和 0.192896，呈现较弱的正相关关系。这表明适度的身体活动对睡眠有一定的积极影响，但这种影响相对有限。身体活动水平与压力水平的相关系数几乎为 0，说明两者之间线性关系不明显。而与心率的相关系数为 0.136971，有较弱的正相关，可能是身体活动水平较高时会引起心率一定程度的上升The correlation coefficients between Physical Activity Level and Sleep Duration, Quality of Sleep are 0.212360 and 0.192896 respectively, showing a weak positive correlation. This indicates that moderate physical activity has a certain positive impact on sleep, but this impact is relatively limited. The correlation coefficient between Physical Activity Level and Stress Level is almost 0, indicating that the linear relationship between the two is not obvious. The correlation coefficient with Heart Rate is 0.136971, showing a weak positive correlation, which may be that a higher physical activity level can cause a certain increase in heart rate.。
3. **压力水平与其他特征Stress Level and other features**：压力水平与睡眠时长、睡眠质量分别呈现 - 0.811023 和 - 0.898752 的强负相关关系。这表明压力是影响睡眠的重要因素，当人们处于高压力状态时，身体分泌的应激激素会干扰入睡和睡眠质量。压力水平与心率的相关系数为 0.670026，呈正相关关系，说明压力越大，心率可能越高，反映出压力对身体生理指标的影响Stress Level shows a strong negative correlation with Sleep Duration and Quality of Sleep, with correlation coefficients of - 0.811023 and - 0.898752 respectively. This indicates that stress is an important factor affecting sleep. When people are under high stress, the stress hormones secreted by the body will interfere with falling asleep and sleep quality. The correlation coefficient between Stress Level and Heart Rate is 0.670026, showing a positive correlation, indicating that the greater the stress, the higher the heart rate may be, reflecting the impact of stress on physical physiological indicators。
4. **心率与其他特征Heart Rate and other features**：心率与睡眠时长、睡眠质量呈负相关关系，意味着心率较高时，睡眠时长和质量可能较差。这可能是由于心率异常反映了身体的某种不适状态，进而影响睡眠Heart Rate shows a negative correlation with Sleep Duration and Quality of Sleep, meaning that when the heart rate is high, the sleep duration and quality may be poor. This may be because an abnormal heart rate reflects a certain uncomfortable state of the body, which in turn affects sleep.。
5. **年龄与其他特征Age and other features**：年龄与睡眠时长、睡眠质量呈负相关关系，说明随着年龄的增长，睡眠时长可能减少，睡眠质量可能变差。年龄与压力水平呈正相关，可能年龄较大者面临的生活压力相对更大Age shows a negative correlation with Sleep Duration and Quality of Sleep, indicating that as age increases, sleep duration may decrease and sleep quality may deteriorate. Age shows a positive correlation with Stress Level, and it is possible that older people face relatively greater life pressure.。
6. **每日步数与其他特征Daily Steps and other features**：每日步数与睡眠时长、睡眠质量呈正相关关系，表明步数较多可能有助于延长睡眠时长和提高睡眠质量。每日步数与身体活动水平呈较强的正相关，符合预期，步数越多身体活动水平越高Daily Steps show a positive correlation with Sleep Duration and Quality of Sleep, indicating that more steps may help to extend sleep duration and improve sleep quality. Daily Steps show a strong positive correlation with Physical Activity Level, which is as expected, the more steps, the higher the physical activity level。

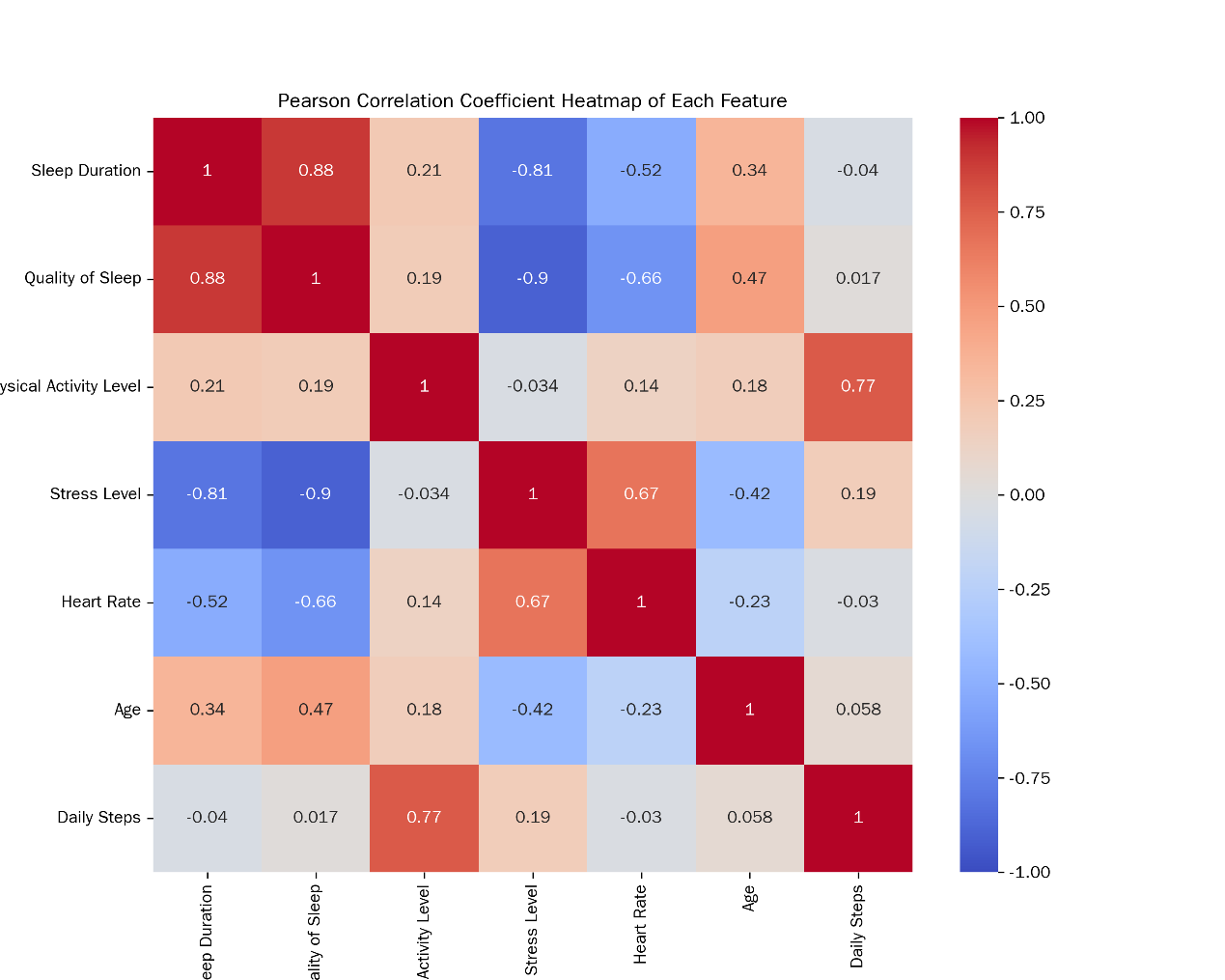


图 3：各特征之间皮尔逊相关系数热力图Figure 3: Heatmap of Pearson correlation coefficients among features

通过对数据集中各特征的分布情况进行统计分析和可视化展示，我们可以更深入地了解数据的特点和规律，为后续利用大语言模型进行睡眠障碍分类提供有力的数据支持。By conducting statistical analysis and visualizing the distribution of each feature in the dataset, we can gain a deeper understanding of the characteristics and patterns of the data, providing strong data support for the subsequent use of large language models in sleep disorder classification.在后续的研究中，我们可以根据这些分析结果，选择合适的特征和模型，提高睡眠障碍分类的准确性和可靠性In subsequent research, we can select appropriate features and models based on these analysis results to improve the accuracy and reliability of sleep disorder classification 。

## 四、研究方法Research Methods

### 4.1 大语言模型选择Selection of Large Language Mode

在睡眠障碍自动分类这一研究领域，大语言模型的选型是影响研究成果准确性与效率的关键因素。经过综合考量，本研究最终选定豆包模型，主要基于以下几方面的深度剖析In the research field of automatic sleep disorder classification, the selection of a large language model is a key factor affecting the accuracy and efficiency of research results. After comprehensive consideration, this study finally selects the Doubao model, mainly based on the following in - depth analysis.。

* 豆包模型构建于 Transformer[14] 架构之上，通过在海量的文本数据中进行无监督预训练，广泛且深入地学习了语言在语义、语法、语用等多方面的知识体系。其独特的多头注意力机制，能够并行地关注输入文本的不同部分，从而捕捉到文本中丰富的语义关联和上下文信息；多层神经网络结构则进一步对这些信息进行深度加工与特征提取，使得豆包模型能够从词汇、语句、篇章等多个维度深入剖析文本内容。
* The Doubao model is built on the Transformer [14] architecture. Through unsupervised pre - training on massive text data, it has extensively and deeply learned the knowledge systems of language in semantics, grammar, pragmatics, and other aspects. Its unique multi - head attention mechanism can concurrently focus on different parts of the input text, thus capturing rich semantic associations and contextual information in the text. The multi - layer neural network structure further deeply processes and extracts features from this information, enabling the Doubao model to deeply analyze text content from multiple dimensions such as vocabulary, sentences, and passages
* 在自然语言处理任务中，豆包模型能充分理解睡眠健康领域专业的医学术语，如 “睡眠呼吸暂停低通气指数”“周期性肢体运动障碍” 等，并且，凭借良好的泛化性能，豆包模型可以快速适应睡眠健康领域复杂多变的文本数据，包括医疗记录、睡眠监测设备的日志数据等，这些数据来源广泛，格式和语言风格差异较大，但豆包模型都能有效处理。
* In natural language processing tasks, the Doubao model can fully understand professional medical terms in the field of sleep health, such as "apnea - hypopnea index" and "periodic limb movement disorder". Moreover, with good generalization performance, the Doubao model can quickly adapt to the complex and variable text data in the field of sleep health, including medical records, log data from sleep monitoring devices, etc. These data come from a wide range of sources, with significant differences in format and language style, but the Doubao model can handle them effectively.
* 此外，豆包模型还具备一些独特的优势，极大地助力了睡眠障碍自动化分类的实现。在交互层面，部分版本（如 PC 版）的豆包模型拥有友好的用户接口，能够直接接受分析 CSV 文件等常见的数据格式，这使得研究人员可以便捷地将睡眠健康与生活方式数据集导入模型进行处理，无需花费大量时间进行数据格式转换等预处理工作。在技术实现层面，豆包模型具备自动生成代码和执行代码的能力，在睡眠障碍自动化分类流程中，该能力可依据文本分析结果快速生成实现分类算法的代码逻辑，并直接执行，不仅大大提高了分类效率，还能够修正人工编写代码过程中可能出现的语法错误、逻辑漏洞等问题，为实现高效、准确的睡眠障碍自动化分类提供了有力支持。
* In addition, the Doubao model has some unique advantages that greatly facilitate the realization of automated sleep disorder classification. At the interaction level, some versions (such as the PC version) of the Doubao model have a user - friendly interface that can directly accept and analyze common data formats such as CSV files. This enables researchers to conveniently import the sleep health and lifestyle dataset into the model for processing without spending a lot of time on pre - processing tasks such as data format conversion. At the technical implementation level, the Doubao model has the ability to automatically generate and execute code. In the automated sleep disorder classification process, this ability can quickly generate the code logic for implementing the classification algorithm based on the text analysis results and directly execute it. This not only greatly improves the classification efficiency but also can correct potential syntax errors and logical loopholes that may occur during manual code writing, providing strong support for the realization of efficient and accurate automated sleep disorder classification

综上所述，豆包模型凭借其强大的自然语言处理能力、对睡眠健康领域数据的高度适配性以及独特的交互和技术实现优势，成为本研究中睡眠障碍自动分类的不二之选In summary, with its powerful natural language processing capabilities, high adaptability to data in the field of sleep health, and unique interaction and technical implementation advantages, the Doubao model becomes the top choice for automatic sleep disorder classification in this study。

### 4.2 精妙提示设计策略Exquisite Prompt Design Strategy

如表1，提示细节， 我们设计了三种提示策略As shown in Table 1, details of prompts, we designed three prompting strategies：

* + 零样本提示（Zero-shot Prompting）规则引导的分类探索: 核心概念是在不提供任何具体示例的情况下，仅凭借对任务的清晰描述和模型自身的预训练知识，引导模型生成相应的回答或完成特定任务 。这一策略的原理基于大语言模型在大规模数据上的预训练，使其具备了广泛的语言理解和知识储备能力，能够根据任务描述中的语义信息，从已学习的知识中提取相关内容并进行推理和判断,
  + Zero - shot Prompting: Rule - guided Classification Exploration: The core concept is to guide the model to generate corresponding answers or complete specific tasks based only on a clear description of the task and the model's own pre - training knowledge without providing any specific examples. The principle of this strategy is based on the pre - training of large language models on large - scale data, enabling them to have extensive language understanding and knowledge reserve capabilities. They can extract relevant content from the learned knowledge and conduct reasoning and judgment according to the semantic information in the task description.
* 少样本提示（Few-shot Prompting）数据模式驱动的分类优化: 其设计思路基于对数据模式的深入挖掘和利用，旨在通过提供一定数量的样本数据，引导大语言模型学习数据中的特征与分类结果之间的对应关系，从而实现更准确的分类. 90 样本提示的优势在于它能够利用训练集中的数据模式，为大语言模型提供更丰富的信息，从而提高分类的准确性 。
* Few - shot Prompting: Data - pattern - driven Classification Optimization: Its design idea is based on the in - depth exploration and utilization of data patterns. It aims to guide the large language model to learn the corresponding relationship between features and classification results in the data by providing a certain amount of sample data, so as to achieve more accurate classification. The advantage of few - shot prompting is that it can utilize the data patterns in the training set to provide more abundant information for the large language model, thereby improving the accuracy of classification
* 分解提示（Decomposed Prompting 任务拆解的分类进阶: 其核心在于将复杂的任务进行细致的拆解，转化为一系列更易于处理的子任务，从而引导大语言模型更高效地完成任务。在自然语言处理领域，这种策略能够充分发挥大语言模型的优势，提升任务完成的质量和效率.
* Decomposed Prompting: Classification Advancement through Task Decomposition: Its core lies in carefully decomposing complex tasks into a series of more manageable sub - tasks, thus guiding the large language model to complete tasks more efficiently. In the field of natural language processing, this strategy can give full play to the advantages of large language models and improve the quality and efficiency of task completion.

每种策略包含两个子任务，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. 任务二（三种提示相同）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。Each strategy contains two sub - tasks. Task 1 is to perform multi - class classification on the data in the specified CSV file and generate a new CSV file. Task 2 (the same for the three prompts) is to evaluate he classification results using specific evaluation indicators and draw relevant charts.

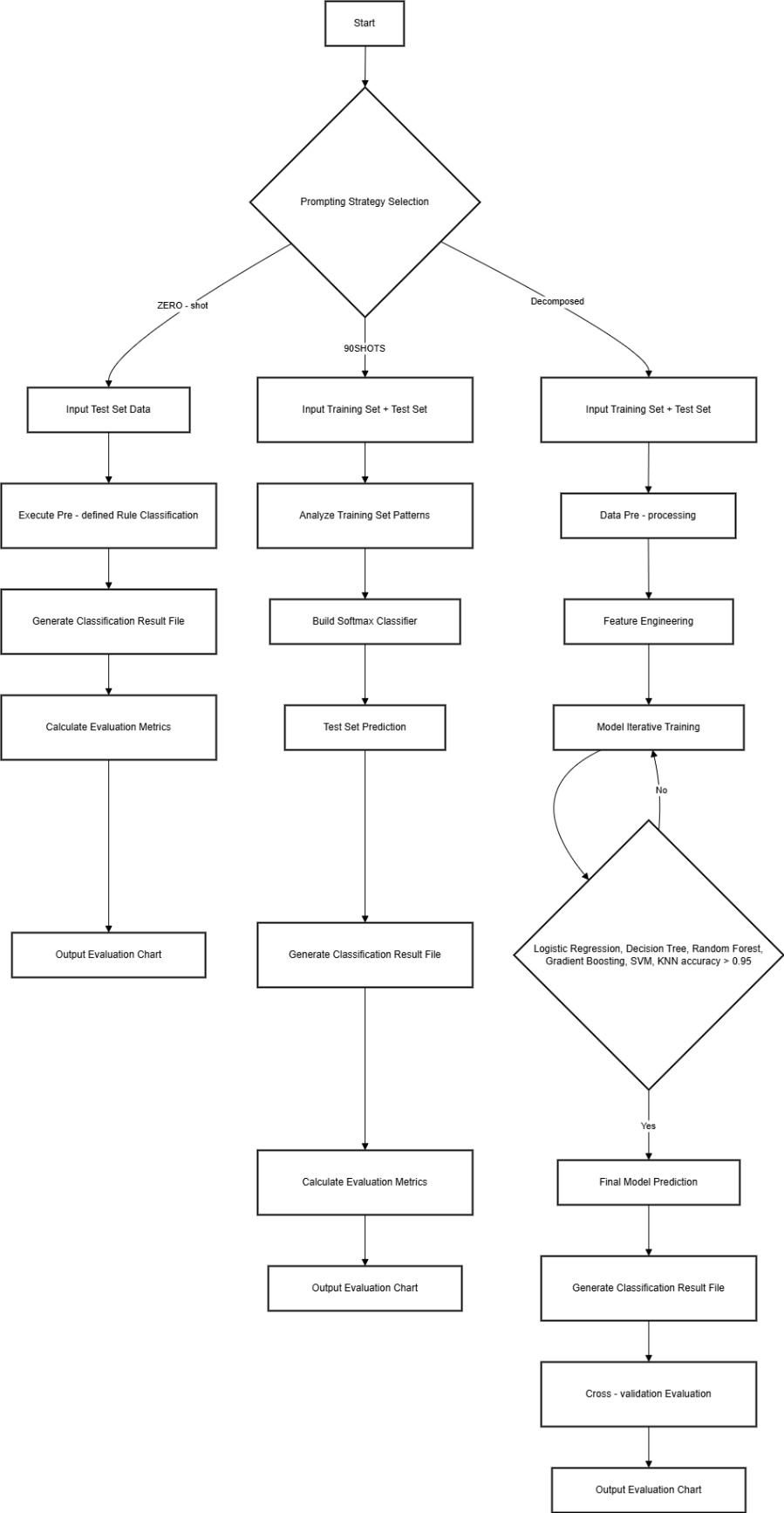


图4：提示策略流程图Figure 4: Flowchart of Prompting Strategies

## 五、实验设计与实施Experimental Design and Implementation

### 5.1 实验环境搭建Construction of the Experimental Environment

实验环境的搭建是确保研究顺利进行的基础，其硬件和软件配置对实验结果的准确性和可靠性有着重要影响。The construction of the experimental environment is the foundation to ensure the smooth progress of the research. Its hardware and software configurations have a significant impact on the accuracy and reliability of the experimental results.在本次睡眠障碍分类研究中，实验选用的处理器为 Intel (R) Core (TM) i5 - 6400T CPU @ 2.20GHz，该处理器具备 2201 Mhz 的主频，拥有 4 个内核和 4 个逻辑处理器 In this sleep disorder classification research, the selected processor for the experiment is Intel (R) Core (TM) i5 - 6400T CPU @ 2.20GHz. This processor has a main frequency of 2201 Mhz, with 4 cores and 4 logical processors.。操作系统采用的是 Microsoft Windows 10 家庭中文版，版本为 10.0.19045 内部版本 19045 The operating system used is Microsoft Windows 10 Home Chinese Edition, with the version number 10.0.19045 (build 19045).。大语言模型选用的是豆包PC版本 1.41.6 .The large language model selected is the Doubao PC version 1.41.6.。

### 5.2 实验步骤流程Experimental Procedure

1. 数据集的划分及样本的选取Division of the dataset and selection of samples:：
   * 从原数据集中按三种类别（正常、睡眠呼吸暂停、失眠）随机各选取 30 个样本，共 90 个样本作为prompts 90 examples 。这一随机选取的方式确保了样本的代表性和随机性，能够在一定程度上反映数据集中不同睡眠障碍类型的特征 。将这 90 个样本保存为Sleep\_health\_and\_lifestyle\_dataset\_selected\_90.csv作为训练集
   * Randomly select 30 samples from each of the three categories (normal, sleep apnea, and insomnia) in the original dataset, resulting in a total of 90 samples as "prompts 90 examples". This random selection method ensures the representativeness and randomness of the samples, which can reflect the characteristics of different sleep disorder types in the dataset to a certain extent. Save these 90 samples as Sleep\_health\_and\_lifestyle\_dataset\_selected\_90.csv as the training set.
   * 在原数据集文件中删去以上 90 个样本，保存为Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv，作为真实标签（GROUND Truth）。真实标签是评估模型分类准确性的重要依据，通过将模型的预测结果与真实标签进行对比，可以准确地评估模型的性能
   * Delete the above 90 samples from the original dataset file and save it as Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv as the ground truth. The ground truth is an important basis for evaluating the classification accuracy of the model. By comparing the model's prediction results with the ground truth, the performance of the model can be accurately evaluated.
   * 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 。测试集用于评估模型在未见过的数据上的表现，能够检验模型的泛化能力和分类准确性
   * Make a copy of Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90.csv, delete the last column ("Sleep Disorder"), and save it as Sleep\_health\_and\_lifestyle\_dataset\_remaining\_90\_without\_last\_column.csv as the test set. The test set is used to evaluate the performance of the model on unseen data and can test the generalization ability and classification accuracy of the model.
2. 手动上传以上 3 个csv文件到豆包 。这一步骤确保了大语言模型能够获取到所需的数据，为后续的分类器设计、训练和评估提供数据基础Manually upload the above three CSV files to Doubao. This step ensures that the large language model can obtain the required data, providing a data foundation for the subsequent design, training, and evaluation of the classifier. 。
3. 根据图4（详情见附件一）不同提示策略进行分类器设计、训练和评估Design, train, and evaluate the classifier according to different prompting strategies in Figure 4 (details can be found in Annex 1).

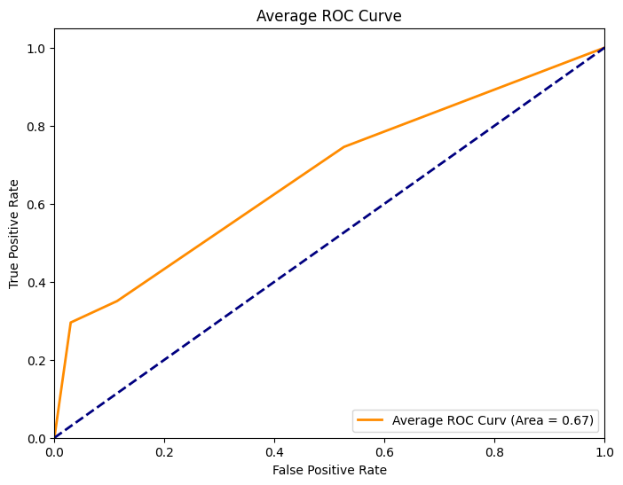
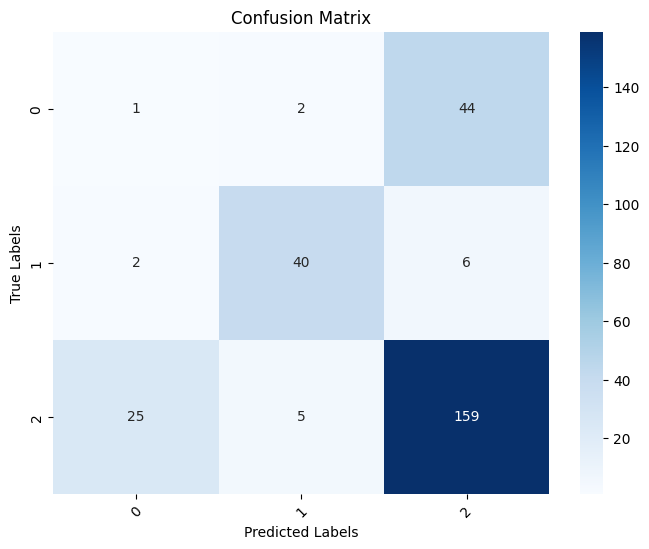
。

### 5.3 实验结果Experimental Results

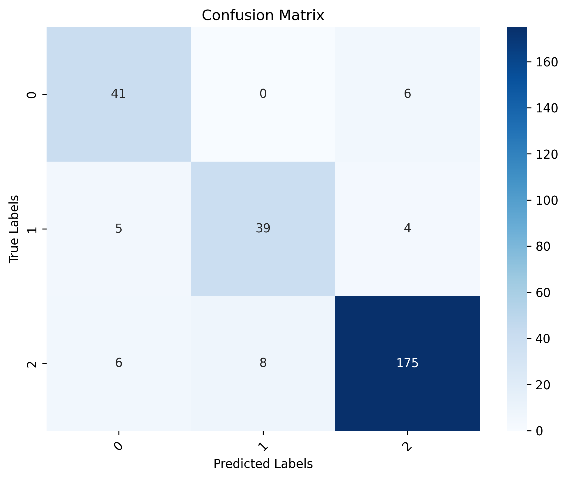
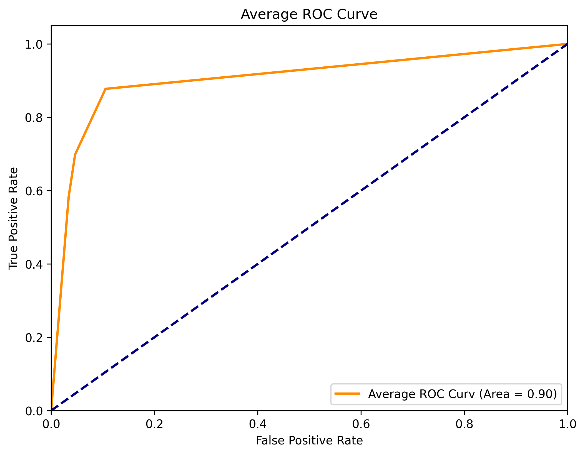
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Prompting Strategy | 准确率Accuracy | 精确率Precision | 召回率Recall | F1 分数F1 Score | AUC 值AUC Value |
| 零样本提示Zero-shot Prompting | 0.704225352112676 | 0.6560363530055422 | 0.704225352112676 | 0.6784663247048188 | 0.6709964855486846 |
| 90 样本提示90-sample Prompting | 0.897887323943662 | 0.9002509272025304 | 0.897887323943662 | 0.8986319612644122 | 0.9041985643947756 |
| 分解提示Decomposed Prompting | 0.9190140845070423 | 0.9191754537248555 | 0.9190140845070423 | 0.9188775418205605 | 0.9163083064019824 |

为了更直观地展示不同提示策略下模型的分类性能，我们绘制了混淆矩阵图和 ROC 曲线（ROC 曲线则通过展示模型在不同阈值下的真正率和假正率之间的关系，评估模型的分类性能 。曲线越靠近左上角，说明模型的分类性能越好 ）To more intuitively display the classification performance of the model under different prompting strategies, we drew confusion matrix diagrams and ROC curves. (The ROC curve evaluates the classification performance of the model by showing the relationship between the true positive rate and the false positive rate of the model at different thresholds. The closer the curve is to the upper - left corner, the better the classification performance of the model.) 。混淆矩阵图以矩阵的形式展示了模型预测结果与真实标签之间的关系，能够清晰地反映出模型在不同类别上的分类准确性和误分类情况 The confusion matrix diagram shows the relationship between the model's prediction results and the ground truth in the form of a matrix, which can clearly reflect the classification accuracy and misclassification situations of the model for different categories.。

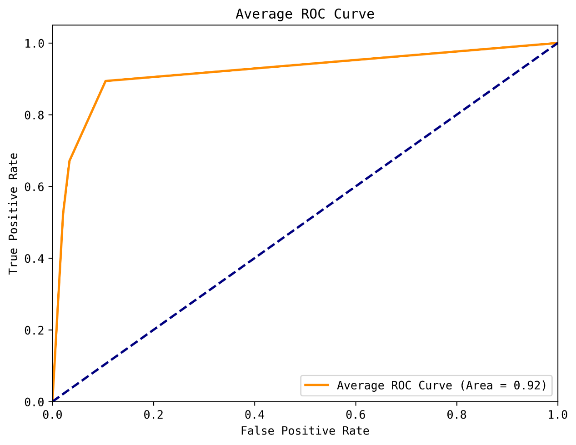
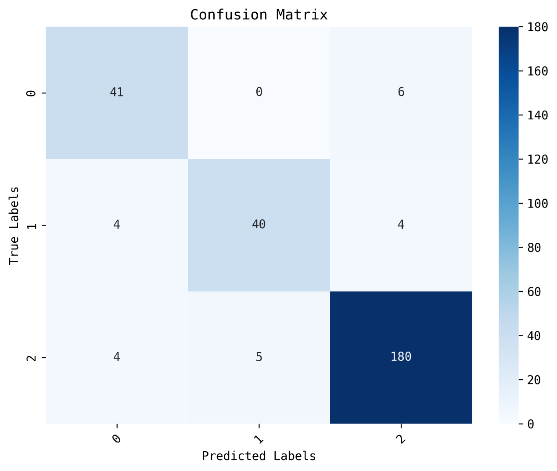
在零样本提示的混淆矩阵中，我们可以看到，模型在区分正常、睡眠呼吸暂停和失眠这三种睡眠障碍类型时，存在较多的误分类情况 。正常样本被误判为睡眠呼吸暂停或失眠的数量较多，睡眠呼吸暂停和失眠样本之间也存在一定程度的误判 In the confusion matrix of zero - shot prompting, we can see that when the model distinguishes between the three types of sleep disorders: normal, sleep apnea, and insomnia, there are relatively many misclassification situations. A large number of normal samples are misjudged as sleep apnea or insomnia, and there is also a certain degree of misjudgment between sleep apnea and insomnia samples。



在 90 样本提示的混淆矩阵中，误分类情况有了明显改善 。正常样本和睡眠呼吸暂停样本的正确分类数量增加，误判情况减少 。但在失眠样本的分类上，仍存在一定的误判情况In the confusion matrix of 90 - sample prompting, the misclassification situation has been significantly improved. The number of correctly classified normal samples and sleep apnea samples has increased, and the misjudgment situation has decreased. However, there are still some misjudgment situations in the classification of insomnia samples.。

分解提示的混淆矩阵显示，模型在各个类别上的分类准确性都有了显著提高 。正常样本、睡眠呼吸暂停样本和失眠样本的正确分类数量都达到了较高水平，误分类情况最少 。这体现出分解提示策略在睡眠障碍分类任务中的有效性和优越性 The confusion matrix of decomposed prompting shows that the classification accuracy of the model for each category has been significantly improved. The number of correctly classified normal samples, sleep apnea samples, and insomnia samples has reached a relatively high level, with the least misclassification situations. This reflects the effectiveness and superiority of the decomposed prompting strategy in the sleep disorder classification task.。



从绘制的 ROC 曲线可以看出，分解提示策略下的曲线最靠近左上角，AUC 值最大，表明其分类性能最佳 。90 样本提示策略的 ROC 曲线次之，零样本提示策略的 ROC 曲线最远离左上角，AUC 值最小，分类性能最差From the drawn ROC curves, it can be seen that the curve under the decomposed prompting strategy is closest to the upper - left corner, with the largest AUC value, indicating the best classification performance. The ROC curve of the 90 - sample prompting strategy is the second - best, and the ROC curve of the zero - shot prompting strategy is the farthest from the upper - left corner, with the smallest AUC value and the worst classification performance 。

### 5.4 实验结果分析Analysis of Experimental Results

* 零样本提示在睡眠障碍分类任务中表现欠佳，主要原因在于其对数据利用的局限性和模型学习能力的不足 。零样本提示仅依据简单的预定义规则进行分类，缺乏对训练数据的深入学习 。在睡眠障碍分类任务中，睡眠障碍的判定受到多种因素的综合影响。这些因素相互交织，形成了复杂的非线性关系 。零样本提示无法从训练数据中学习到这些复杂的特征与分类结果之间的关系，难以捕捉数据中的潜在模式 ，导致分类性能欠佳 。
* The zero - shot prompting performs poorly in the sleep disorder classification task, mainly due to its limitations in data utilization and insufficient model learning ability. Zero - shot prompting classifies only according to simple predefined rules and lacks in - depth learning of the training data. In the sleep disorder classification task, the determination of sleep disorders is comprehensively affected by various factors. These factors are intertwined to form complex non - linear relationships. Zero - shot prompting cannot learn the relationships between these complex features and classification results from the training data and has difficulty capturing the potential patterns in the data, resulting in poor classification performance.
* 90 样本提示相比零样本提示，分类性能显著提升。大语言模型参考训练集数据模式、特征与分类结果的对应关系，能学习到不同特征组合和睡眠障碍类型的潜在联系，像分析训练集中性别、年龄、职业等因素与睡眠障碍类型的关系，可更好理解这些因素在分类中的作用。并且用逻辑回归模型分类，通过对类别型特征编码、数值型特征标准化等预处理提升数据质量，优化学习效果，在面对常见睡眠障碍类型时判断更准确。但逻辑回归模型是线性的，面对特征间复杂非线性关系的数据，分类能力有限，处理睡眠障碍与多种因素的复杂非线性关系时，可能无法准确捕捉，导致分类错误。
* Compared with zero - shot prompting, the 90 - sample prompting has a significantly improved classification performance. The large language model can learn the potential relationships between different feature combinations and sleep disorder types by referring to the data patterns, and the correspondence between features and classification results in the training set. For example, by analyzing the relationships between factors such as gender, age, and occupation in the training set and sleep disorder types, it can better understand the roles of these factors in classification. Moreover, when using the logistic regression model for classification, pre - processing such as encoding categorical features and standardizing numerical features can improve data quality and optimize the learning effect, making the judgment more accurate when facing common sleep disorder types. However, the logistic regression model is linear, and its classification ability is limited when dealing with data with complex non - linear relationships between features. When dealing with the complex non - linear relationships between sleep disorders and various factors, it may not be able to accurately capture these relationships, resulting in classification errors.
* 分解提示策略在睡眠障碍分类任务中取得了最佳效果，这主要归功于其对任务的有效拆解和对多种分类器的优化选择 。分解提示将任务进行分解，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 。支持向量机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 。在判断失眠与压力水平、生活习惯等因素的关系时，SVM 可以通过寻找最优分类超平面，准确地将失眠样本与其他样本区分开来 。分解提示对任务的分解使得模型能够更深入地理解任务要求，逐步完成数据处理、模型训练和评估等环节，提高了模型的分类准确性和可靠性 。
* The decomposed prompting strategy has achieved the best results in the sleep disorder classification task, mainly due to its effective task decomposition and optimized selection of multiple classifiers. The decomposed prompting decomposes the task, tries multiple classifiers such as logistic regression, decision tree, random forest, gradient boosting, support vector machine, and K - nearest neighbor, and performs parameter tuning. Through the comparison and screening of different classifiers, the support vector machine (SVM) is determined to be the most suitable model for this dataset. The support vector machine can find the optimal classification hyperplane in high - dimensional space and effectively handle complex non - linear relationships. In the sleep disorder classification problem, there are complex non - linear relationships between many features, such as sleep duration and sleep quality, stress level and daily activity volume. SVM can accurately capture these relationships and thus obtain the best classification performance. When judging the relationships between insomnia and factors such as stress level and living habits, SVM can accurately distinguish insomnia samples from other samples by finding the optimal classification hyperplane. The task decomposition of decomposed prompting enables the model to more deeply understand the task requirements and gradually complete the links such as data processing, model training, and evaluation, improving the classification accuracy and reliability of the model.

通过对不同提示策略的实验结果进行分析与对比，我们可以看出，在睡眠障碍分类任务中，充分利用训练数据、提升模型学习能力以及合理分解任务和选择分类器是提高模型性能的关键因素 。分解提示策略在处理复杂的睡眠障碍分类问题时具有明显的优势，为睡眠障碍分类提供了更有效的方法和思路 By analyzing and comparing the experimental results of different prompting strategies, we can see that in the sleep disorder classification task, making full use of training data, enhancing the model's learning ability, and reasonably decomposing tasks and selecting classifiers are the key factors to improve the model's performance. The decomposed prompting strategy has obvious advantages in dealing with complex sleep disorder classification problems, providing more effective methods and ideas for sleep disorder classification。

## 六、研究结论与展望Research Conclusions and Prospects

### 6.1 研究结论Research Conclusions

在本次睡眠障碍分类研究中，我们深入探索了大语言模型在睡眠健康与生活方式数据集上的应用，通过精心设计的实验方案和严谨的分析过程，取得了一系列成果In this sleep disorder classification research, we deeply explored the application of large language models on the sleep health and lifestyle dataset. Through a carefully designed experimental plan and a rigorous analysis process, we have achieved a series of results.。大语言模型在睡眠障碍自动分类任务中展现出了显著的潜力。不同的提示策略对大语言模型的性能产生了关键影响Large language models have demonstrated significant potential in the automatic sleep disorder classification task. Different prompting strategies have a crucial impact on the performance of large language models：

1.零样本提示作为一种简单直接的提示策略，在睡眠障碍分类任务中表现相对较弱 。这表明，在缺乏具体示例和深入学习的情况下，大语言模型仅凭预训练知识和简单规则，难以准确应对睡眠障碍分类的复杂性As a simple and direct prompting strategy, zero - shot prompting performs relatively weakly in the sleep disorder classification task. This indicates that without specific examples and in - depth learning, large language models can hardly accurately cope with the complexity of sleep disorder classification relying only on pre - training knowledge and simple rules 。

2.90 样本提示策略通过引入一定数量的样本数据，为大语言模型提供了更丰富的学习信息 。与零样本提示相比，其的分类性能有了显著提升 。这充分说明，利用训练数据中的模式和对应关系，能够增强大语言模型对睡眠障碍分类任务的理解和执行能力 。然而，90 样本提示策略仍存在一定的局限性 。由于其使用的逻辑回归模型具有线性特性，对于特征间存在复杂非线性关系的数据，难以准确捕捉和处理这些关系，从而限制了模型的分类能力The 90 - sample prompting strategy provides the large - language model with more abundant learning information by introducing a certain amount of sample data. Compared with the zero - shot prompting, its classification performance has been significantly improved. This fully demonstrates that leveraging the patterns and corresponding relationships in the training data can enhance the large - language model's understanding and execution ability for the sleep disorder classification task. However, the 90 - sample prompting strategy still has certain limitations. Due to the linear nature of the logistic regression model it uses, it is difficult to accurately capture and handle the relationships between features when there are complex non - linear relationships in the data, thus limiting the model's classification ability.。

3.分解提示策略在睡眠障碍分类任务中取得了最为优异的成绩 。对多种分类器进行了全面的尝试和深入的参数调优 。经过评估和比较，合理分解任务和选择合适的分类器，表明其能够能够充分发挥大语言模型的优势，提高睡眠障碍分类的准确性和可靠性 The decomposed prompting strategy has achieved the most outstanding results in the sleep disorder classification task. A comprehensive attempt has been made on various classifiers, along with in - depth parameter tuning. Through evaluation and comparison, the reasonable decomposition of tasks and the selection of appropriate classifiers indicate that it can give full play to the advantages of the large - language model, improving the accuracy and reliability of sleep disorder classification.。

在利用大语言模型进行睡眠障碍分类研究中，数据与任务理解对模型性能影响重大，直接关乎分类准确性。提供详细示例、参考信息和合理分解任务，有助于提升模型分类的准确性。因此，未来研究应着重深度剖析数据和任务，优化提示策略，以此增强模型性能。同时，在分解提示过程中，根据任务特性选择适配模型并加以优化是提升分类性能的核心要点。由于不同分类器适用场景各异，面对复杂的睡眠障碍分类任务时，需综合权衡多种因素，精准挑选最佳分类器并细致调整参数，从而实现最优分类效果In the research of using large - language models for sleep disorder classification, the understanding of data and tasks has a significant impact on the model's performance, which is directly related to classification accuracy. Providing detailed examples, reference information, and reasonably decomposing tasks can help improve the classification accuracy of the model. Therefore, future research should focus on in - depth analysis of data and tasks, and optimize the prompting strategy to enhance the model's performance. At the same time, in the process of decomposed prompting, choosing an appropriate model according to the task characteristics and optimizing it is the core point for improving the classification performance. Since different classifiers are suitable for different scenarios, when facing complex sleep disorder classification tasks, it is necessary to comprehensively consider various factors, accurately select the best classifier, and carefully adjust the parameters to achieve the optimal classification effect.。

### 6.3 研究局限及未来方向Research Limitations and Future Directions

本研究在睡眠障碍分类中虽然取得了一定的成果，但在数据、模型、实验设计、提示工程等方面仍存在局限性。如数据的多样性不足[15]、分类器模型的选择没有利用最新的深度学习强大的特征学习能力、对比实验不够充分和没有外部验证、优化提示策略不足。未来的研究可以从以上多个方向展开，以进一步提升大语言模型在睡眠障碍自动分类任务中的性能和应用价值Although this research has achieved certain results in sleep disorder classification, there are still limitations in aspects such as data, model, experimental design, and prompt engineering. For example, the data lacks diversity [15], the selection of the classifier model does not utilize the powerful feature - learning ability of the latest deep learning, the comparative experiments are insufficient, there is no external verification, and the optimization of the prompting strategy is inadequate. Future research can be carried out in multiple directions mentioned above to further improve the performance and application value of large - language models in the automatic sleep disorder classification task 。

* 扩充数据集是当务之急，需要广泛收集各类睡眠障碍的样本，特别是罕见或特殊类型的睡眠障碍数据。同时，改善数据分布不均衡问题也是关键，可利用数据重采样技术，如过采样（SMOTE 算法）和欠采样[16]，调整数据集中不同类别样本的比例，让模型在训练过程中能更均衡地学习各类睡眠障碍的特征从而提高模型对特定群体的泛化能力。
* Expanding the dataset is an urgent task. It is necessary to widely collect samples of various sleep disorders, especially data on rare or special types of sleep disorders. At the same time, it is crucial to address the issue of unbalanced data distribution. Data resampling techniques such as oversampling (SMOTE algorithm) and undersampling [16] can be used to adjust the proportion of different types of samples in the dataset, enabling the model to learn the characteristics of various sleep disorders more evenly during the training process, thus improving the model's generalization ability for specific groups.
* 在分类器模型改进方面，探索更先进的架构和特征工程方法至关重要 。深度学习模型具有强大的特征学习能力[17]，如卷积神经网络（CNN）[18]、循环神经网络（RNN）[19]及其变体长短期记忆网络（LSTM）[20]、门控循环单元（GRU）[21]等 。这些模型在处理时间序列数据和复杂数据结构方面具有优势，能够自动学习这些数据中的特征和模式，更好地挖掘睡眠障碍与难以量化因素之间的潜在关系。
* In terms of improving the classifier model, it is essential to explore more advanced architectures and feature - engineering methods. Deep - learning models have powerful feature - learning capabilities [17], such as convolutional neural networks (CNN) [18], recurrent neural networks (RNN) [19], and their variants long - short - term memory networks (LSTM) [20], gated recurrent units (GRU) [21], etc. These models have advantages in processing time - series data and complex data structures. They can automatically learn the features and patterns in these data, and better explore the potential relationships between sleep disorders and difficult - to - quantify factors.
* 在实验设计完善方面，应开展更全面的对比实验和外部验证 。全面对比不同的模型架构（包括LLM和分类模型）、提示策略等因素对分类性能的影响，能够帮助我们深入了解各种因素的作用机制，找到最适合睡眠障碍分类任务的方法和参数设置 。通过在多个不同来源的外部数据集上进行测试，可以评估模型在不同数据分布和场景下的泛化能力，避免模型出现过拟合问题。
* In terms of improving the experimental design, more comprehensive comparative experiments and external verification should be carried out. Comprehensively comparing the impacts of different model architectures (including LLM and classification models), prompting strategies, and other factors on the classification performance can help us deeply understand the action mechanisms of various factors and find the most suitable methods and parameter settings for the sleep disorder classification task. By testing on multiple external datasets from different sources, the generalization ability of the model under different data distributions and scenarios can be evaluated, and the problem of overfitting can be avoided.
* 在提示策略优化方面，可以探索使用更复杂的提示方式，如链式思维提示（Chain of Thought Prompting[22]）、情境提示（Contextual Prompting）[23]等，引导模型进行更深入的推理和分析。链式思维提示可以让模型在回答问题时，逐步展示其推理过程，提高模型的可解释性[24]。情境提示可以根据任务的具体情境，为模型提供更多的背景信息和上下文，帮助模型更好地理解任务要求，从而更准确地进行睡眠障碍分类 。
* In terms of optimizing the prompting strategy, more complex prompting methods can be explored, such as Chain of Thought Prompting [22], Contextual Prompting [23], etc., to guide the model to conduct more in - depth reasoning and analysis. Chain of Thought Prompting can enable the model to gradually display its reasoning process when answering questions, improving the interpretability of the model [24]. Contextual Prompting can provide the model with more background information and context according to the specific situation of the task, helping the model better understand the task requirements and thus more accurately classify sleep disorders.

七、致谢Acknowledgments

在完成这项研究的过程中，我得到了许多个人和机构的帮助，在此，我要向他们表达我最诚挚的感谢During the process of completing this research, I have received help from many individuals and institutions. Here, I would like to express my sincere gratitude to them。

我要衷心感谢我的导师。在整个研究过程中，导师凭借其深厚的学术造诣和丰富的经验，为我指明了研究方向。当我在研究中遇到困惑和难题时，导师总是耐心地倾听我的问题，给予我宝贵的建议和指导。在数据集的分析和提示策略的设计阶段，导师的专业见解让我对研究有了更深入的理解，避免了许多弯路。导师严谨的治学态度和对科研的执着追求，也一直激励着我不断前进，是我学术道路上的榜样I would like to wholeheartedly thank my supervisor. Throughout the research process, with profound academic attainments and rich experience, my supervisor has pointed out the research direction for me. When I encountered confusion and difficulties in the research, my supervisor always listened to my questions patiently and provided me with valuable suggestions and guidance. During the analysis of the dataset and the design stage of the prompting strategy, my supervisor's professional insights gave me a deeper understanding of the research and helped me avoid many detours. My supervisor's rigorous academic attitude and persistent pursuit of scientific research have always inspired me to keep moving forward and served as a role model on my academic path.。

我同样要感谢数据提供方。他们为我们提供了宝贵的睡眠健康与生活方式数据集，这是整个研究的基础I also want to thank the data providers. They have provided us with the valuable sleep health and lifestyle dataset, which is the foundation of the entire research。

在未来的学术道路上，我将继续努力，不断探索，为睡眠医学领域的发展贡献自己的力量On the future academic path, I will continue to work hard, keep exploring, and contribute my efforts to the development of the field of sleep medicine。

八参考文献References

[1] Ibomoiye Domor Mienye, N., Jere, N. (Year Unknown). Survey of Decision Trees: Concepts, Algorithms, and Applications [Literature Type]. In IEEE Xplore . <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10562290>  
[2] Kumari, A., Akhtar, M., Shah, R., et al. Support matrix machine: A review[EB/OL]. <https://arxiv.org/abs/2310.19717>.  
[3] Curth, A., Jeffares, A., van der Schaar, M. Why do Random Forests Work? Understanding Tree Ensembles as Self-Regularizing Adaptive Smoothers[EB/OL]. <https://arxiv.org/abs/2402.01502>.  
[4] Kim, Y., Xu, X., McDuff, D., 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] Katz, D. M., Bommarito, M. J., Gao, S., et al. GPT-4 passes the bar exam[J]. Philosophical Transactions of the Royal Society A, 2024, 382(2270):20230254.  
[6] Nori, H., Lee, Y. T., Zhang, S., et al. Can generalist foundation models outcompete special-purpose tuning? Case study in medicine[EB/OL]. arXiv preprint arXiv:2311.16452, 2023.  
[7] Saab, K., Tu, T., Weng, W.-H., et al. Capabilities of Gemini models in medicine[EB/OL]. arXiv preprint arXiv:2404.18416, 2024.  
[8] Singhal, K., Tu, T., Gottweis, J., et al. Towards expert-level medical question answering with large language models[EB/OL]. arXiv preprint arXiv:2305.09617, 2023.  
[9] McDuff, D., Schaekermann, M., Tu, T., 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>.  
[11] Wang, G., Zhao, W., Han, J., et al. MedFound: The First Medical Large Language Model Passing the Physician Qualification Examination [J]. Journal of Artificial Intelligence, 2024, 5(1): 1 - 12.  
[12] McDuff, D., Xu, X., Kim, Y., et al. Personal Health Large Language Model (PH-LLM): leveraging large language models for personalized health insights [EB/OL]. arXiv preprint arXiv:2311.17133, 2023.  
[13] Zhang, Y., Maziarka, P., Klicpera, J., et al. DiffSBDD: Equivariant diffusion for structure-based drug design [EB/OL]. arXiv preprint arXiv:2403.14338, 2024.

[4]Vaswani, A., Shazeer, N., Parmar, N., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems 30 (NIPS 2017).

[15] Liu, X., Uchiyama, M., Okawa, M., et al. Prevalence and correlates of insomnia in the Japanese general population: Results from the Japan epidemiological sleep study[J]. Sleep, 2000, 23(4):497 - 506.  
[16] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. SMOTE: Synthetic Minority Over - sampling Technique[J]. J Artif Intell Res, 2002, 16:321 - 357.  
[17] LeCun, Y., Bengio, Y., Hinton, G. Deep Learning[J]. Nature, 2015, 521(7553):436 - 444.  
[18] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. Gradient - Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278 - 2324.  
[19] Graves, A., Mohamed, A. - R., Hinton, G. Speech Recognition with Deep Recurrent Neural Networks[C]. In Proceedings of the 2013 IEEE international conference on acoustics, speech and signal processing (pp. 6645 - 6649), 2013.  
[20] Hochreiter, S., Schmidhuber, J. Long Short - Term Memory[J]. Neural Comput, 1997, 9(8):1735 - 1780.  
[21] Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y. On the properties of neural machine translation: Encoder–decoder approaches[EB/OL]. arXiv preprint arXiv:1409.1259, 2014.  
[22] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichtertz, R., Joshi, S., Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models[EB/OL]. arXiv preprint arXiv:2201.11903, 2022.  
[23] Zhou, X., Huang, M., Wang, H., Zhang, Z. Contextual Prompting for Few - Shot Text Classification[C]. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 9312 - 9327), 2022.  
[24] Ribeiro, M. T., Singh, S., Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier[C]. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1135 - 1144), 2016.