利用深度学习可视化检测社交媒体用户抑郁倾向变化

杨雨坤[1] 乔晚馨[2] 林钰婷[1] 刘蕊[1]
（1.北京理工大学管理学院，2.北京工业大学经济与管理学院）

摘要

研究目的：利用深度学习模型检测社交媒体用户抑郁倾向，并可视化其情绪变化轨迹，解决传统抑郁症检测方法存在主观偏见、实现成本高和可解释性有限等问题。

研究方法：笔者提出了一种基于 LSTM (Long Short-Term Memory)与多示例学习的深度学习模型，该模型采用自动编码器从文本中提取时间序列特征，构建二元分类器筛选抑郁症相关的推文，确定用户抑郁倾向，并基于时间序列数据的分布生成可视化图谱，以展示用户情绪的动态变化。

研究结果：我们的模型不仅能实现传统的基于社交媒体用户推文对其是否抑郁进行分类功能，还可以输出用户每条推文的抑郁倾向“分数”并可视化，且其性能优于基本方法。

边际贡献：本研究提出了一种低成本且易于扩展的大规模抑郁症筛查方案，突破了传统单时间点检测的局限，实现了抑郁倾向的动态可视化监测。

关键词：深度学习、社交媒体、抑郁倾向检测、LSTM 自动编码器
文章类型：研究论文

1 引言

抑郁症是全球公共卫生领域的重大挑战(Yu *et al.*, 2020)，其典型症状包括持续性抑郁情绪、睡眠障碍、注意力涣散以及对生活兴趣减退(American Psychiatric Association, 2013)。严重时可能引发躯体症状。该疾病的成因复杂，涉及社会压力、心理健康问题、历史创伤及躯体疾病等多种因素。2005 至 2015 年间，全球抑郁症患者数量激增 18.4%，确诊人数突破 3.32 亿大关(World Health Organization, 2017)。随着患者数量攀升，自杀率也同步走高。世界卫生组织(2023a)数据显示，全球每年有超过 70 万人死于自杀，这一数字已超越战争、他杀和疟疾致死人数，成为人类主要死因之一。2019 年数据显示，约 1.3%的死亡案例由自杀造成。更值得关注的是，自杀已成为 15-29 岁人群第二大死因(World Health Organization, 2019)，青少年群体中抑郁症引发的自杀念头日益普遍。因此，早期识别抑郁症状对降低自杀风险也具有关键意义。

传统抑郁症检测方法依赖标准化量表测量、患者主观报告及主治医师的临床诊断。但是这些方法存在明显缺陷：例如，参与者对量表的反应可能受当时环境、情绪和心理状态影响，医患关系及对疾病的羞耻感可能导致症状漏报，造成量表假阴性结果(Liu *et al.*, 2022)；大规模实施量表测量也缺乏时间粒度和成本高昂的挑战(Kumar *et al.*, 2015)。随着微博、推特等社交媒体的发展，人们能在此表达真实情感与观点。凭借匿名性、实时性和海量用户生成数据，社交媒体使研究人员能够收集推文内容，为心理健康检测提供新视角(Biradar and Totad, 2019)，特别是年轻人更倾向于在匿名环境中通过社交媒体表达个人想法与负面情绪(Ma *et al.*, 2016)。数据显示，20%的自杀未遂者和 50%的自杀身亡者会留下遗言(Dejong *et al.*, 2010)。

与此同时，考虑到专业规模评估的局限性（如成本高昂、无法支持大规模抑郁症筛查），越来越多的研究人员开始探索利用社交媒体数据进行抑郁症检测。通过分析用户发布的动态，这些研究旨在实现更可扩展的筛查，并推动早期干预。

本研究旨在构建多示例长短期记忆网络模型，并利用微博公开文本数据集识别抑郁用户，同时可视化抑郁倾向的变化趋势，其大致结构如下：1.基于微博数据构建词嵌入向量，生成输入 LSTM (Long Short-Term Memory)自编码器的初始向量，并获取经过 LSTM 层处理后的输出向量。2.将输出向量输入预训练二元分类器，通过对比三种基准模型，获得最优分类器下每条推文对应的抑郁概率。3.采用多示例学习策略，设定抑郁判定阈值，判断用户是否处于抑郁状态。4.根据二元分类得出的抑郁概率，绘制用户抑郁倾向分布图，直观展现动态变化趋势。

2 文献综述

2.1 识别抑郁的传统方法

据估计，全球约有 3.8%的人口患有抑郁症，其中成年人占比 5%，60 岁以上老年人群中占比 5.7% (World Health Organization, 2023b)。抑郁症已成为心理学和医学领域的重要研究课题。传统抑郁症诊断方法主要依赖量表、心理咨询和临床诊断。

抑郁检测量表作为普遍的诊断手段，包括自测量表与他测量表两种类型。常见的自测量表包括 PHQ-9 (Kroenke *et al.*, 2001)、SPS (Social Phobia Scale)、DASS-21 (Crawford and Henry, 2003)、BDI (Beck, 1961)等；他测量表包括 HAMD/HDRS (Hamilton Depression Rating Scale) (Hamilton, 1960)、MADRS (Montgomery-Asberg Depression Rating Scale) (Montgomery and Åsberg, 1979)、Suicidal Affect Behavior-Cognition Scale (Harris *et al.*, 2015)等。自测量表要求患者根据自身情绪和行为习惯进行自我评估，而他测量表则需由专业心理学家或精神科医师操作。使用标准化抑郁量表的优势在于操作简便、快速且可量化评估。但这类量表也存在局限性，例如个体对自身情绪和行为的主观评估可能受记忆偏差、情绪波动及社会期望偏见的影响；研究表明进行自杀评估可能对抑郁症患者产生负面影响(Harris and Goh, 2017)。

心理咨询也是识别抑郁症的常用方法。通过面对面访谈，心理学家能够评估个体的情绪状态、生活习惯和行为模式。这种方法通常具有较高的准确性，但其实施高度依赖诊断者的专业能力，且患者情绪表达可能受限，导致适用性较差。

临床诊断通常依据《精神疾病诊断与统计手册第五版》(DSM-5, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) (American Psychiatric Association, 2013)等权威指南进行，主要通过评估患者症状和生活史展开。这种模式具有高度权威性和专业性，但实施效率低、成本高且普及度有限。

当前人们仍主要依赖传统方法识别抑郁症，但它们存在如下缺陷：1.部分抑郁患者可能未主动寻求医疗帮助，导致错过最佳治疗时机；2.在使用量表或接受心理咨询时，患者的量表准确性易受患者主观因素影响；3.多数传统方法都只能提供静态的时点评估，缺乏监测抑郁情绪波动所需的时间粒度；4.在许多地区，获取心理健康资源的渠道仍然相对有限。这些缺陷凸显了开发低成本且时间敏感的抑郁症检测方法的需求——尤其是那些能充分利用社交媒体平台海量实时行为数据的方案。

2.2 社交媒体的崛起与抑郁情感表达

近年来，社交媒体的迅猛发展对情感表达和心理健康也产生了深远影响，人们常常在社交媒体上分享自己的真实感受和生活状态，这为抑郁检测提供了新的可能。

Statista (2025)的数据显示，截止至 2025 年 2 月，全球互联网用户数量达到 55.6 亿，占全球总人口的 67.9%，其中社交媒体用户数量达到 52.4 亿，占全球总人口的 63.9%。社交媒体（如 Facebook、Twitter、Instagram、Weibo 和 Reddit）已经成为人们日常生活的一部分。研究表明，社交媒体中的数据能包含抑郁症患者发布出的消极信号，这为抑郁识别提供了重要的数据源(De Choudhury *et al.*, 2013)，如 Facebook、Twitter 上的用户通过发布表达孤独、焦虑等负面情绪来暗示其存在抑郁症状(Guntuku *et al.*, 2017)。此外，社交平台通过文字、图片及视频等形式，提供多维度的情感表达途径，为抑郁情绪识别提供更多信息。

2.3 基于社交媒体数据的抑郁识别方法

随着机器学习和深度学习技术的发展，研究人员开始尝试使用自动化的方式识别抑郁情绪。这些方法主要依赖于文本分析、图像识别或音视频分析等特征提取手段，通过算法进行情绪识别。

Wei 等人(2023)提出了一种基于子注意力机制的多模态融合方法，该方法整合了音频、视觉和文本三种模态，在针对重度抑郁症检测的 DAIC-WOZ (Distress Analysis Interview Corpus - Wizard of Oz)数据集上进行评估时，该方法达到了 0.89 的精确率和 0.70 的 F1 分数。Yan 等人(2025)提出了一种名为“Depressive Emotion-Context Enhanced Network (DECEN)”的深度学习模型，该模型结合了抑郁情绪识别模块和上下文感知表示机制，以提高从社交媒体内容中检测抑郁症的能力。Roy 等人(2020)提出一种名为“Suicide Artificial Intelligence Prediction Heuristic (SAIPH)”的算法，将 Twitter 数据输入一系列神经网络的，并用输出训练随机森林模型，AUC (area under curve)达到 0.88。Mann 等人(2021)将抑郁症识别任务转化为多示例学习问题，并利用多模态社交媒体数据进行检测。Tapotosh Ghosh 等人(2023)提出一种基于注意力的 BiLSTM-CNN 的模型，来检测孟加拉语社交媒体文本的抑郁状态。Guo 等人(2023)聚焦中国社交媒体，提出了一种基于领域知识的抑郁症检测方法。他们构建了符合中文语言特性的抑郁症相关词汇表，并提取了抑郁症相关术语的频率和情感倾向等特征。这些特征通过基于相关性的加权融合后，被输入到多个机器学习模型中。Peng 等人(2019)提出一种基于多核支持向量机的方法，利用用户社交媒体数据提取三类特征，并构建情感词典，使用 TF-IDF (Term Frequency-Inverse Document Frequency)提取微博文本词频统计，所有特征被合并起来分类用户是否抑郁。Zhang 等人(2024)提出了一种名为 DKDD(Deep Knowledge-aware Depression Detection)的知识感知型深度学习模型，该模型通过实体识别、医学本体对齐及注意力机制，从社交媒体用户的数据中提取具有临床意义的实体及其时间分布特征。Tadesse 等人(2019) 基于 Reddit 数据检测在线用户抑郁情绪的相关因素，他们在模型中使用了 LIWC (Linguistic Inquiry and Word Count)、LDA (Latent Dirichlet Allocation)、N-gram 和其他经典的机器学习算法。Deshpande 等人(2017)基于 Twitter 平台，将推文分为中性或负面，以检测抑郁倾向，并使用朴素贝叶斯和支持向量机作为对比分类器。Yoon 等人(2022)构建了一个包含 951 个视频博客的多模态抑郁症数据集，并采用音频-视频跨模态注意力机制来捕捉不同模态间的特征关联。Kumar 和 Venkatram (2024)基于 Twitter 数据集，使用包括年龄、性别、粉丝数、在线时长等特征，预测自杀行为。该研究提出了一种基于规则的分类算法，通过快速排序方法确定最佳分割点，构建决策树。Islam 等人(2018)使用 Facebook

数据，利用 LIWC 处理文本数据，引入 KNN (K-Nearest Neighbors)对文本进行抑郁分类。Chen 等人(2025)提出了一种基于跨模态特征重构与解耦的文本引导多模态抑郁症检测框架。该模型通过从联合文本-音频嵌入中分离出共享和潜在表示，并采用双向交叉注意力模块来增强跨模态交互。Uddin 等人(2019)采用了基于 GRU (Gate Recurrent Unit)的方法，创建了 GRU 和 Dense 层的几种组合方式，并最终找到了具有 512 个神经元的 5 层 GRU 作为最优模型。Cai 等人(2023)提出了一种基于用户抑郁症状的多变量时间序列特征的抑郁症检测方法。该方法通过从社交媒体帖子中构建用户的情感行为序列，提取了包括发帖频率、情感波动和抑郁相关关键词在内的多个维度的动态特征。这些特征随后被输入到分类器中进行建模。实验结果表明，这种方法在捕捉用户抑郁状态的时间变化方面非常有效。

已有研究在利用社交媒体数据进行抑郁症检测方面已取得显著进展，但仍面临一些局限性。首先，许多方法依赖静态特征聚合，无法捕捉用户情感随时间的变化，这限制了它们反映抑郁症症状发展变化的能力。其次，虽然多模态模型显示出潜力，但这些模型通常依赖于昂贵或难以收集的数据类型，如音频和视频，这使得它们在大规模应用中不太可行。而且，尽管一些模型融入了“领域知识”，但这种整合通常仅限于手工构建的词汇表或简单的加权方案，缺乏深度的语义理解和广泛的适用性。此外，大多数现有研究仅关注最终的分类性能，而忽视了用户情感变化的可解释性。这些不足之处要求我们开发一种基于社交媒体的抑郁症检测框架，该框架需要更加时间敏感且易于理解。

2.4 本研究边际贡献

为解决上述局限性，本研究提出了一种基于社交媒体的抑郁症检测新框架，整合了时序建模、弱监督学习和可视化解读三大核心要素。具体而言，我们开发了多示例 LSTM 学习模型，通过 LSTM 自编码器无需依赖帖子标签即可提取用户社交媒体内容的时间特征。采用弱监督学习策略时，利用用户层面的标注数据指导二元分类器训练，输出每条帖子的抑郁倾向概率，从而实现对情绪变化趋势的精细评估。此外，多示例学习机制用于整合帖子层面的预测结果，确定用户抑郁症状态；同时设计的时间序列可视化模块可直观呈现用户抑郁倾向的动态演变轨迹。

相较于已有方法，本框架可捕捉纵向情绪变化模式，避免依赖高成本模态或人工标注，显著提升了系统的可扩展性和解释性。通过深度时序特征提取、弱监督学习与可视化分析的有机结合，本研究为基于社交媒体数据的大规模抑郁症筛查提供了经济高效、可扩展且具有洞察力的解决方案。

3 模型构建

与现有深度学习模型相比，我们的研究提出了一种新型抑郁症检测框架——多示例 LSTM 学习模型。该模型首先从用户发布的帖子中提取时间序列特征，评估帖子层面的抑郁风险概率。随后基于多示例学习（Multi-instance Learning, MIL）原理，模型会预测用户是否处于抑郁症风险状态，并可视化其抑郁倾向随时间演变的过程。

本研究的优势之一在于仅需依赖粗粒度的用户级标注数据，无需进行成本高昂且主观性强的推文级标注（这类标注在实际场景中也难以获取）。基于多示例学习的核心理念，我们将同一用户的全部帖子视为多个实例，而用户本身则被建模为一个语义包。该语义包会被标记为抑郁或正常状态，而各实例仍保持未标注状态。模型完全基于用户级标签进行训练，从

而在弱监督条件下实现抑郁症检测。

具体而言，所提出的方法包括三个关键步骤：

1.采用无监督 LSTM 自编码器对每篇帖子进行建模并提取时间特征。该模块保留了帖子的语义和情感信息，从而增强了模型捕捉与抑郁症相关的行为特征的能力。

2.将 LSTM 输出的帖子后层表征输入二元分类器，预测每篇帖子的抑郁概率，随后在多示例框架下对这些概率进行聚合，推断用户的整体心理状态。

3.设计可视化模块，绘制用户帖子级抑郁概率的时间动态，提供其精神状态随时间波动的直观表现。

总体而言，我们的方法在低标签要求、强表征能力和高可解释性之间取得了平衡，且在比较实验中优于几个现有的基线模型。

3.1 模型总体结构

首先，让我们给出要研究问题的定义：设有 $N = N_d + N_u$条标注的推文数据，包括 P_d 个抑郁倾向用户的 N_d 条推文数据 $\boldsymbol{TD} = \{\boldsymbol{td_1}, \boldsymbol{td_2}, \ldots, \boldsymbol{td_{Nd}}\}$ 和标签数据 $\boldsymbol{YD} = \{yd_1, yd_2, \ldots, yd_{Nd}\}$，以及 P_u 个无抑郁倾向用户的 N_u条推文数据 $\boldsymbol{TU} = \{\boldsymbol{tu_1}, \boldsymbol{tu_2}, \ldots, \boldsymbol{tu_{Nu}}\}$ 和标签数据$\boldsymbol{YU} = \{yu_1, yu_2, \ldots, yu_{Nu}\}$，其中$\boldsymbol{T} = \boldsymbol{TD} \cup \boldsymbol{TU} = \{t_1, t_2, \ldots, t_N\}$，$\boldsymbol{Y} = \boldsymbol{YD} \cup \boldsymbol{YU} = \{y_1, y_2, \ldots, y_N\}$。期望采用非监督方法对$\boldsymbol{TD}$和$\boldsymbol{TU}$分别抽取时间序列特征$\boldsymbol{SD} = \{\boldsymbol{sd_1}, \boldsymbol{sd_2}, \ldots, \boldsymbol{sd_{Nd}}\}$和 $\boldsymbol{SU} = \{su_1, su_2, \ldots, su_{Nu}\}$，其中 $\boldsymbol{S} = \boldsymbol{SD} \cup \boldsymbol{SU} = \{\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_N}\}$，并通过这些数据训练二值分类器 F，用于未来对其他相关人员进行抑郁症的分类检测。

该方法的结构如图 1 所示：首先，基于给定用户的时间轴推文 T，将其标记化为可用的推文编码 X；其次，使用无监督的 LSTM，将每个用户的推文编码为时间序列特征 S；然后，使用训练好的二进制分类器，对时间序列特征进行分类；接着，利用多示例学习来检测给定用户的某一特定推文是否被判定为抑郁，并且遍历其每一条推文；最后，模型输出用户各条推文的抑郁倾向概率，并以时间序列图的形式展示其动态变化。

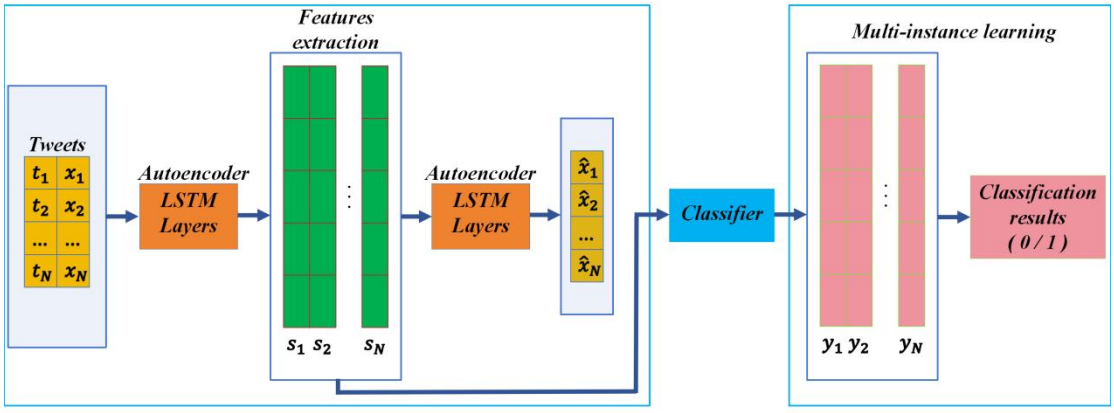

图 1 模型总体结构

3.2 特征提取

为了训练出多示例学习模型，使用时间序列特征作为输入，该特征基于每个用户的时间轴推文。众所周知，时间序列数据是指按时间顺序索引的数据点序列，因此可用无监督的 LSTM 作为从每个推文中提取时间序列特征的方法。在这种情况下，为了表示用户时间轴推文中的时间序列信息，利用无监督的 LSTM (也称为 LSTM 自动编码器)，从每个推文中提取了一组向量 $\boldsymbol{s}_i$ 。其中，自动编码器是一种重建式的神经网络，以无监督的方式学习每条推文的矢量化表示。

LSTM 自动编码器至少需要两个 LSTM 层。以两个 LSTM 层模型为例，第一层可以视为编码器，第二层可以视为解码器。首先，模型通过第一 LSTM 层输入矢量化的推文，以输出形状良好的向量; 然后，将此向量作为输入，以便通过第 2 个 LSTM 层，使输出具有与输入向量化推文相同的形状; 最后，将对模型进行优化，以使输入和输出尽可能相似。因此，LSTM 自动编码器将来自输入层的标记化推文 $\boldsymbol{x}_i$ 压缩为格式良好的代码 $\boldsymbol{s}_i$ ，然后将代码解压缩为向量 $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$ 的形式。

LSTM 自动编码器的步骤如下:

1.推文标记化。由于词干推文是一种字符串，因此自动编码器的第一步是在输入之前对推文进行标记化。为了使每个推文具有相同的尺寸，使用单词嵌入方法。此方法首先计算推文的最大词语数 θ ，然后在 tweet 向量前面嵌入 0，以使每个推文具有相同的维数。通过这样的方式，可以将推文 $T = \{t_1, t_2, \ldots, t_N\}$ 标记为 $X = \{x_1, x_2, \ldots, x_N\}$ 。

2.LSTM 自编码器。输入层顺序输入标记化的文本序列 x_i，进入到 LSTM 层中进行处理并编码为序列编码 $\boldsymbol{s}_{\mathrm{i}}$ ，然后这些序列编码再通过 LSTM 层进行解码为 $\hat{x}_i$ 。其中，LSTM 自动编码器层旨在生成与输入标记化推文具有相同形状的矩阵。因此，可以将输入和输出之间的差定义为损失函数，有：

$$loss = \sum_{i=1}^{N} ||x_i - \hat{x}_i||_2 \quad (1)$$

为了使输入与输出相似，该自动编码器力求使损失函数最小化。

3.3 分类器训练

考虑给定的数据集，其中非抑郁的样本数量大于抑郁的样本数量。此外，根据标签的规定，用户的推文包含严格的模式，例如“我被诊断为抑郁症”，才会被标记为抑郁。因此，多示例学习是一种合适的机器学习方法。另外，在这项工作中，抑郁症检测模型用来估算用户抑郁症的可能性。在机器学习领域，支持向量机和逻辑回归是二进制分类领域的基本方法。因此，采用了这两种方法来训练分类器。

1.支持向量机(Support Vector Machine，SVM)。本研究还训练了两个具有不同内核功能的 SVM 模型。根据 Representer 定理，SVM 中的参数可以写成训练数据的线性组合，那么最终的分类器可以表示如下:

$$f(\boldsymbol{s}_i) = \sum_{j=1}^{N} \alpha_i \gamma_i \varphi(s_j)^T \varphi(s_i) + b \quad (2)$$

其中， $\varphi(s_j)^T \varphi(s_i)$ 为核函数，也可以表示成 $K(s_i, s_j)$ 。

根据经验，采用了线性核(L-SVM, Linear-Support Vector Machines)和 RBF 核(R-SVM,

Radial Basis Function-Support Vector Machines)。其中，线性核的核函数可以表示为：

$$K(s_i, s_j) = s_i{}^T s_j \quad (3)$$

此外，RBF 核的核函数可表示为：

$$K(s_i, s_j) = exp(-\frac{||\boldsymbol{s}_i - \boldsymbol{s}_j||_2}{2\sigma^2}) \quad (4)$$

2.逻辑回归(Logistic Regression，LR) 。根据这一概念，逻辑回归分类器$f(\boldsymbol{s}_i)$可以定义如下:

$$f(\boldsymbol{s}_i) = \frac{1}{1+e^{-(\boldsymbol{W}^T s_i + \boldsymbol{b})}} \quad (5)$$

其中，$\boldsymbol{W}$、$\boldsymbol{b}$是待学习的参数。

同时，为了测量二元分类器的逻辑回归模型的学习损失，这里采用交叉熵损失函数，可以表示如下：

$$\boldsymbol{loss}(f(\boldsymbol{s}_i), y_i) = \frac{1}{N}\sum_{i=1}^{N}[y_i lg\, f(\boldsymbol{s}_i) + (1 - y_i) lg(1 - f(\boldsymbol{s}_i))] \quad (6)$$

式中，y_i 为 $\boldsymbol{s}_i$ 所对应的真实标签。

为了训练分类器，采用了基本的优化器算法，即梯度下降法，该方法也称为最速下降法，以寻找损失函数的最小值。

3.4 多示例学习

在完成抑郁推文检测模型的训练后，下一步需要预测微博用户是否处于抑郁状态。为解决数据集不平衡问题，我们采用多示例学习(Multi-instance learning)框架构建用户层面的检测模型。本文提出的算法结构具体如下：

1.经过自编码，用户 u_i 推文的时间序列特征被提取为 $\boldsymbol{S}_i = \{s_{i1}, s_{i2}, \ldots, s_{ij}, \ldots, s_{in}\}$，其中 n 为用户的推文数。采用选定的$f(\boldsymbol{s_i})$来检测每个特征，输出结果表示为 $\widehat{X}_i = \{\hat{x}_{i1}, \hat{x}_{i2}, \ldots, \hat{x}_{ij}, .. \hat{x}_{in}\}$，$j = 1, 2, 3, \ldots, n$。

2.下一步，遍历 $\widehat{X}_i = \{\hat{x}_{i1}, \hat{x}_{i2}, \ldots, \hat{x}_{ij}, .. \hat{x}_{in}\}$ 中的各个元素。我们定义函数 $T(\hat{x}_{ij})$ 并外生添加一个阈值参数 p，若某一特定 $\hat{x}_{ij} > p$，则 $T(\hat{x}_{ij}) = 1$；反之若 $\hat{x}_{ij} \leq p$，则 $T(\hat{x}_{ij}) = 0$。然后再将 n 个 $T(\hat{x}_{ij})$ 加总为 l。

$$T(\hat{x}_{ij}) = \begin{cases} 1, \hat{x}_{ij} > p \\ 0, x_{ij} \leq p \end{cases},\ l = \sum_{j=1}^{n} T(\hat{x}_{ij}) \quad (7)$$

3.最后，将用户 u_i 分类为抑郁与非抑郁。再外生添加一个权重参数 ω，如果该用户预测的推文被判定为抑郁的比重大于所有时间轴推文的权重，即 $l > \omega \times n$，则将其分类为抑郁；反之若 $l \leq \omega \times n$，则将其分类为非抑郁。因此，该算法的最终结果可以定义如下:

$$\widehat{\boldsymbol{Y}}_i = \begin{cases} 1\ , l > \omega \times n \\ 0\ , l \leq \omega \times n \end{cases} \quad (8)$$

以上 $\widehat{\boldsymbol{Y}}_i = 1$ 代表用户 u_i 被分类为抑郁；$\widehat{\boldsymbol{Y}}_i = 0$ 代表用户 u_i 被分类为非抑郁。

4 实验结果

4.1 数据集描述

为了验证本文提出的方法，我们使用了一个名叫 WU3D (Wang *et al.*, 2020)的公开数据集，它从微博上搜集大量的抑郁和正常用户的推文并打上标签。每个用户样本包含用户的昵称、帖子、发帖时间、用户性别。对它的总体统计如表 I 所示。我们从中随机选取 2000 个用户（包括 400 个抑郁用户和 1600 个正常用户）作为实验样本，每次将他们按照 8：2 的比例随机分为训练集和测试集。

表 I. WU3D 数据集描述

标签	用户数	推文总数	用户平均推文数
正常	22245	1564349	70.32
抑郁	10325	372377	36.07
合计	32570	1936726	59.46

4.2 模型可视结果

在多示例学习中，我们提取了每个用户推文的时间序列特征，并检测每个特征。其中用户 u_i 的时间序列特征 $S_i = \{s_{i1}, s_{i2}, \ldots, s_{ij}, \ldots, s_{in}\}$，输出结果为 $\hat{X}_i = \{\hat{x}_{i1}, \hat{x}_{i2}, \ldots, \hat{x}_{ij}, .. \hat{x}_{in}\}$。不妨设一共有 m 个用户，依次完成对对他们的检测后，分别找出 $max(\hat{x}_{ij})$ 和 $min(\hat{x}_{ij})$, $i = 1,2,\ldots,m, j = 1,2,\ldots,n.$ 接下来对每个 $\hat{x}_{ij}$ 进行归一化处理得到归一化预测值 $\widetilde{\hat{x}_{ij}}$ ，如式(9)所示。

$$\widetilde{\hat{x}_{ij}} = \frac{\hat{x}_{ij} - min(\hat{x}_{ij})}{max(\hat{x}_{ij}) - min(\hat{x}_{ij})} \quad (9)$$

下一步，得到归一化后的向量。用户 u_i 的归一化向量表示为$\widetilde{\hat{\boldsymbol{X}}_i} = \{\widetilde{\hat{x}_{i1}}, \widetilde{\hat{x}_{i2}}, \ldots, \widetilde{\hat{x}_{ij}} \ldots, \widetilde{\hat{x}_{in}}\}, j = 1,2,\ldots,n. \forall \widetilde{\hat{x}_{ij}} \in [0,1].$ 最后，在图像中输出向量各元素值。

我们通过可视化分析对比了大量被标记为抑郁或非抑郁用户的抑郁症概率随时间变化趋势。结果显示，抑郁用户的预测概率波动较大，表明其抑郁症状随时间推移更为不稳定。相比之下，正常用户的抑郁症概率保持相对稳定，多数数值低于 0.3。这些发现与 Cai 等人(2023)和 Seabrook 等人(2018)的前期研究结果一致。

为便于演示，我们从数据集中随机选取一位抑郁症患者和一位非抑郁症用户，分别展示其抑郁概率评分随时间变化的可视化结果。具体推文内容及其发布时间戳分别列于表II和表III中，可视化图表则呈现于图 2 和图 3。

选中的抑郁症用户的推文记录共有 10 条，时间跨度为 2019 年 5 月 8 日至 6 月 12 日。抑郁概率评分范围在 0.22 到 0.97 之间，差异达 0.75，这表明数据波动较大且变化趋势更为剧烈。选中的非抑郁症用户则有 18 条推文记录，时间跨度从 2019 年 9 月 18 日延续至 2020 年 3 月 21 日，多数预测值保持在 0.2 以下。整体波动幅度较小，概率变化趋势显得更加稳

定和规律。

此外，通过分析推文内容与模型预测抑郁概率之间的关联性，我们发现两者存在显著对应关系。那些包含明显负面情绪或心理困扰表达的推文往往获得更高预测值。例如，用户“最后最后我还是选择了药物治疗”的推文被赋予 0.97 分，表明其与抑郁状态存在强关联。相比之下，描述日常作息、情感中立或积极体验的推文通常抑郁概率较低。以非抑郁用户发布的“今日份晚餐”为例，该推文仅获得 0.08 分。这说明模型能够捕捉语言中的情感线索，从而将特定语言特征与抑郁症发生概率相关联。

表II. 某被标记为抑郁的用户的推文内容

Order number	The time tweet was posted	Content of the tweet	Normalized predicted values $\widetilde{x}$
1	2019-06-12 12:25:43	狗咬你一口 你会选择怎么办？	0.22
2	2019-06-09 00:45:04	最后最后 我还是选择了药物治疗	0.97
3	2019-06-04 03:16:27	本以为回家会好的 可是在家里越来越严重 我真的不知道是自己心里固执的抗争 还是身体的毛病 是真的整夜整夜的失眠啊 失眠打卡四十三天 中卫	0.90
4	2019-05-21 19:36:30	小人得势 他呢就悲贱的像条狗 你不知道他跳起来摇尾巴的时候有多乖 兄弟 我劝你做个人 生而为人 劝你善良	0.35
5	2019-05-21 18:50:26	又是一夜无眠 还不如早起看太阳	0.71
6	2019-05-15 23:50:41	其实没有人影响你的情绪是你自己放不过你自己	0.76
7	2019-05-09 01:24:53	十多岁的孩子承受了那么多	0.63
8	2019-05-08 01:01:56	不是因为患有抑郁症才会导致精力很弱 是因为成天失眠 抑郁症患者的精力很弱吗	0.94
9	2019-05-08 01:01:03	让自己忙起来 抑郁症如何走出心理阴影	0.96
10	2019-05-08 00:55:23	我真的希望在与抑郁症抗争的日子里看得到那种从心底里散发出来的阳光☀而不再是所有的阴暗同时不再与失眠继续斗争不再惧怕黑夜的来临	0.93

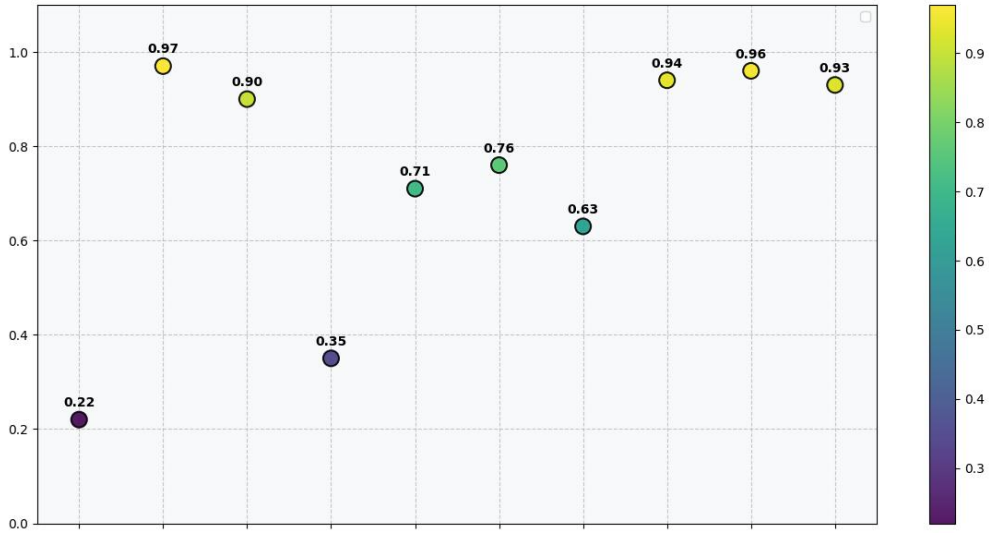

图 2. 某被标记为抑郁的用户的可视化结果

表III. 某被标记为非抑郁的用户的推文内容

Order number	The time tweet was posted	Content of the tweet	Normalized predicted values $\tilde{\tilde{x}}$
1	2019-9-18 22:45	最喜欢医院每天下午的加餐了	0.64
2	2019-9-24 22:49	中午的饭忘了拍 不是很满意今天的饭 点了两顿外卖 但是南瓜汤真的是好好喝 明天正好要来 不过关系不大本来就打算看阅兵 提前煮好了明日份奶茶 加点龟苓膏就可以愉快的吃吃喝喝啦	0.23
3	2019-9-25 23:11	这是昨晚的加餐	0.05
4	2019-9-27 8:00	早餐照样是这么简单 午餐点了老汤烧鸭面 幸好自己又准备了份水煮菜 不然今日份蔬菜都没吃够 现在晚上天天都顶胃 晚上就梨➕奶茶解决了 奶茶里加了冰糖	0.08
5	2019-9-28 23:19	月子中心的晚餐和加餐 完全有理由怀疑会回到生前体重	0.11
6	2019-9-29 22:51	以前和我爸只有有事的时候才聊天 有了小汤圆之后就 哈哈哈	0.09
7	2019-9-30 22:08	先记录下昨天 昨天太累了 准备下午空腹四小时抽血 所以早上吃好点 抽完血吃了两个贝果 晚上点的牛肉炒饭料很多 可是好吃是需要付出代价的 然后就躺下睡了	0.44
8	2019-10-1 19:59	无痛真的是人类之光	0.25
9	2019-10-2	头发洗了真酥服	0.06

	13:46		
10	2019-10-3 18:04	酸汤肥牛底下还有你家的豆皮	0.09
11	2019-10-4 20:17	老王买的橙子真是酸倒我了 还不如贝果甜 贝果还是喜欢放平底锅煎一下表面脆脆的 午餐晚餐都超满足的 晚饭的乌冬面只放了一半蔬菜 剩下的可以明天吃 上午去剪了头发 不忍心剪太多 这样过年还可以烫一下	0.21
12	2019-10-5 19:55	快要十月了明天该写个总结了 每天早上最没食欲所以能吃的很清淡 早点把事情做完的感觉真好 台风又要来了 不会在温岭登陆了吧 老王担心不会在这两天生吧哈哈哈 anyway 明天去囤点水	0.32
13	2019-10-6 20:51	今日份晚餐	0.08
14	2019-10-7 20:54	见红了 医生说宫缩规律开一指了 都还没做好心理准备筋膜训练该了	0.35
15	2019-10-9 21:23	恭喜发财鸭	0.10
16	2019-10-9 21:24	刚吃完最后一个贝果新的一批又到啦 午餐很满足 本来打算一个人吃的老王妹妹临时说回来 连忙加了牛筋丸和荷包蛋 猪肉就喜欢这种原始的味道 晚餐吃了两个香梨感觉刚刚好不顶胃 但是现在又饿了 吃个今天刚到的红豆碱水球 嗯还会再回购的	0.18
17	2019-10-11 20:27	感觉练瑜伽有点拖延症了 学习倒是一直在进程中 碱水结感觉一般啊 还是贝果比较 yummy ps 现在觉得蔬菜最好烧了	0.21
18	2020-3-21 23:04	hello 小汤圆	0.09

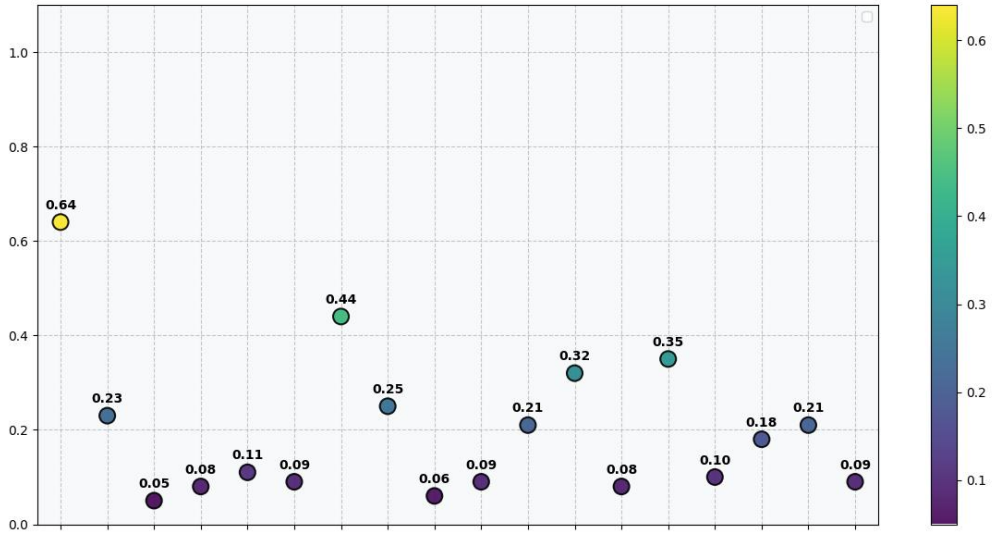

图 3. 某被标记为非抑郁的用户的可视化结果

5 模型效果评估

为了评价所提出模型的效果，我们将本模型与传统方法进行比较。并设置了准确率、召回率、精度和 F1 得分这四项评价指标。

5.1 对比算法

为了验证所提出的深度学习方法的性能，我们把模型与另外 3 个的基本方法进行比较，对它们的描述如下：

1.朴素贝叶斯(Naive Bayes, NB)。在机器学习中，朴素贝叶斯是一种基于贝叶斯定理的简单概率分类器算法。它的思想基础是对给定的项目进行分类，解决条件下每个类别的概率，在给定的项目中考虑最大值，其基本定义如下:

$$\hat{y} = \arg_k \max p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k) \qquad (10)$$

朴素贝叶斯有 3 种类型，分别是高斯朴素贝叶斯、多项式朴素贝叶斯和伯努利朴素贝叶斯。由于伯努利朴素贝叶斯适合于短文本的分类，因此选择此事件模型进行比较。

2.随机决策森林(Random Decision Forest, RDF)。在机器学习中，随机决策森林是包含多个决策树的分类器，其输出类别由各个树状结构的输出类别的模式编号(分类)或均值预测(回归)确定。这种算法被广泛用于分类、回归和其他任务。

3.BERT (Bidirectional Encoder Representations from Transformers)。它是一种基于 Transformer 架构的预训练语言模型。自注意力机制（Self-Attention）是其核心优势，它允许模型关注序列中所有位置，大幅提升效率和建模能力，对其原理介绍如下。

对于输入序列 $X = \{x_1, x_2, \ldots, x_n\}$，首先通过线性变换，分别对其中的每个元素计算：

$$Q = XW_Q, K = XW_K, V = XW_V \qquad (11)$$

其中，W_Q, W_K 与 W_V 是学习得到的网络参数。

接下来，计算元素 x_i 相对于 x_j 的注意力得分：

$$S_{ij} = \frac{Q_i K_j}{\sqrt{d_k}} \quad (12)$$

然后得出元素 x_i 相对于 x_j 的权重：

$$A_{ij} = softmax(S_{ij}) = \frac{exp(S_{ij})}{\sum_{k=1}^{n} exp(S_{ik})} \quad (13)$$

再输出元素 x_i 的注意力得分：

$$o_i = \sum_{j=1}^{n} A_{ij} V_j \quad (14)$$

最终，输出序列各元素的注意力得分向量：

$$O_i = \{o_1, o_2, \ldots, o_n\} \quad (15)$$

这种机制使 BERT 能高效捕捉全局上下文。例如，在句子“她在图书馆借了一本书后，去咖啡馆喝了一杯咖啡”中，BERT 通过计算“她”与“图书馆”“咖啡馆”等的注意力权重，准确理解“她”的行为序列。而其他模型可能难以关联“图书馆”和“咖啡馆”。

5.2 评价指标

在二元分类领域，数据的统计主要采取 4 个指标:

TP (True Rositive)，表示实际值是抑郁的，而预测值也是抑郁的;

TN (Rrue Negative)，表示实际值是非抑郁的，而预测值也是非抑郁的;

FP (False Positives)，表示实际值为非抑郁，而且预测值为抑郁的;

FN (False Negative)，表示实际值为抑郁的，而预测值为非抑郁的。

借助于以上 4 个统计数据，通过比较准确率、召回率、精度和 F1 得分，评价比较方法和比较特征的识别性能，对它们的具体描述如下：

1.准确性: 这是正确预测的值与总值的比率.几乎可以肯定，准确性是最直观的性能指标，有：

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (16)$$

2.召回率: 也称为灵敏度，是正确预测的正值与实际值中所有值的比率，有：

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (17)$$

3.精度: 也称为正预测值，是正确预测的正值与总预测的正值之比，有：

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (18)$$

4.F1 分数: 这是精度和召回率的加权平均值。通常它表示精度和召回率的和谐平均值，有：

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \times 100\% \quad (19)$$

5.3 对比实验结果

为了使模型发挥出最佳性能，我们将本模型与上述的不同对比算法在相同的数据集上进行比较，并且探究不同分类器、不同超参数设置、不同特征维度和不同学习率下的性能。

以下所有实验均在 Windows 11(64 位)操作系统上运行，使用 NVIDIA GeForce RTX 3060 显卡。软件配置为 Python 3.13, PyTorch 2.7.1, scikit-learn 1.6.0, numpy 2.2.1 and pandas 2.2.3.

以下实验的默认设置为：R-SVM 分类器，超参数 $p = 0.20, \omega = 0.50$，特征维度为 128，学习率为 0.06。每次比较时，我们只改变以上的一种条件，控制其他条件不变。

5.3.1 不同模型的比较

首先，本研究进行了一系列实验，以验证提出的多示例学习模型的性能。为了训练这些方法，在开始此实验之前，已将提取的特征格式化为相同的维度。这些方法需要使用的特征是社交网络特征、情感特征、主题特征和领域特征，每个特征都可以视为模态。由于方法自身的特殊性，仅采用时间序列特征作为输入数据来预测抑郁症。表IV显示了两个基本方法与本方法之间的性能比较。可以看出，本方法的准确率和精度最高。

表IV. 不同方法的性能

Models	Accuracy/%	Precision/%	Recall/%	F1_Score/%
NB	81.9	80.1	83.2	81.6
RDF	89.0	73.7	70.0	71.8
BERT	86.5	85.0	**88.4**	**86.7**
Our Model	**95.0**	**94.4**	79.5	86.3

为何模型存在以上性能差异？我们认为，传统机器学习方法（如朴素贝叶斯和随机森林）表现欠佳，主要源于其依赖表面词汇特征且难以建模语境语义——尤其在短小且含噪的社交媒体文本中。这些方法往往难以捕捉隐性情感表达和动态用户行为。相比之下，基于深度学习的模型通过从数据中学习语义表征展现出更强性能。本模型在此基础上进一步优化：采用多示例学习将推文级预测聚合为用户级推断，并融入时间动态特征以反映抑郁症状的波动。这种设计使模型能更精准识别抑郁用户，特别是那些表现出间歇性或间接性痛苦表达的群体。实验结果表明，BERT 在召回率和 F1 分数上表现更优，验证了其从用户生成内容中提取丰富语义特征的优势。但本模型在准确性和精确度方面超越 BERT（且 F1 分数未相差很多），这得益于其整合用户层面情境并建模时序模式的能力：尽管 BERT 能够捕捉单条推文的语言上下文并赋予其不同的注意力，但我们的模型可通过综合用户推文历史记录来检测随时间变化的行为趋势。

5.3.2 不同分类器之间的比较

表V所示为 3 个分类器之间的性能比较。可以看出，所有分类器均具有良好的性能，其中带有 RBF 内核的 SVM 的分类器在所有 4 个指标上均优于其他分类器。这些实验证明 LSTM 功能可用于检测抑郁症的推文并表现良好。由于 RBF 内核 SVM 具有最佳性能，且时间序列具有线性不可分的特征，这也意味着当尺寸不是太大并且样本为中等大小时，RBF

核方法比线性核方法更好。

表V. 不同分类器下的性能

Classifiers	Accuracy/%	Precision/%	Recall/%	F1_Score/%
LR	94.6	94.1	79.1	86.1
L-SVM	94.8	94.2	79.3	86.1
R-SVM	**95.0**	**94.4**	**79.5**	**86.3**

5.3.3 不同超参数的比较

模型中的超参数 p 与 ω 是外生给定的。为了找出合适的参数设置组合，参考 Jahnavi 等人(2023)，我们使用网络搜索方法。它的原理是在预定义的超参数空间中，系统地遍历各种不同的超参数组合，记录每个组合的得分，以找到最佳设置。

为了明确超参数 p 的范围，我们先遍历多示例学习输出结果中的各 $\hat{x}_{ij}$ 的值，其最小值接近 0，最大值约为 0.332，所以理论上我们只需在这个范围内探究不同 p 下的得分。我们不妨把在区间设为 $[0, 0.35]$ 。而超参数 ω 的范围依赖于 p ，不易直接确定，所以为了严谨起见，我们全面地探究它的不同值下得分。

根据本模型的设计，我们猜测：如果我们减少 p 与 ω 的值，就意味着同等情况下用户被识别为抑郁的概率加大，这有益于提高召回率，但是精度可能降低；反之也成立。也就是说，召回率与 p 与 ω 的值是负相关的；精度与 p 与 ω 的值是正相关的。（以下实验验证了该猜想。）

图 4、图 5、图 6 和图 7 分别显示了不同超参数 p 与 ω 下的四种得分（由于 Python 程序

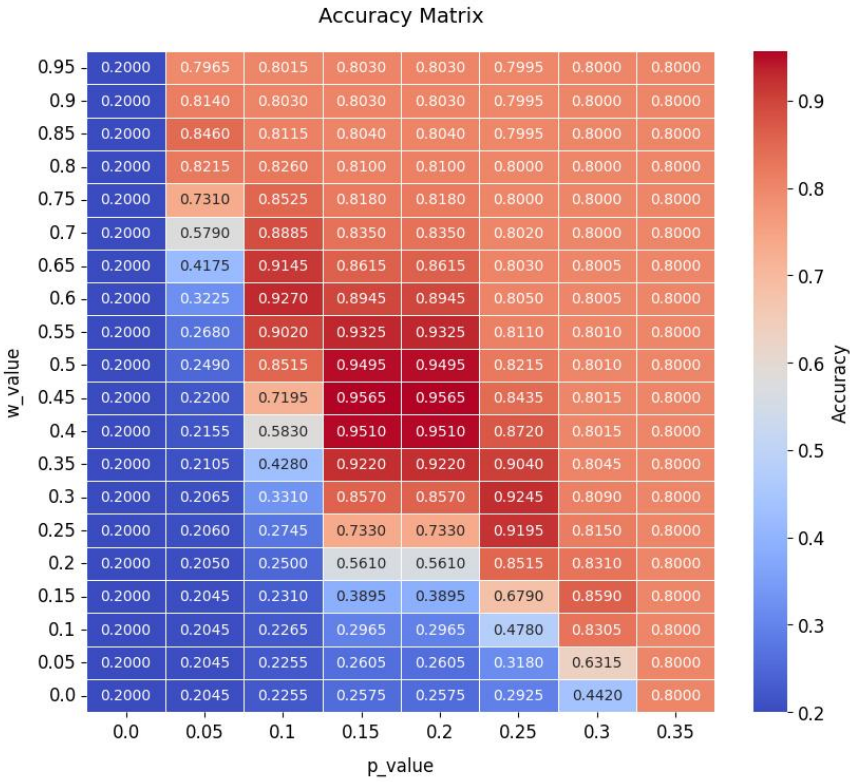

图 4 不同超参数组合下的准确性得分

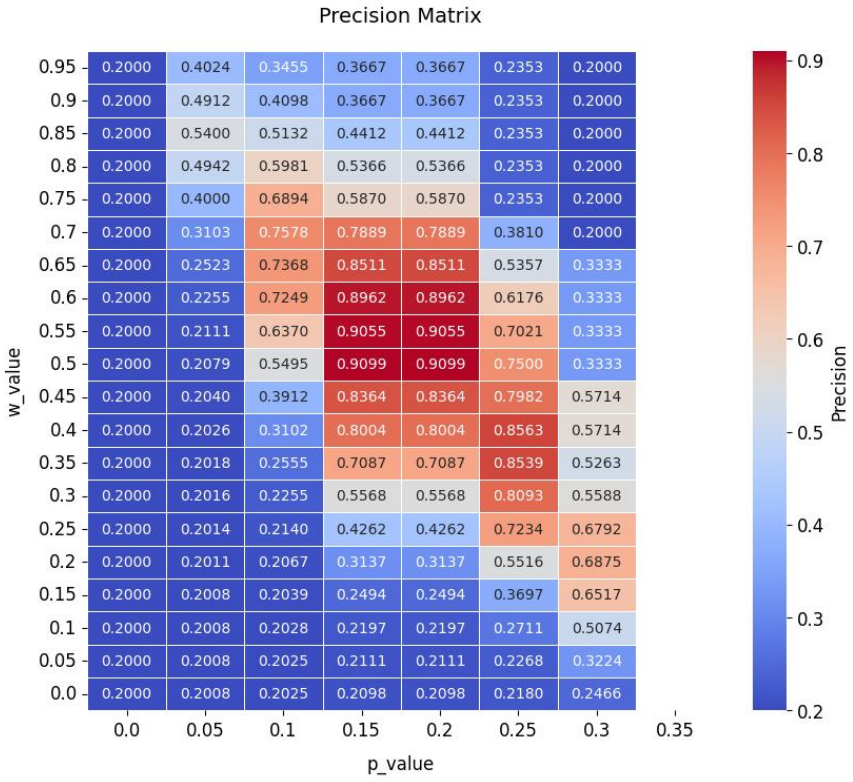

图 5 不同超参数组合下的精度得分

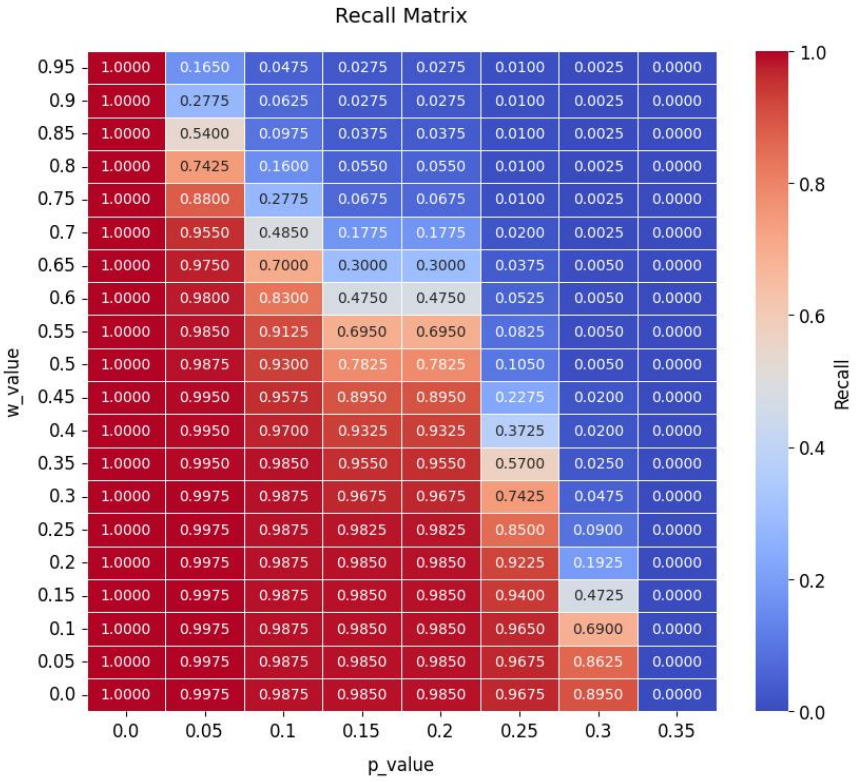

图 6 不同超参数组合下的召回率得分

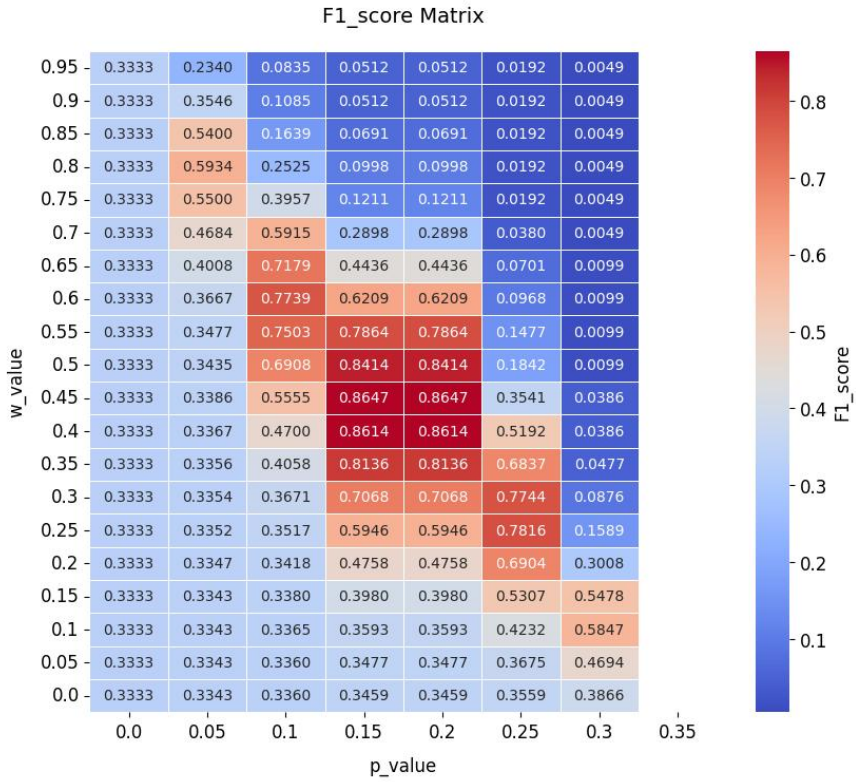

图 7 不同超参数组合下的 F1 得分

无法直接识别希腊字母，图中使用 w 代替 ω）。图 5 和图 7 中存在的空白格代表因为 p 值过大，所有的用户都被预测为非抑郁，精度和 F1 分数不存在，对应的召回率为 0；因数据集中 80%用户标签为正常，所以准确性为 0.8。

5.3.4 不同学习率的比较

作为机器学习模型训练中的一个重要参数，学习率决定了模型在训练过程中每一次参数更新的幅度。适当的学习率能够加快收敛速度，提高模型的性能；相反，过大或过小的学习率则可能导致训练效果不佳。之前的实验将学习率设置为 0.06。为了实现模型的最佳性能，我们再进行一系列不同学习率的下的实验。表VI给出了本模型在不同学习率下的性能。

表VI. 不同学习率下的得分

Learning Rate	Accuracy/%	Precision/%	Recall/%	F1_Score/%
0.04	87.9	88.2	87.5	87.9
0.05	90.3	90.0	89.5	90.3
0.06	**95.0**	**94.4**	79.5	86.3
0.07	92.0	91.5	89.7	**90.6**
0.08	91.5	91.0	**90.0**	90.5
0.09	88.5	88.0	87.0	87.8

5.3.5 不同特征维度的比较

在之前的实验中，采用了 128 维 LSTM 进行实验。为了实现模型的最佳性能，我们再进行一系列不同维度下的实验。若特征维度过小，意味着能够提供的信息量有限，可能导致模型无法捕捉到数据中的重要特征、或者无法充分拟合训练数据，出现欠拟合现象；若特征维度过大，模型可能因为在训练过程中难以处理大量的特征，导致效果不佳、也可能会过度拟合训练数据，导致在新的的数据上表现不佳。

表VII为模型分别在特征维度下的得分。若提取 256 维特征，我们发现损失值无法降低。可以看到，在尺寸 128 可获得最佳性能，并且随着尺寸的增加，性能会变得更好。

表VII. 不同特征维度下的得分

Feature Dimensions	Accuracy/%	Precision/%	Recall/%	F1_Score/%
16	82.5	79.8	84.2	81.9

32	87.3	84.5	**89.1**	86.7
64	91.8	93.5	88.2	**90.8**
128	**95.0**	**94.4**	79.5	86.3

5.3.6 不同训练周期的比较

在之前的实验中，训练周期初始值被设定为 100。当训练周期不足时，模型可能无法充分学习数据中的周期性特征和趋势模式，特别是在长期依赖关系对准确预测至关重要的场景下。反之，过多的训练周期会增加过拟合风险——尤其当训练数据有限时，模型容易学习到噪声或样本特定伪影，从而降低泛化能力。为确定最佳训练周期，我们通过对比不同周期设置下的模型性能进行评估，相关结果汇总于表VIII。

Table VIII. 不同训练周期下的得分

Training Epochs	Accuracy/%	Precision/%	Recall/%	F1_Score/%
70	82.3	81.5	73.8	77.4
80	89.1	88.2	79.4	83.5
90	93.5	92.8	**84.6**	**88.3**
100	**95.0**	**94.4**	79.5	86.3
110	94.5	94.1	76.5	84.4

6 结论

抑郁症作为一种长期的精神障碍，通常需要较长时间的观察才能确诊，其临床上的诊断需要至少两周的观察期(Shen *et al.*, 2017)。依据 Eichstaedt 等人(2018)的研究，三个月的社交媒体活动窗口足以有效识别用户是否正在经历抑郁症。一些研究开发了用于可视化抑郁概率的模块。例如，Cai 等人(2023)提出了一种基于多变量时间序列的方法来跟踪和可视化用户抑郁症状的时间动态，为未来的个人层面的心理学研究提供数据和方法论支持。

在传统的临床诊断中，心理医生需要通过与来访者进行面对面的访谈来诊断其是否患有抑郁症以及病情的程度。然而，这种方法会有很多潜在的问题。鉴于有抑郁倾向的患者更愿意向社交媒体倾述自己的心情及状态，所以借助社交媒体中的文本信息识别用户抑郁倾向可能会更有利于发现抑郁症患者。

本研究提出了一种可视化展示用户抑郁倾向变化的方法，并在一个通用标注数据集上进

行了测试，以验证所提出方法的有效性。通过可视化每条推文被预测为抑郁的概率,直观地反映用户的情绪变化轨迹。这对于及时发现和诊断抑郁症患者非常有帮助。与之前仅关注单一时间点预测的研究不同,本文的可视化分析能更好地洞察用户情绪的动态变化,为临床诊断提供更丰富的信息；与基于图像分析或生理信号的抑郁症检测方法相比,本文利用文本数据的优势在于数据获取更加便捷,可以更广泛地应用于大规模人群筛查。

与现有研究相比,本文在提出可视化分析方法方面有创新性突破。这些研究成果为利用社交媒体数据进行心理健康监测提供了新的思路和方法,对于改善抑郁症的早期发现和精准诊断具有重要的实践价值。与以往仅关注单一时间点预测的研究不同，本文的可视化分析可以更好地洞察用户情绪的动态变化，为临床诊断提供更丰富的信息。同时，与基于图像分析或生理信号的抑郁检测方法相比，本研究也具有数据采集更方便，更容易被广泛应用于大规模筛查的优点。

参考文献

American Psychiatric Association. (2013), *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition., American Psychiatric Association, doi: 10.1176/appi.books.9780890425596.

Beck, A.T. (1961), “An Inventory for Measuring Depression”, *Archives of General Psychiatry*, Vol. 4 No. 6, p. 561, doi: 10.1001/archpsyc.1961.01710120031004.

Biradar, A. and Totad, S.G. (2019), “Detecting Depression in Social Media Posts Using Machine Learning”, *2nd International Conference on Recent Trends in Image Processing and Pattern Recognition, RTIP2R 2018, December 21, 2018 - December 22, 2018*, Vol. 1037, Springer Verlag, Solapur, India, pp. 716–725, doi: 10.1007/978-981-13-9187-3_64.

Cai, Y., Wang, H., Ye, H., Jin, Y. and Gao, W. (2023), “Depression detection on online social network with multivariate time series feature of user depressive symptoms”, *Expert Systems with Applications*, Vol. 217, p. 119538, doi: 10.1016/j.eswa.2023.119538.

Chen, Z., Wang, D., Lou, L., Zhang, S., Zhao, X., Jiang, S., Yu, J., *et al.* (2025), “Text-guided multimodal depression detection via cross-modal feature reconstruction and decomposition”, *Information Fusion*, Elsevier, Amsterdam, Vol. 117, p. 102861, doi: 10.1016/j.inffus.2024.102861.

Crawford, J.R. and Henry, J.D. (2003), “The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample”, *The British Journal of Clinical Psychology*, Vol. 42 No. Pt 2, pp. 111–131, doi: 10.1348/014466503321903544.

De Choudhury, M., Gamon, M., Counts, S. and Horvitz, E. (2013), “Predicting depression via social media”, *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, Cambridge, MA, United states, pp. 128–137, doi: 10.1609/icwsm.v7i1.14432.

Dejong, T.M., Overholser, J.C. and Stockmeier, C.A. (2010), “Apples to oranges?: A direct comparison between suicide attempters and suicide completers”, *Journal of Affective Disorders*, Elsevier, Amsterdam, Vol. 124 No. 1–2, pp. 90–97, doi: 10.1016/j.jad.2009.10.020.

Deshpande, M. and Rao, V. (2017), “Depression detection using emotion artificial intelligence”, *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, presented at

the 2017 International Conference on Intelligent Sustainable Systems (ICISS), IEEE, Palladam, pp. 858–862, doi: 10.1109/ISS1.2017.8389299.

Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D.A., *et al.* (2018), "Facebook language predicts depression in medical records", *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, Vol. 115 No. 44, pp. 11203–11208, doi: 10.1073/pnas.1802331115.

Ghosh, T., Banna, Md.H.A., Nahian, Md.J.A., Uddin, M.N., Kaiser, M.S. and Mahmud, M. (2023), "An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla", *Expert Systems with Applications*, Vol. 213, doi: 10.1016/j.eswa.2022.119007.

Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H. and Eichstaedt, J.C. (2017), "Detecting depression and mental illness on social media: an integrative review", *Current Opinion in Behavioral Sciences*, Vol. 18, pp. 43–49, doi: 10.1016/j.cobeha.2017.07.005.

Guo, Z., Ding, N., Zhai, M., Zhang, Z. and Li, Z. (2023), "Leveraging Domain Knowledge to Improve Depression Detection on Chinese Social Media", *IEEE Transactions on Computational Social Systems*, Vol. 10 No. 4, pp. 1528–1536, doi: 10.1109/TCSS.2023.3267183.

Hamilton, M. (1960), "A RATING SCALE FOR DEPRESSION", *Journal of Neurology, Neurosurgery & Psychiatry*, Vol. 23 No. 1, pp. 56–62, doi: 10.1136/jnnp.23.1.56.

Harris, K.M. and Goh, M.T.-T. (2017), "Is suicide assessment harmful to participants? Findings from a randomized controlled trial", *International Journal of Mental Health Nursing*, Wiley, Hoboken, Vol. 26 No. 2, pp. 181–190, doi: 10.1111/inm.12223.

Harris, K.M., Syu, J.-J., Lello, O.D., Chew, Y.L.E., Willcox, C.H. and Ho, R.H.M. (2015), "The ABC's of Suicide Risk Assessment: Applying a Tripartite Approach to Individual Evaluations", *PLoS ONE*, Public Library Science, San Francisco, Vol. 10 No. 6, p. e0127442, doi: 10.1371/journal.pone.0127442.

Islam, M.R., Kamal, A.R.M., Sultana, N., Islam, R., Moni, M.A. and Ulhaq, A. (2018), "Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique", *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering, IC4ME2 2018*, Rajshahi, Bangladesh, doi: 10.1109/IC4ME2.2018.8465641.

Jahnavi, Y., Elango, P., Raja, S.P., Parra Fuente, J. and Verdú, E. (2023), "A new algorithm for time series prediction using machine learning models", *Evolutionary Intelligence*, Vol. 16 No. 5, pp. 1449–1460, doi: 10.1007/s12065-022-00710-5.

Kroenke, K., Spitzer, R.L. and Williams, J.B.W. (2001), "The PHQ-9: Validity of a brief depression severity measure", *Journal of General Internal Medicine*, Vol. 16 No. 9, pp. 606–613, doi: 10.1046/j.1525-1497.2001.016009606.x.

Kumar, E.R. and Venkatram, N. (2024), "Predicting and analyzing suicidal risk behavior using rule-based approach in Twitter data", *Soft Computing*, Vol. 28 No. 23, pp. 13821–13829.

Kumar, M., Dredze, M., Coppersmith, G. and De Choudhury, M. (2015), "Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides", *Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15*, presented at the the 26th ACM Conference, ACM Press, Guzelyurt, Northern Cyprus, pp. 85–94, doi: 10.1145/2700171.2791026.

Liu, D., Feng, X.L., Ahmed, F., Shahid, M. and Guo, J. (2022), "Detecting and Measuring

Depression on Social Media Using a Machine Learning Approach: Systematic Review", *Jmir Mental Health*, Jmir Publications, Inc, Toronto, Vol. 9 No. 3, p. e27244, doi: 10.2196/27244.

Ma, X., Hancock, J. and Naaman, M. (2016), "Anonymity, intimacy and self-disclosure in social media", *34th Annual Conference on Human Factors in Computing Systems, CHI 2016, May 7, 2016 - May 12, 2016*, Vol. 0, Association for Computing Machinery, San Jose, CA, United states, pp. 3857–3869, doi: 10.1145/2858036.2858414.

Mann, P., Paes, A. and Matsushima, E.H. (2021), "Screening for Depressed Individuals by Using Multimodal Social Media Data", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35 No. 18, pp. 15722–15723, doi: 10.1609/aaai.v35i18.17858.

Montgomery, S.A. and Åsberg, M. (1979), "A New Depression Scale Designed to be Sensitive to Change", *British Journal of Psychiatry*, Vol. 134 No. 4, pp. 382–389, doi: 10.1192/bjp.134.4.382.

Peng, Z., Hu, Q. and Dang, J. (2019), "Multi-kernel SVM based depression recognition using social media data", *International Journal of Machine Learning and Cybernetics*, Springer Heidelberg, Heidelberg, Vol. 10 No. 1, pp. 43–57, doi: 10.1007/s13042-017-0697-1.

Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W. and Kaminsky, Z.A. (2020), "A machine learning approach predicts future risk to suicidal ideation from social media data", *Npj Digital Medicine*, Vol. 3 No. 1.

Seabrook, E.M., Kern, M.L., Fulcher, B.D. and Rickard, N.S. (2018), "Predicting Depression From Language-Based Emotion Dynamics: Longitudinal Analysis of Facebook and Twitter Status Updates", *Journal of Medical Internet Research*, Vol. 20 No. 5, p. e168, doi: 10.2196/jmir.9267.

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S., *et al.* (2017), "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution", *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, presented at the Twenty-Sixth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, Melbourne, Australia, pp. 3838–3844, doi: 10.24963/ijcai.2017/536.

Statista. (2025), "Internet and social media users in the world 2025| Statista", available at: https://www.statista.com/statistics/617136/digital-population-worldwide/ (accessed 18 September 2025).

Tadesse, M.M., Lin, H., Xu, B. and Yang, L. (2019), "Detection of Depression-Related Posts in Reddit Social Media Forum", *IEEE Access*, Vol. 7, pp. 44883–44893, doi: 10.1109/ACCESS.2019.2909180.

Uddin, A.H., Bapery, D. and Mohammad Arif, A.S. (2019), "Depression Analysis of Bangla Social Media Data using Gated Recurrent Neural Network", *1st International Conference on Advances in Science, Engineering and Robotics Technology, ICASERT 2019, May 3, 2019 - May 5, 2019*, Institute of Electrical and Electronics Engineers Inc., Dhaka, Bangladesh, doi: 10.1109/ICASERT.2019.8934455.

Wang, Y., Wang, Z., Li, C., Zhang, Y. and Wang, H. (2020), "A Multimodal Feature Fusion-Based Method for Individual Depression Detection on Sina Weibo", *2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC)*, presented at the 2020 IEEE 39th International Performance Computing and

Communications Conference (IPCCC), pp. 1–8, doi: 10.1109/IPCCC50635.2020.9391501.

Wei, P.-C., Peng, K., Roitberg, A., Yang, K., Zhang, J. and Stiefelhagen, R. (2023), "Multi-modal Depression Estimation Based on Sub-attentional Fusion", *Workshops Held at the 17th European Conference on Computer Vision, ECCV 2022, October 23, 2022 - October 27, 2022*, Vol. 13806 LNCS, Springer Science and Business Media Deutschland GmbH, Tel Aviv, Israel, pp. 623–639, doi: 10.1007/978-3-031-25075-0_42.

World Health Organization. (2017), "Depression and Other Common Mental Disorders: Global health estimates Internet", available at: https://www.who.int/publications/i/item/depression-global-health-estimates (accessed 10 October 2025).

World Health Organization. (2019), "Suicide", available at: https://www.who.int/news-room/fact-sheets/detail/suicide (accessed 10 October 2025).

World Health Organization. (2023a), "WHO launches new resources on prevention and decriminalization of suicide", available at: https://www.who.int/news/item/12-09-2023-who-launches-new-resources-on-prevention-and-decriminalization-of-suicide (accessed 18 September 2025).

World Health Organization. (2023b), "Depressive disorder (depression)", 19 February, available at: https://www.who.int/news-room/fact-sheets/detail/depression (accessed 18 September 2025).

Yan, Z., Peng, F. and Zhang, D. (2025), "DECEN: A deep learning model enhanced by depressive emotions for depression detection from social media content", *Decision Support Systems*, Vol. 191, p. 114421, doi: 10.1016/j.dss.2025.114421.

Yoon, J., Kang, C., Kim, S. and Han, J. (2022), "D-vlog: Multimodal Vlog Dataset for Depression Detection", *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, Vol. 36, Virtual, Online, pp. 12226–12234, doi: 10.1609/aaai.v36i11.21483.

Yu, B., Zhang, X., Wang, C., Sun, M., Jin, L. and Liu, X. (2020), "Trends in depression among Adults in the United States, NHANES 2005–2016", *Journal of Affective Disorders*, Vol. 263, pp. 609–620, doi: 10.1016/j.jad.2019.11.036.

Zhang, W., Xie, J., Zhang, Z. and Liu, X. (2024), "Depression Detection Using Digital Traces on Social Media: A Knowledge-aware Deep Learning Approach", *Journal of Management Information Systems*, Vol. 41 No. 2, pp. 546–580, doi: 10.1080/07421222.2024.2340822.