利用由 LLM 衍生的嵌入从社交媒体文本中解释性地检测抑郁症

Samuel Kim
Department of Computer Science
Earlham College
Richmond, IN, USA
skim24@earlham.edu

Oghenemaro Imieye
Department of Computer Science
Earlham College
Richmond, IN, USA
ooimieye23@earlham.edu

Yunting Yin
Department of Computer Science
Earlham College
Richmond, IN, USA
yinyu@earlham.edu

***Abstract*—在社交媒体中准确且可解释地检测抑郁语言对于早期干预心理健康状况非常有用，并且对临床实践和更广泛的公共卫生工作有重要影响。在本文中，我们研究了大型语言模型（LLMs）和传统机器学习分类器在涉及社交媒体数据的三个分类任务中的表现：二元抑郁症分类、抑郁症严重程度分类以及在抑郁症、创伤后应激障碍（PTSD）和焦虑之间的鉴别诊断分类。我们的研究将零射 LLMs 与基于常规文本嵌入和 LLM 生成的摘要嵌入进行训练的监督分类器进行了比较。我们的实验表明，虽然零射 LLMs 在二元分类中表现出强大的泛化能力，但它们在细粒度的序数分类中表现不佳。相比之下，使用 LLM 生成的摘要嵌入进行训练的分类器在这些分类任务中表现出竞争力，且在某些情况下表现优于使用传统文本嵌入的模型。我们的研究结果展示了 LLMs 在心理健康预测中的优势，并提出了更好利用其零射能力和基于上下文的总结技术的有希望的方向。**

***Index Terms*—depression detection, large language model, clinical natural language processing**

I. 简介

精神健康疾病如抑郁症影响了全球数亿人，许多病例因社会污名、费用或缺乏护理渠道而未经诊断或治疗。随着越来越多的人在社交媒体上表达他们的思想和情感，这些平台已成为评估心理健康状况的实时数据的宝贵来源。从社交媒体帖子中自动检测抑郁语言可以成为一种有希望的大规模、低成本精神健康筛查和干预工具。以往的抑郁症分类方法通常依赖于两种类型的特征：心理语言学标记和从预训练语言模型中提取的文本嵌入。虽然像从语言研究及词汇计数（LIWC）词库中提取的心理语言学特征提供了可解释性，但它们在表达能力上存在局限性。另一方面，传统的句子嵌入捕捉丰富的语义信息，但可能缺乏心理健康预测任务所需的具体情感线索。

在这项工作中，我们提出了一种新颖的基于提示的嵌入方法，该方法利用大型语言模型的推理能力来生成具有更多可解释性和语义丰富的嵌入。我们并非直接对原始输入文本进行嵌入，而是通过一个与心理健康相关的问题来提示大型语言模型，同时提供用户的社交媒体帖子。然后，我们使用句子编码器从该大型语言模型的摘要中提取嵌入，并将其作为分类器的输入。该方法通过迫使大型语言模型生成超越表层语法的语义丰富的解释来引入推理。它还通过允许大型语言模型过滤掉不相关的信息来减少噪音。此外，它通过生成可以作为干预或诊断一部分向临床医师展示的中间摘要来提高可解释性。我们在五个基于社交媒体的抑郁症数据集上评估了我们的方法，发现与使用原始文本嵌入的模型相比，基于大型语言模型总结的嵌入改善了预测性能。

本文的其余部分组织如下。第 II 节回顾相关工作，包括基于传统文本的方法预测社交媒体中的心理健康情况，以及在此领域中大语言模型的最新应用。第 **??** 节介绍我们的方法学，包括数据预处理、使用文本嵌入和心理语言学特征进行特征提取，以及生成基于大语言模型的摘要嵌入。第 III 节报告了三个分类任务的实验结果：二分类抑郁症分类、抑郁症严重程度分类，以及抑郁症、焦虑和创伤后应激障碍之间的差异诊断。最后，第 IV 节通过讨论关键发现和未来方向总结了本文。

II. 相关工作

A. 文本作为心理健康的预测因素

文本数据，无论是源自书面语言、转录的语音，还是在线互动，都提供了一个有力的途径来了解心理健康。许多研究表明，语言模式可以反映与心理障碍相关的情绪状态和临床症状。压力检测已经通过各种文本来源进行了探索，包括在线博客和论坛帖子 [1]–[3] 以及社交媒体互动 [4]–[7]。创伤后应激障碍（PTSD）的诊断也通过临床患者叙述 [8], [9] 、在线调查 [10] 和转录语音邮件 [11] 进行了实施。Sawalha 等人 [12] 认为，利用半结构化虚拟访谈的转录文本进行情感分析，可以有效识别患有 PTSD 的个人，其方法使用随机森林模型分析 VADER 情感分数。基于深度学习方法，Zeberga 等人 [13] 提出了一种框架，通过使用 BERT 和 Bi-LSTM 模型检测社交媒体帖子中的抑郁和焦虑，以保留上下文和语义意义，并结合知识蒸馏方法以提高效率和准确性。Mansoor 等人 [14] 引入了一种多模态 AI 模型，通过分析多语言社交媒体数据来检测心理健康危机的早期迹象，并强调了现实世界心理健康系统中道德保障和文化敏感应用的需要。Althoff 等人 [15] 使用计算话语方法对基于短信的咨询对话进行了大规模定量分析。Ewbank 等人。[16] 开发了一种深度学习模型，用于在互联网认知行为疗法中自动对患者话语进行分类。Bantilan 等人 [17] 提出了一种自然语言处理模型，通过使用治疗师干预模式和专家注释来标记风险水平，检测远程治疗过程中患者信息中的自杀风险。这些研究强调了自然语言处理（NLP）在心理健康评估和干预中的日益潜力，并展示了上下文和语言特征在这些模型的现实世界应用中的重要性。

This research was supported by the Lemann Student/Faculty Collaborative Research Fund at Earlham College. We gratefully acknowledge their funding and support.

B. 用于心理健康预测的大型语言模型

最近在大型语言模型方面的进展使它们能够应用于广泛的领域，包括心理健康状况的分析和预测。徐等人 [18] 评估了几种大型语言模型在使用在线文本数据进行心理健康预测任务中的表现。他们的研究结果表明，尽管零次和少次提示获得的结果有限，但指令微调显著提高了准确性。Boggavarapu 等人 [19] 探索了使用增强检索生成的 LLMs 从临床笔记中预测心理健康相关的 ICD-10-CM 代码，并发现当前的 LLMs 在准确解释这些复杂代码方面仍然存在困难。他们的研究结果表明，需要更好地整合结构化医学知识到这些模型中。Malgaroli 等人 [20] 讨论了 LLMs 在通过改善诊断、监测和治疗方面推进心理健康护理的潜力。他们还指出了偏见、可访问性以及数据表示等挑战。钱等人 [21] 探索了基础模型（如 LLMs）如何通过个性化诊断、实时监控、情绪识别和使用多模态数据的自适应干预来变革数字化心理健康。他们提出了一个社会技术框架，该框架将脑启发 AI 和临床监督与伦理考量相结合。花等人 [22] 回顾了 LLM 在心理健康护理中的应用现状，并得出结论认为在咨询和临床支持方面有很有前途的案例，但大多数研究缺乏标准化的评估方法。这些研究共同表明了 LLMs 在心理健康诊断和护理中的巨大潜力。解决模型可靠性、可解释性和评估严谨性方面的挑战对于将 LLMs 融入现实临床环境至关重要。

我们的方法分为四个主要阶段：数据预处理、文本嵌入和心理语言特征提取、零样本大型语言模型提示、模型总结嵌入生成，以及模型训练和评估。

我们对五个公开可用的基于社交媒体的心理健康数据集进行了预处理，以用于我们的实验：MHB [23] 、CAMS [24] 、HelaDepDet [25] 、RMHD [26] 和 DepressionEmo [27] 。每个数据集由带有心理健康标签的用户生成的简短文本条目组成，主要与抑郁症相关。为了评估模型区分抑郁和非抑郁语言的能力，我们还包括了一个通用领域的社交媒体数据集 AITA [28] ，其中包含与心理健康无关的文本，作为合并数据集中的非抑郁示例。

我们首先通过一系列预处理步骤对每个数据集进行清理。去除重复条目，并仅保留文本长度位于第 10 和第 90 百分位之间的帖子以排除异常值。对于每个数据集，仅保留与下游任务相关的列。我们的实验设置支持三个分类任务：

1) 二分类抑郁症分类：我们结合所有五个与抑郁相关的数据集和一个额外的非抑郁社交媒体数据集，以训练模型区分抑郁和非抑郁内容。
2) 抑郁严重程度分类：使用 HelaDepDet [25] 数据集，该数据集提供分级的抑郁严重程度标签，我们训练模型来预测包括最低、轻度、中度和重度在内的抑郁程度。
3) 鉴别诊断分类：我们使用 MHB [23] 和 RMHD [26] 数据集，这些数据集包含关于抑郁、焦虑和 PTSD 的多分类注释，以评估模型区分相关心理健康状况的能力。

表格 I 总结了预处理后的数据集的描述性统计，包括帖子数量、标签类别和以单词计量的平均文本长度。对于每个分类任务，相应的数据集被划分为 70 % 的训练集和 30 % 的测试集，两者之间没有重叠，以便能够训练传统的监督分类器。值得注意的是，我们评估的一种方法，基于零样本的大型语言模型（LLM）分类，不需要训练数据，直接应用于测试集。

为了建立基线性能，我们评估了基于两种传统特征表示组合训练的分类器：

我们使用来自 SentenceTransformers 库的 `all-mpnet-base-v2` [29] 模型提取上下文感知的句子级别嵌入。每篇社交媒体帖子都通过预训练模型，以获得一个 768 维的固定大小嵌入向量。

我们使用语言调查和词汇计数（LIWC）词典计算心理语言学特征。每个帖子都会被分析以得出相关类别的标准化频率，包括情感过程、认知过程和代词使用。所得的特征向量通过 z 分数标准化进行标准化。

1) 分类模型： 我们在特征集合的连接上训练了三种传统的机器学习分类器。对数回归使用 L2 正则化（C = 1.0）和最多 1000 次迭代进行训练。支持向量机使用线性核，C = 1.0。随机森林分类器配置了 100 棵决策树，并设定了固定的随机种子以确保可重现性。我们在测试集上计算了准确率，以评估模型性能。

C. 基于零样本大型语言模型的分类

为了评估大型语言模型作为直接零样本分类器的效果，我们使用 OpenAI GPT-4o API。对于每个社交媒体帖子，根据任务设置，我们发送一个请求二分类或多分类标签的提示。二分类任务的提示格式如下：

> 你是一位心理健康专家。阅读以下社交媒体帖子，并判断用户的心理健康状况。从以下标签中选择：抑郁症、非抑郁症。

对于严重程度检测和鉴别诊断，标签选项相应地进行了修改。模型的文本响应被解析为预测的标签。GPT-4o 未经过任何训练或微调，预测结果直接在保留的测试集上生成。

在这种方法中，我们提示 GPT-4o 根据社交媒体帖子的内容来解释和总结用户的心理状态。目标是生成一个简洁、临床导向的描述，以捕捉与心理健康评估相关的信号。用于 LLM 总结生成的提示格式如下：

> 你是一位心理健康专家。请阅读以下社交媒体帖子，并用一到两句话描述用户的心理状态。重点关注情感语气、认知状态以及任何心理健康状况的迹象。避免逐字引用帖子内容。

我们使用 `all-mpnet-base-v2` [29] 句子嵌入模型对 LLM 生成的响应进行嵌入，得到一个 768 维的特征向量。这个向量捕捉了从原始文本中抽象出来的与任务相关的情感语义。然后，我们训练在第 II-B1 节中描述的相同的一组分类器来评估 LLM 生成的改写表示是否改善预测性能。

D. 评估指标

我们使用适合二元和多类设置的标准指标来评估模型在所有分类任务上的表现。对于二元抑郁症分类任务，我们报告在保留测试集上计算的准确率、精确率、召回率和 F1 得分。对于抑郁症严重程度检测和鉴别诊断分类任务，我们关注类内 F1 得分，以便更全面地了解模型在各个类别上的表现。对于所有方法，包括零样本 LLM 分类，我们采用相同的测试集拆分以保持一致性。零样本 LLM 响应被解析为离散标签，并使用与监督模型相同的指标与真实

TABLE I: 各分类任务中使用数据集的统计数据

Dataset	Used For	Size (posts)	Label Categories	Avg. Text Length (words)
MHB [23]	Binary, Differential Diagnosis	7,452	Depression, Anxiety, PTSD	253
CAMS [24]	Binary Only	4,042	Depression	179
HelaDepDet [25]	Binary, Severity Detection	33,498	Depression, Minimum, Mild, Moderate, Severe	120
RMHD [26]	Binary, Differential Diagnosis	658	Depression, Anxiety	236
DepressionEmo [27]	Binary Only	4,830	Depression	95
AITA [28]	Binary Only	24,795	Non-Depression (control)	386
Total (unique)	—	75,275	—	—

TABLE II: 按模型和特征类型划分的二分类指标

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression (Text+LIWC)	0.89	0.94	0.89	0.91
SVM (Text+LIWC)	0.88	0.93	0.88	0.91
Random Forest (Text+LIWC)	0.91	0.96	0.90	0.93
Zero-Shot LLM-Based Classification	0.96	0.97	0.97	0.97
Logistic Regression (LLM Summary)	0.93	0.96	0.93	0.95
SVM (LLM Summary)	0.91	0.95	0.90	0.93
Random Forest (LLM Summary)	0.92	0.96	0.91	0.94

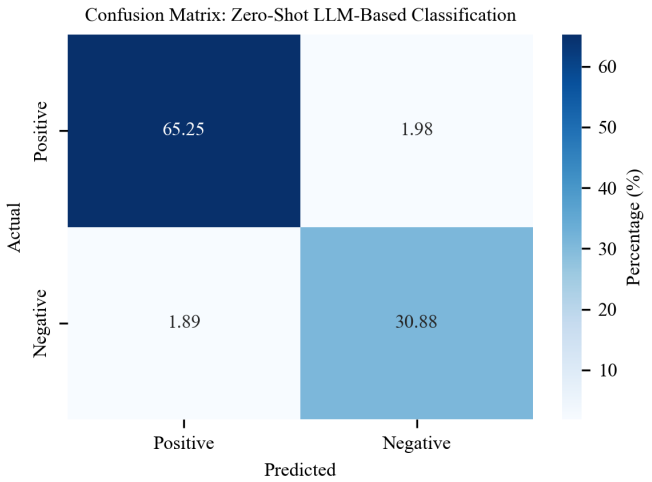


Fig. 1: 零样本 LLM 二分类器在测试集上的混淆矩阵，百分比在所有预测中进行了归一化处理。

标签进行比较。所有指标均使用 `scikit-learn` [30] 库计算。

III. 实验结果

在本节中，我们展示了我们提出的模型在三个心理健康分类任务中的评估结果。每个任务旨在评估模型处理不同类型诊断的能力。首先，我们进行二元抑郁症分类，其目标是使用多个数据集区分抑郁内容和非抑郁内容。接下来，我们研究抑郁症严重程度分类，其中模型使用 HelaDepDet [25] 数据集预测细化的严重程度等级。最后，我们评估模型和嵌入在抑郁症、焦虑症和 PTSD 之间的鉴别诊断分类任务中的性能，使用 MHB [23] 和 RMHD [26] 数据集。每项任务的性能指标将在后续小节中报告和分析。

A. 二元抑郁分类

对于判断社交媒体帖子是否具有抑郁倾向的二分类任务，性能指标汇总在表格 II 中。在所有评估的模型中，零样本 LLM 分类器实现了最高的整体准确率，优于使用心理语言学和基于文本的嵌入的传统机器学习模型，以及基于 LLM 生成的摘要嵌入进行预测的模型。

值得注意的是，这个零样本 LLM 并没有针对抑郁症检测进行专门的微调，但其表现依然优异，这可能归因于其在大规模、多样化数据集上进行的广泛预训练。这证明了模型令人印象深刻的泛化能力，并支持了最近关于 LLM 在零样本环境中表现出色的发现。

对于零样本 LLM 的对应混淆矩阵如图 1 所示。该模型显示出较低的假阳性率和假阴性率，并在精度和召回率之间保持良好的平衡，这表明它不会偏向于通过牺牲一种错误来最小化另一种错误。然而，它确实显示出将非抑郁帖子分类为抑郁帖子的轻微倾向。

使用大型语言模型生成的摘要嵌入的机器学习模型，其性能优于直接从原始社交媒体文本中提取特征的模型。这个结果是意料之中的，因为摘要提供了每个帖子的浓缩的、更高层次的解释，这可能使得隐含的抑郁信号更容易被识别。

B. 抑郁严重程度分类

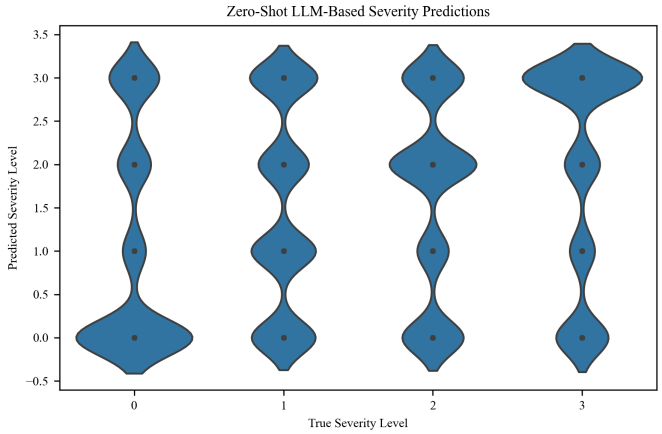
为了评估不同模型和嵌入如何从用户的社交媒体帖子中评估抑郁症的严重程度，我们使用 HelaDepDet [25] 数据集进行了多类分类任务。该数据集包括四个有序标签，表示抑郁严重程度逐步增加的水平，从 0（最低）到 3（严重）。

我们应用了与先前任务相同的分类框架，比较了零样本 LLM（大型语言模型）分类和传统机器学习模型中使用各种特征表示的方法。在所有方法中，使用 LLM 生成的摘要嵌入的逻辑回归分类器取得了最高的准确率 58 %，略微优于使用其他特征的模型，并显著优于零样本 LLM 方法。

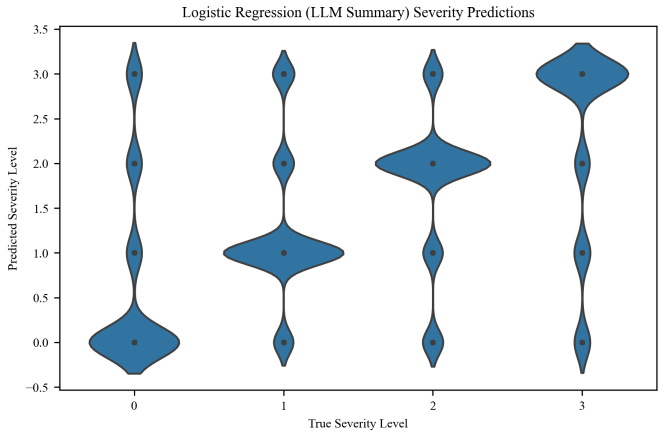
图 2a 和 2b 呈现了针对表现最差和最佳的模型的预测和真实严重程度水平的提琴图。图 2c 显示了每个分类器的每类 F1 分数。我们观察到，零样本 LLM 分类器难以直接从原始文本推断细粒度的严重程度水平，常常无法反映标签的序数结构。相比之下，经过训练的机器学习模型得益于有监督学习，能够捕捉数据中的语义和序数关系，从而能够准确且一致地预测抑郁严重程度。

C. 鉴别诊断分类

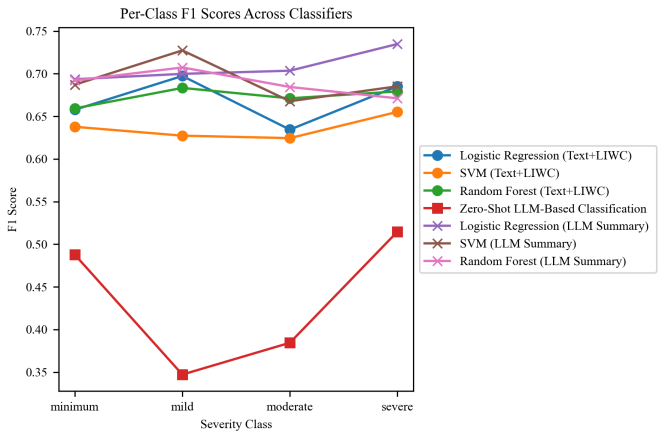
我们使用 MHB [23] 和 RMHD [26] 数据集评估模型在鉴别诊断分类任务上的表现，这些数据集包含抑郁症、PTSD 和焦虑症的多类别注释。该任务评估模型基于社交媒体文本区分相关但临床上不同的心理健康状况的能力。



(a) 零样本大语言模型（LLM）基础的分类器预测的抑郁严重程度与真实标签的比较。



(b) 使用 LLM 摘要嵌入通过逻辑回归分类器预测的抑郁严重程度与真实标签比较。



(c) 按抑郁严重程度和分类器划分的 F1 分数

Fig. 2: 模型在抑郁严重程度分类任务上的性能比较。

在所有评估方法中，零样本 LLM 基于的分类器取得了最高的整体准确率 65 %，略微超过了下一最佳模型——使用 LLM 摘要嵌入的逻辑回归，其整体准确率为 59 %。使用文本嵌入和 LIWC 特征的机器学习模型在所有方法中整体准确率最低。

图 3 展示了所有分类器和所有三种诊断类别的 F1 分数分布。我们观察到，大多数分类器往往将抑郁症与焦虑症混淆，这在意料之中，因为这两种心理健康状况有重叠的语言和情感模式，社交媒体上的表达可以反映这一点。此外，PTSD 通常能够更一致地被区分开，可能是由于更具体的症状语言，例如对创伤的引用，使其与其他两种状况有所不同。

IV. 结论与未来工作

本研究在基于社交媒体数据的抑郁症分类任务中，对零样本大语言模型和传统机器学习模型进行了比较评估。我们发现零样本大语言模型在二元抑郁症分类中表现强劲，证明了其从预训练知识中进行泛化的能力。然而，其在诸如严重程度预测等任务中的表现有所下降，而在这些任务中，使用大语言模型生成的摘要嵌入的监督模型表现出更准确和稳定的性能。

我们的评估结果表明，大型语言模型（LLMs）在心理健康预测任务中表现出色，它们的上下文摘要有助于得出更好的特征。由大型语言模型生成的摘要嵌入捕捉了重要的语义线索，可以改进传统模型以做出更准确和一致的预测。这些发现证明了将大型语言模型的概括能力与基于精心挑选特征训练的轻量级、可解释分类器结合的混合方法的潜力。通过探索先进的提示策略、应用少样本学习以及在领域特定的心理健康数据上微调大型语言模型，可能会获得进一步的性能提升。

References

[1] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, “Machine learning driven mental stress detection on Reddit posts using natural language processing,” Hum-Cent Intell Syst 3, 80–91, 2023.

[2] L. Zhao, J. Jia, and L. Feng, “Teenagers’ stress detection based on time-sensitive micro-blog comment/response actions,” In IFIP International Conference on Artificial Intelligence in Theory and Practice, pp. 26–36, 2015.

[3] B. Desmet, G. Jacobs, and V. Hoste, “Mental distress detection and triage in forum posts: the lt3 clpsych 2016 shared task system,” In Proc. Third Workshop on Computational Linguistics and Clinical Psychology, pp. 148–152, 2016.

[4] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, “Stress detection using natural language processing and machine learning over social interactions,” J Big Data 9, 33, 2022.

[5] N. Li, H. Zhang, and L. Feng, “Incorporating forthcoming events and personality traits in social media based stress prediction,” IEEE Trans. Affect. Comput, 2021.

[6] S. C. Guntuku, A. Buffone, L. Jaidka, J. C. Eichstaedt, and L. H. Ungar, “Understanding and measuring psychological stress using social media,” In Proc. International AAAI Conference on Web and Social Media, vol. 13, pp. 214–225, 2019.

[7] M. H. Kabir, N. Samrat, A. Al Mahmud, R. Akter and M. Raihan, “Mental stress prediction from the text of social media using machine learning techniques,” 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023.

[8] Q. He, B. P. Veldkamp, C. A. W. Glas, and T. de Vries, “Automated assessment of patients’ self-narratives for posttraumatic stress disorder screening using natural language processing and text mining,” Assessment 24, 157–172, 2017.

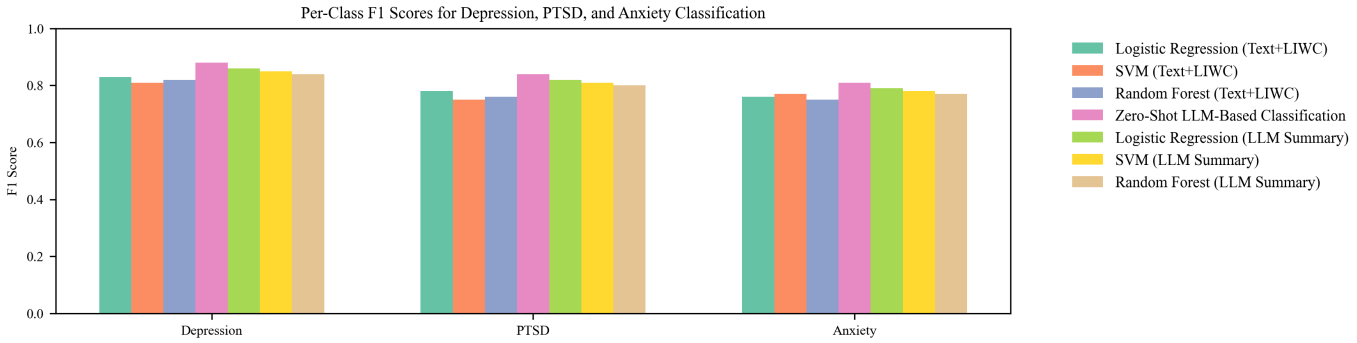


Fig. 3: 多类别心理健康诊断分类中模型性能的比较。

[9] S. Wshah S, C. Skalka C, and M. Price, "Predicting posttraumatic stress disorder risk: a machine learning approach," JMIR Ment Health 2019;6(7):e13946.

[10] D. Marengo, C. M. Hoeboer, B. P. Veldkamp, and M. Olff, "Text mining to improve screening for trauma-related symptoms in a global sample," Psychiatry Research, Volume 316, 2022.

[11] J. R. Oltmanns, H. A. Schwartz, C. Ruggero, Y. Son, J. Miao, M. Waszczuk, S. A. P. Clouston, E. J. Bromet, B. J. Luft, and R. Kotov, "Artificial intelligence language predictors of two-year trauma-related outcomes," Journal of psychiatric research, 143, 239–245, 2021.

[12] J. Sawalha, M. Yousefnezhad, Z. Shah, M. R. G. Brown, A. J. Greenshaw, and R. Greiner, "Detecting presence of PTSD using sentiment analysis from text data," Frontiers in psychiatry, 12, 811392, 2022.

[13] K. Zeberga, M. Attique, B. Shah, F. Ali, Y. Z. Jembre, and T. S. Chung, "A novel text mining approach for mental health prediction using Bi-LSTM and BERT model," Computational intelligence and neuroscience, 2022, 7893775.

[14] M. A. Mansoor and K. H. Ansari, "Early detection of mental health crises through artifical-intelligence-powered social media analysis: a prospective observational study," Journal of personalized medicine, 14(9), 958, 2024.

[15] T. Althoff, K. Clark, and J. Leskovec, "Large-scale analysis of counseling conversations: an application of natural language processing to mental health," Transactions of the Association for Computational Linguistics, 4:463–476, 2016.

[16] M. P. Ewbank, R. Cummins, V. Tablan, A. Catarino, S. Buchholz, and A. D. Blackwell, "Understanding the relationship between patient language and outcomes in internet-enabled cognitive behavioural therapy: A deep learning approach to automatic coding of session transcripts," Psychotherapy research : journal of the Society for Psychotherapy Research, 31(3), 326–338, 2021.

[17] N. Bantilan, M. Malgaroli, B. Ray, and T. D. Hull, "Just in time crisis response: suicide alert system for telemedicine psychotherapy settings," Psychotherapy Research, 31(3), 289–299, 2020.

[18] X. Xu, B. Yao, Y. Dong, S. Gabriel, H. Yu, J. Hendler, M. Ghassemi, A. K. Dey, and D. Wang, "Mental-LLM: leveraging large language models for mental health prediction via online text data," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 8, 1, Article 31 (March 2024), 32 pages, 2024.

[19] L. Boggavarapu, V. Srivastava, A. M. Varanasi, Y. Lu, R. Bhaumik, "Evaluating enhanced LLMs for precise mental health diagnosis from clinical notes," medRxiv, 2024, unpublished.

[20] M. Malgaroli, K. Schultebraucks, K. J. Myrick, A. A. Loch, L. Ospina-Pinillos, T. Choudhury, R. Kotov, M. De Choudhury, and J. Torous, "Large language models for the mental health community: framework for translating code to care," The Lancet. Digital health, 7(4), e282–e285, 2025.

[21] K. Qian, H. Zhang, X. Jing, B. Hu, Y. Yamamoto, and B. W. Schuller, "Foundation models for digital mental health: igniting the dawn," Medicine Plus, Volume 2, Issue 2, 2025.

[22] Y. Hua, H. Na, Z. Li, F. Liu, X. Fang, D. Clifton, and J. Torous, "A scoping review of large language models for generative tasks in mental health care," npj Digit. Med. 8, 230, 2025.

[23] S. Boinepelli, T. Raha, H. Abburi, P. Parikh, N. Chhaya, and V. Varma, "Leveraging mental health forums for user-level depression detection on social media," In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 5418–5427, Marseille, France. European Language Resources Association, 2022.

[24] M. Garg, C. Saxena, S. Saha, V. Krishnan, R. Joshi, and V. Mago, "CAMS: an annotated corpus for causal analysis of mental health issues in social media posts," In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 6387–6396, Marseille, France. European Language Resources Association, 2022.

[25] Y. H. P. P. Priyadarshana, Z. Liang, and I. Piumarta, "HelaDepDet: a novel multi-class classification model for detecting the severity of human depression," In Collaboration Technologies and Social Computing: 29th International Conference, CollabTech 2023, Osaka, Japan, August 29–September 1, 2023, Proceedings. Springer-Verlag, Berlin, Heidelberg, 3–18.

[26] S. Rani, K. Ahmed, and S. Subramani, "From posts to knowledge: annotating a pandemic-Era Reddit dataset to navigate mental health narratives," Applied Sciences, 2024; 14(4):1547.

[27] A. B. S. Rahman, H. Ta, L. Najjar, A. Azadmanesh, A. S. Gönul, "DepressionEmo: a novel dataset for multilabel classification of depression emotions," Journal of Affective Disorders, Volume 366, 2024, Pages 445-458.

[28] A. Alhassan, J. Zhang, and V. Schlegel, " 'Am I the Bad One' ? predicting the moral judgement of the crowd using pre–trained language models," In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 267–276, Marseille, France. European Language Resources Association, 2022.

[29] K. Song, X. Tan, T. Qin, J. Lu, and T. Liu, "MPNet: masked and permuted pre-training for language understanding," In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS '20), Curran Associates Inc., Red Hook, NY, USA, Article 1414, 16857–16867, 2020.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: machine learning in Python," J. Mach. Learn. Res. 12, 2825–2830, 2011.