

Early Depression Detection in Social Media: Monitoring of Individual Nighttime Dynamics and LLM Analysis

Bicheng Yu, Zhichang Zhang, Lulu Ma, Jiongfu Cai, Yuanyuan Zhang

Submitted to: JMIR Mental Health
on: October 23, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 27

 Figures 28

 Figure 1..... 29

 Figure 2..... 30

 Figure 3..... 31

Early Depression Detection in Social Media: Monitoring of Individual Nighttime Dynamics and LLM Analysis

Bicheng Yu¹; Zhichang Zhang¹ PhD; Lulu Ma¹; Jiongfu Cai¹; Yuanyuan Zhang¹

¹ Northwest Normal University Lanzhou, Gansu, CN CN

Corresponding Author:

Zhichang Zhang PhD

Northwest Normal University
967 anning east road, lanzhou city, gansu province, China
Lanzhou, Gansu, CN
CN

Abstract

Background: Depression has become a major global public health challenge, and early intervention is critical for improving patient outcomes. Current depression detection techniques based on social media data (Traditional Risk Detection, TRD) rely heavily on users' complete historical information, which cannot meet the timeliness requirements of early intervention. This underscores the need for Early Risk Detection (ERD) methods emphasizing early-stage and real-time warning. However, existing ERD studies have notable limitations: (1) they overlook sleep–wake rhythms hidden in posting timestamps, missing vital warning signals; and (2) they depend on static templates or resource-intensive sequence models, resulting in limited interpretability and inefficient use of early data, ultimately constraining their clinical applicability for early intervention.

Objective: To address these issues, this study aims to develop an efficient, reliable, and interpretable ERD model. The core objectives are: to extract sleep–wake rhythm features from posting timestamps, thereby to enrich the feature dimensions for risk warning; to leverage large language models (LLM) for improved text filtering precision and depression-related factor analysis; and ultimately to achieve accurate early detection of depression, supporting early clinical intervention.

Methods: We propose the Monitoring of Individual Nighttime Dynamics and LLM Analysis (MIND) model, which integrates two key innovations:

(1) Sleep Dynamics: posting timestamps are transformed into sleep–wake rhythms, analyzing fluctuations in posting frequency and time to derive sleep-related features, thereby compensating for the limitations of text-only approaches.

(2) LLM Depression Profiler: LLM is used for dynamic text filtering, automatically removing irrelevant noise and focusing on potential depression-related cues. Based on LLM semantic understanding, latent depression risk factors are identified, enhancing interpretability for clinical treatment and robustness to noise.

Results: Experiments on the eRisk2017 benchmark dataset demonstrated that MIND significantly outperformed existing baseline models in early detection sensitivity, specificity, and accuracy. By combining sleep features with text analysis, the model achieved interpretable, traceable predictions that can support clinical treatment. Relevant experimental code is publicly available.

Conclusions: The MIND model innovatively combines sleep–wake rhythm features with LLM-based text analysis, addressing the challenges of poor interpretability and inefficient use of early-stage data in existing ERD methods. It significantly enhances early detection performance, offering a new paradigm for applying social media data in ERD task, thereby enabling earlier intervention and reducing the public health burden of depression.

(JMIR Preprints 23/10/2025:86367)

DOI: <https://doi.org/10.2196/preprints.86367>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/



Original Manuscript

Early Depression Detection in Social Media: Monitoring of Individual Nighttime Dynamics and LLM Analysis

Abstract

Background: Depression has become a major global public health challenge, and early intervention is critical for improving patient outcomes. Current depression detection techniques based on social media data (Traditional Risk Detection, TRD) rely heavily on users' complete historical information, which cannot meet the timeliness requirements of early intervention. This underscores the need for Early Risk Detection (ERD) methods emphasizing early-stage and real-time warning. However, existing ERD studies have notable limitations: (1) they overlook sleep–wake rhythms hidden in posting timestamps, missing vital warning signals; and (2) they depend on static templates or resource-intensive sequence models, resulting in limited interpretability and inefficient use of early data, ultimately constraining their clinical applicability for early intervention.

Objective: To address these issues, this study aims to develop an efficient, reliable, and interpretable ERD model. The core objectives are: to extract sleep–wake rhythm features from posting timestamps, thereby to enrich the feature dimensions for risk warning; to leverage large language models (LLM) for improved text filtering precision and depression-related factor analysis; and ultimately to achieve accurate early detection of depression, supporting early clinical intervention.

Methods: We propose the **Monitoring of Individual Nighttime Dynamics and LLM Analysis (MIND)** model, which integrates two key innovations:

(1) **Sleep Dynamics:** posting timestamps are transformed into sleep–wake rhythms, analyzing fluctuations in posting frequency and time to derive sleep-related features, thereby compensating for the limitations of text-only approaches.

(2) **LLM Depression Profiler:** LLM is used for dynamic text filtering, automatically removing irrelevant noise and focusing on potential depression-related cues. Based on LLM semantic understanding, latent depression risk factors are identified, enhancing interpretability for clinical treatment and robustness to noise.

Results: Experiments on the eRisk2017 benchmark dataset demonstrated that MIND significantly outperformed existing baseline models in early detection sensitivity, specificity, and accuracy. By combining sleep features with text analysis, the model achieved interpretable, traceable predictions that can support clinical treatment. Relevant experimental code is publicly available[1].

Conclusions: The MIND model innovatively combines sleep–wake rhythm features with LLM-based text analysis, addressing the challenges of poor interpretability and inefficient use of early-stage data in existing ERD methods. It significantly enhances early detection performance, offering a new paradigm for applying social media data in ERD task, thereby enabling earlier intervention and reducing the public health burden of depression.

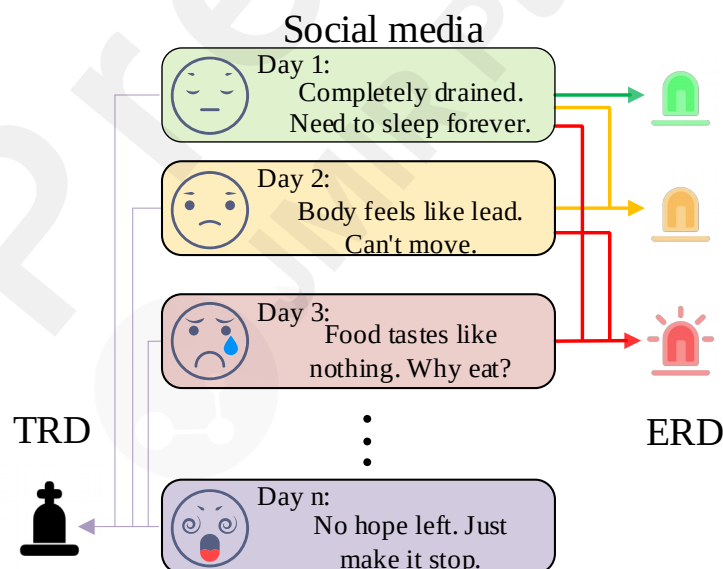
Keywords: Depression, Early Risk Detection, Sleep Analysis, Large Language Model

Introduction

Background

Depression affects approximately 260 million people worldwide [2][3]. This mental disorder leads to prolonged negative emotions, feelings of loss, and, in severe cases, life-threatening risks, making it a major public health concern. As social media increasingly becomes a platform for individuals to record their daily experiences and express emotions, user-generated content often reflects genuine psychological states[4]. This has enabled the academic community to explore social media data for depression detection. Over the years, the eRisk workshop has explored various contexts for identifying psychological disorders and predatory conversations [5][6][7], contributing valuable insights to the broader field of mental health monitoring through digital traces. Traditional Risk Detection (TRD) identifies depression from users' sufficient and long-term social media history and has achieved significant success. However, TRD fails to align with the clinical principle of "early detection and early treatment" [8], as it relies on sufficient and long-term user data. To bridge this gap, Early Risk Detection (ERD) has been proposed. ERD aims to predict depression risks based on limited, early-stage user data, allowing timely warnings and interventions. Unlike TRD, ERD prioritizes high precision under constrained data availability, with evaluation criteria focusing not only on final accuracy but also on the earliness of prediction, as illustrated in Figure 1.

Figure 1. The left side illustrates the TRD task, which predicts a user's depression status using all available user information. In contrast, the right side represents the ERD task, which aims to predict depression status by utilizing early-stage information.



Previous Work

ERD research generally follows two strategies:

- **Long-sequence processing methods:** Recurrent neural networks (RNNs)[9] and their variants, along with CNNs[10] and variants of the attention mechanism[11], have been used to process

full user histories[12][13][14]. While timestamp data has occasionally been considered, analyses have been limited to simple statistical summaries, failing to capture rhythmic features. Although comprehensive, such methods are resource-intensive and often ineffective at detecting sparse but meaningful emotional signals. Moreover, they cannot identify specific time points of psychological state changes, limiting their clinical value.

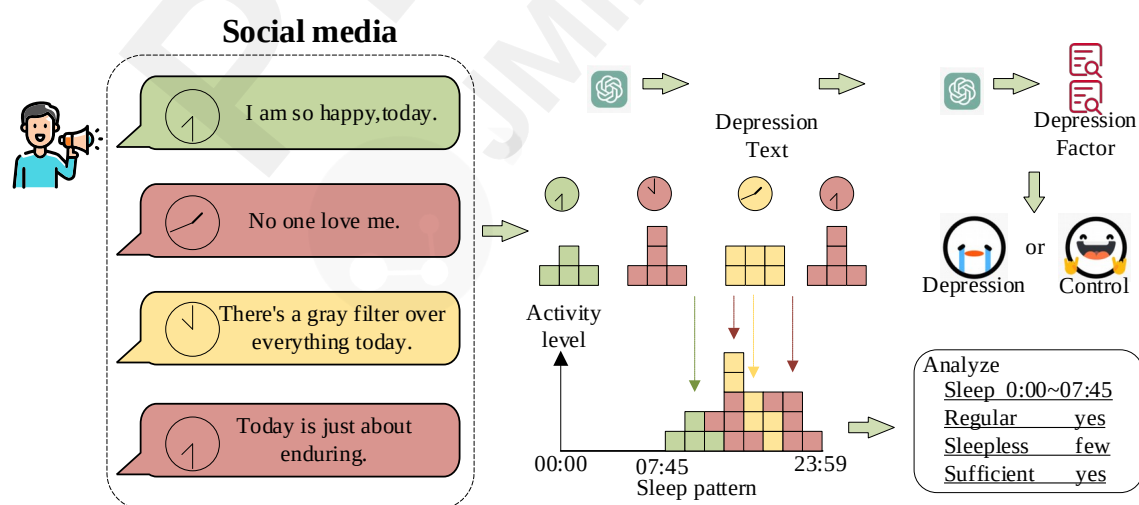
- **Filtering and denoising methods:** Template matching and sliding windows have been employed to filter high-information texts[15][16], reducing noise complexity. However, these static methods lack adaptability to individual differences and developmental psychology. While effective in identifying overt depressive content, they struggle with latent signals due to limitations of earlier NLP methods, offering insufficient support for intervention.

Studies in TRD also continuously provide new inspiration for the advancement of ERD research. Early TRD approaches relied on text feature analysis using traditional machine learning models (e.g., SVMs, decision trees) and LIWC-based features [17], but suffered from noise sensitivity and feature unimodality. Multimodal fusion of text, audio, and images [18][19] improved robustness, while more recent studies employed LLM for data augmentation and graph structures for multimodal fusion, advancing clinical depression dialogue prediction [20]. Nonetheless, TRD methods remain constrained by their reliance on complete histories, making them unsuitable for early intervention.

In summary, current ERD methods face three limitations: (1) neglect of information in timestamps, particularly sleep-related patterns critical to depression; (2) static, template-driven filtering with weak interpretability; and (3) resource-heavy, long-sequence methods that fail to capture sparse but crucial early cues.

Study Objectives

Figure 2. Illustration of innovation points. This figure outlines how user data are transformed into sleep pattern features and user depression profile features.



To overcome these challenges, we propose MIND model, comprising three key modules:

- **Sleep Dynamics (SD):** By modeling users' sleep-wake rhythms from social media posting timestamps and integrating these temporal features with textual information, Sleep Dynamics

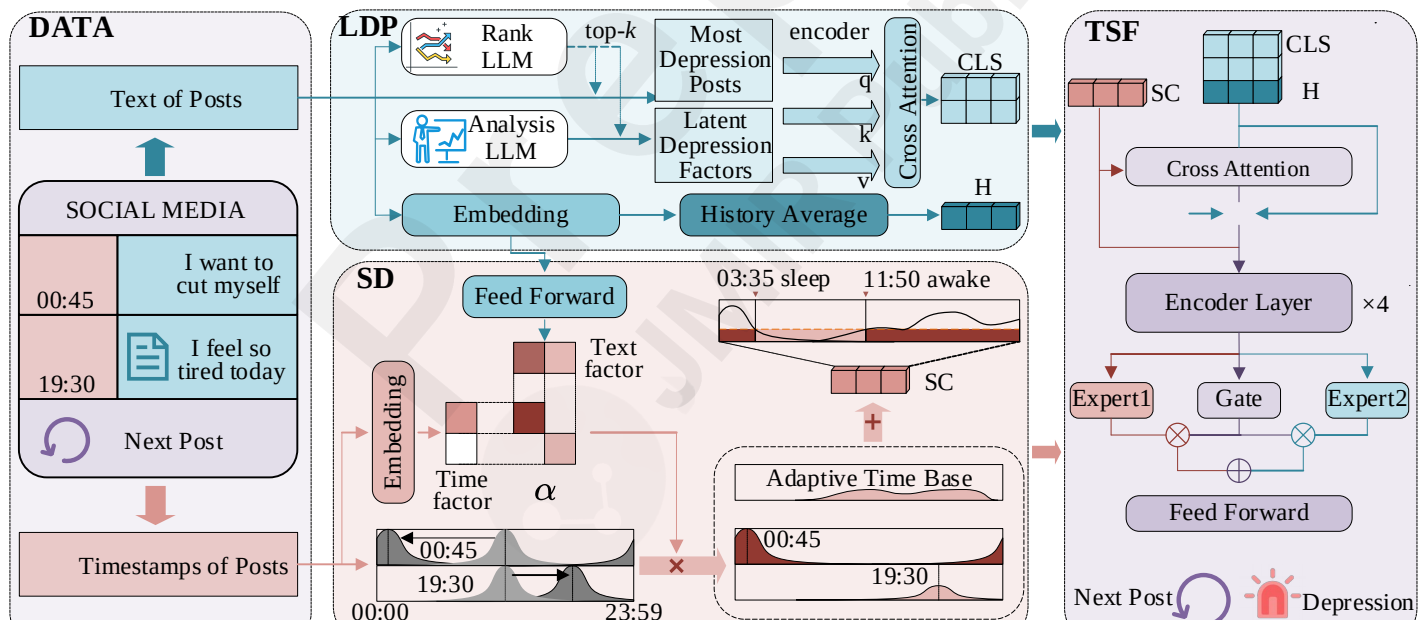
investigates depressive tendencies through the interaction between activity frequency and sleep regularity. This module is grounded in psychological evidence demonstrating a strong association between disrupted sleep patterns and depression. [21][22][23][24][25].

- **LLM Depression Profiler (LDP):** The integration of LLM enables the automatic filtering of texts from users at high risk of depression and mitigates interference. Building on this, the LLM analyzes latent negative psychological factors in the text, which markedly boosts the method's information utilization efficiency and interpretability.
- **Text-Sleep Fusion (TSF):** Fuses multimodal features (sleep and text) for final early risk predictions.

Experiments on the eRisk2017 dataset validate the effectiveness of MIND. Our contributions include: 1. Introducing timestamp-based sleep-wake rhythms inference as an ERD feature. 2. Leveraging LLM for text filtering and depression factor analysis, improving efficiency and interpretability. 3. Designing a novel multimodal fusion model (MIND) that significantly enhances early prediction performance. Figure 2 illustrates how user data is converted into the features we need.

Methods

Figure 3. Overview of the MIND Model.



Task Definition

Let a user's collection of posts on a social media platform be denoted as $P_i = \{p_{i1}, p_{i2}, \dots, p_{iL}\}$, with each post associated with a timestamp forming a corresponding set of posting timestamps $T_i = \{t_{i1}, t_{i2}, \dots, t_{iL}\}$. Each user is labeled according to their psychological state y_i (depressed or control).

In traditional risk detection tasks (TRD), models typically leverage all available user history to predict psychological states. By contrast, the Early Risk Detection (ERD) task is defined as a sequential prediction process, where posts are ordered chronologically, and each post is associated with a time step j (with $j = 1, 2, \dots, L$).

At each time step j (where $j < L$), the model processes available posts $P_i^j = \{p_{i1}, p_{i2}, \dots, p_{ij}\}$ and timestamps $T_i^j = \{t_{i1}, t_{i2}, \dots, t_{ij}\}$ to generate a prediction $\hat{y}_i^j$. This task aims to identify the user's depressive state as early as possible (i.e., at the earliest possible time step $j \in (1, L)$), with only one final prediction $\hat{y}_i$ allowed. The evaluation metric emphasizes whether the model can make an accurate prediction at an earlier step, rather than waiting for the complete history.

Sleep Dynamics

To better simulate user's sleep-wake rhythms, we propose a probabilistic modeling approach based on timestamp data. The core assumption is that a user is most likely asleep during their least active posting hours. Intuitively, user activity level peaks around posting times and decreases with distance from those timestamps. This naturally aligns with a Gaussian distribution.

To capture fine-grained sleep-wake rhythms, we discretize each day (24 hours) into 768 time slots. For each posting timestamp t , the corresponding slot m is used as the Gaussian center to initialize a feature vector v_t . The above process can be expressed by the following formula:

$$m = \lfloor t \times 32 \rfloor \bmod 768$$

(AUTONUM Arabic)

$$v_t(k) = \exp\left(-\frac{k^2}{2\sigma^2}\right)$$

where $v_t(k)$ represents the initial value of the k -th slot in the initialization vector ($k \in [0, 767]$). The decay rate parameter σ controls how quickly activity diminishes away from m . Smaller values yield sharper peaks, while larger values distribute activity more evenly.

Additionally, not all timestamps exert equal influence on activity levels. To address this, we introduce learnable hourly influence factors α_{time} and text-derived influence factors α_{text} , which are combined to compute the final impact score α . Summing across all posts within a time step j yields a user's sleep-wake rhythms representation SC^j . The process outlined above may be formulated as follows:

$$\alpha_n^{time} = \text{embedding}(t_n)$$

(AUTONUM Arabic)

$$\alpha_n^{text} = \text{Linear}(p_n)$$

(AUTONUM Arabic)

$$\alpha_n = \alpha_n^{time} + \alpha_n^{text}$$

(AUTONUM Arabic)

$$SC^j = \frac{4}{j} \sum_{n=0}^j \alpha_n \odot v_n$$

(AUTONUM Arabic)

where $n \in [1, j]$, $\alpha \in R$, $SC \in R^{768}$, $v_0 \in trainable R^{768}$.

At this point, for any time step j , we can calculate the sleep schedule (SC) of user i using $T_i^j = \{t_{i1}, t_{i2}, \dots, t_{ij}\}$ and $P_i^j = \{p_{i1}, p_{i2}, \dots, p_{ij}\}$, enabling the construction of a personalized sleep–wake rhythm feature. To attain improved performance, we refined traditional text information processing approaches and combined them with sleep–wake rhythm features.

LLM Depression Profiler Analysis

To overcome the rigidity and lack of interpretability in traditional text filtering, we redesign the depression-text selection process by leveraging **LLM** with *few-shot* and *zero-shot* learning paradigms[26][27][28].

Table 1. Prompt of LLM

Prompt of Rank LLM	
You are a psychologist who needs to assess whether a patient's text contains depression related emotions in order to score based on the severity of emotions. Don't explain why, just output the result.	
emotion: {sad, hopeless, failure, suicide, expected to be punished, guilty, cry, lose interest, worthless, low energy, change pattern, irritable, change appetite, tired}	
response: 0~5	
<example>	
text: Just it looks so lonely	
score: 3	
text: I still feel awful whenever I make a mistake	
score: 5	
</example>	
{input text}	
Prompt of Analysis LLM	
You are a clinical psychologist. Read the following text and identify any potential psychological issues.	
–Your reply must be very brief (one sentence or less).	
–Do not explain your reasoning.	
–If no issues are found, reply exactly “None”	
{input text}	

Specifically, rank LLM can learn to score the depression level of user posts under the guidance of a small number of expert examples, and select the top- k most relevant depression posts

$P_{i,j}^{topk} = p_{top1}, p_{top2}, \dots, p_{topk}$ based on these scores. Subsequently, we use Analysis LLM to conduct in-depth explanatory analysis on these texts and extract potential depression factors $F_{i,j}^{topk} = f_{top1}, f_{top2}, \dots, f_{topk}$.

Next, we use a pre-trained language model Encoder to embed $P_{i,j}^{topk}$ and $F_{i,j}^{topk}$ into matrices M_p and M_F , respectively. Semantic fusion is performed via Cross-Attention[29], and finally, a more refined user depression text profile CLS is obtained, which can be described by the following formula:

$$M_p = \text{Encoder} \left(P_{i,j}^{topk} \right) \\ (AUTONUM \text{ } \textcolor{red}{\rightarrow} \text{Arabic})$$

$$M_F = \text{Encoder} \left(F_{i,j}^{topk} \right) \\ (AUTONUM \text{ } \textcolor{red}{\rightarrow} \text{Arabic})$$

$$M = \text{softmax} \left(\frac{(W_q M_p)(W_k M_F)}{\sqrt{d}} \right) \times W_v M_F \\ (AUTONUM \text{ } \textcolor{red}{\rightarrow} \text{Arabic})$$

$$CLS = M_{[CLS]} \\ (AUTONUM \text{ } \textcolor{red}{\rightarrow} \text{Arabic})$$

where $M, M_p, M_F \in R^{k \times seq \times dim}$, $CLS \in R^{k \times dim}$, $W \in trainable R^{dim \times dim}$.

At this point, we have obtained the user depression-text profile feature CLS enhanced by the LLM. Additionally, to ensure that the historical information posted by the user is not overlooked by the model, we average the embeddings of all the user's texts and use this as the coarse-grained historical text feature:

$$history = \text{Average} \left(\text{Encoder} \left(P_{ij} \right) \right) \\ (AUTONUM \text{ } \textcolor{red}{\rightarrow} \text{Arabic})$$

Text-Sleep Fusion

First, for sleep-aware text feature enhancement, we leverage the Cross-Attention mechanism and use sleep-wake rhythm features as prior knowledge to guide the fusion of depressive text profile features and historical information features, thereby achieving the goal of targeted text feature enhancement.

$$F_{text} = \text{concat} (CLS, history) \\ (AUTONUM \text{ } \textcolor{red}{\rightarrow} \text{Arabic})$$

$$F_{fusion} = \text{CrossAttention} (F_{text}, SC, SC) \\ (AUTONUM \text{ } \textcolor{red}{\rightarrow} \text{Arabic})$$

$$\text{CrossAttention}(Q, K, V) = \text{softmax}\left(\frac{(W_q Q)(W_k K)}{\sqrt{d}}\right) \times W_v V$$

(AUTONUM Arabic)

where $W_{q,k,v} \in \text{trainable } R^{dim \times dim}$.

Next, to preserve the order of text and distinguish between historical texts and filtered texts, we add trainable positional encoding. The resulting features are then concatenated with sleep–wake rhythm features SC and input into the Transformer Encoder.

$$F = \text{concate}(SC, F_{\text{fusion}} + PE)$$

(AUTONUM Arabic)

$$Fusion = \text{TransformerEncoder}(F)$$

(AUTONUM Arabic)

where $PE \in \text{trainable } R^{k \times dim}$. As global prior knowledge, sleep–wake rhythm features SC are inherently different from the sequential attribute of text sequences due to their time-independence; therefore, SC is excluded from positional encoding to ensure the model can correctly distinguish between the two. After in-depth data mining by the Transformer Encoder, more accurate multimodal features are obtained.

To better address the information compression issue when mapping multimodal features to a single user representation, we adopt the Mixture of Experts (MoE) approach—a framework originally proposed by Jacobs et al.(1991) [30] that has recently re-emerged as a critical technique in LLM research. By dynamically selecting expert combinations to enable more flexible feature fusion, this modernized application of MoE avoids potential information loss caused by traditional pooling operations, making it well-suited for the demands of contemporary multimodal and large-scale model scenarios. Finally, the final prediction is generated through a linear layer.

$$e_i = W_i^e \times Fusion$$

(AUTONUM Arabic)

$$g_i = W_i^g \times Fusion$$

(AUTONUM Arabic)

$$e = \sum_i e_i \odot g_i$$

(AUTONUM Arabic)

$$\hat{y}_i^j = \text{Softmax}(\text{Linear}(e))$$

(AUTONUM Arabic)

where $e, e_i \in R^{dim}$, $g_i \in R^1$, $W^e \in \text{trainbale } R^{dim \times dim}$, $W^g \in \text{trainbale } R^{dim \times 1}$. Since this is a classification task, we use the traditional cross-entropy loss function, defined as follows:

$$Loss = \sum_i -[y_i \times \log(\hat{y}_i) + (1 - y_i) \times \log(1 - \hat{y}_i)]$$

(AUTONUM Arabic)

Results

Dataset

We conducted experiments on the **eRisk-2017** dataset, the standard benchmark for ERD tasks [20]. The dataset contains 137 depressed users and 755 control users, divided into training (486) and testing (406) sets, collected from the Reddit platform. Depressed users were identified through self-reports in depression-related forums, with explicit confirmation of clinical diagnosis. Control users were sampled from non-depression subforums and depression subforums where users exhibited no depressive symptoms.

To ensure privacy, all user IDs were anonymized, and only text content and posting times were preserved. Table 2 summarizes dataset statistics.

Table 2. Statistics of the dataset

	train		test	
	depression	control	depression	control
Users	83	403	52	349
Posts	30,851	172,834	18,706	217,665
AVG Post/User	371.7	655.5	359.7	623.7

Implementation Details

We used GPT-4o-2024-11-20 as the base LLM. For text filtering, the top-16 posts with the highest depression scores were retained. The encoder was implemented with BERT[31], and embeddings were generated using the frozen paraphrase-MiniLM-L6-v2 model. In the Sleep Dynamics module, the Gaussian parameter σ was set to 1. All attention layers adopted multi-head attention with 8 heads.

In the Text-Sleep Fusion module, the Transformer encoder was configured with 4 layers. Training used Adam optimizer with a learning rate of $2e-5$, batch size of 4, and 7–14 epochs. Training was conducted on NVIDIA A30 GPUs with PyTorch 2.3.0. To mitigate randomness, five random seeds were used, and the best-performing model was reported. Additional experiments on encoder depth are included in the appendix.

Comparison Methods

We compared MIND with multiple baselines, including official dataset baselines and subsequent studies:

- **LR:** TF-IDF features + logistic regression, reframing ERD as multi-class classification[32].
- **Feature-Rich:** Feature engineering with topic extraction (LDA) and user statistics [33].
- **BiLSTM+Attention:** Sequence modeling with full user history using LSTMs and attention [16].
- **Risk Window:** Sliding-window strategy to trigger warnings when depressive posts appear consecutively[16].
- **τ -SS3:** Enhanced n-gram approach for dynamic phrase sequence detection[34].
- **Early List:** Reinforcement learning-based gating to encourage early predictions[12].
- **HAN-Psych:** Hierarchical attention networks inspired by clinical diagnosis questionnaires[15].
- **ESDM:** Balancing earliness and accuracy with a loss function combining LSTM and linear attention[13].

For fairness, all models followed their original decision-making rules. For methods without explicit rules, we adopted the default “diagnose-as-soon-as-identified” strategy. Due to hardware constraints, ESDM and Early List were limited to processing 64 user posts. Due to hardware constraints, ESDM and Early List were limited to processing 64 user posts. Therefore, we do not show the time elapsed during the inference process of these methods.

Metrics

ERD introduces a specialized evaluation metric balancing **earliness** and **accuracy**. This metric penalizes late predictions exponentially beyond a certain time threshold, ensuring models are rewarded for early and correct predictions.

$$ERD E_o(j) = \begin{cases} c_{fp}, FP \\ 1, FN \\ l_{c_o}(k) \times c_{t_p}, TP \\ 0, TN \end{cases}$$

(AUTONUM $\rightarrow$ Arabic)

$$l_{c_o}(j) = 1 - \frac{1}{1 + e^{-j+o}}$$

(AUTONUM $\rightarrow$ Arabic)

where $l_{c_o}(j) \in [0,1]$, $c_{fp} = \frac{\text{positive}}{\text{total}}$.

In addition, we also introduced the F1-score of the final early warning results for the ERD task ($F1_{ERD}$) and the F1-score for the TRD task ($F1_{TRD}$) respectively, to more accurately evaluate the model's accuracy and its generalization ability on the TRD task.

Overall, a superior model is characterized by smaller values in metrics $ERDE_5$, $ERDE_{50}$ and larger values in metrics $F1_{ERD}$, $F1_{TRD}$.

Evaluation Outcomes

Table 3. Main results of the experiments.

Model	$\downarrow ERDE_5$	$\downarrow ERDE_{50}$	$\uparrow F1_{ERD}$	$\uparrow F1_{TRD}$	$\downarrow$ Time cost
LR	0.137	0.084	0.405	0.602	4710
Feature-Rich	0.131	0.084	0.358	0.630	7558
BiLSTM+Att	0.124	0.096	0.562	0.629	1101
Riskwindow	0.130	0.097	0.606	0.629	920
τ -SS3	0.133	0.086	0.497	0.546	-
EarlyList	0.164	0.148	0.273	0.175	-
HAN-Psych	0.107	0.081	0.631	0.703	1330
ESDM	0.109	0.077	0.662	0.712	-
MIND	0.080	0.069	0.671	0.727	2208

Table 3 shows that MIND achieves substantial improvements in $ERDE_5$ and $ERDE_{50}$ metrics (25.2% and 10.3% relative gains over existing models), attributed to the introduction of sleep–wake rhythm features. Gains in F1 scores were moderate but consistent. The tradeoff was increased inference time due to LLM-based components.

Ablation Study

To evaluate each component's contribution, we performed ablation experiments:

- **–Sleep:** Removed the Sleep Dynamics, relying solely on text.
- **–Text:** Removed the LLM Depression Profiler, using only sleep–wake rhythm features.
- **–LLM:** Replaced LLM-based filtering with the previous method, template matching.
- **–History:** Removed historical text features.
- **–Fusion:** Removed Cross-Attention and MoE, using only a Transformer encoder.

Table 4. Results of the ablation study. The MIND results are bolded.

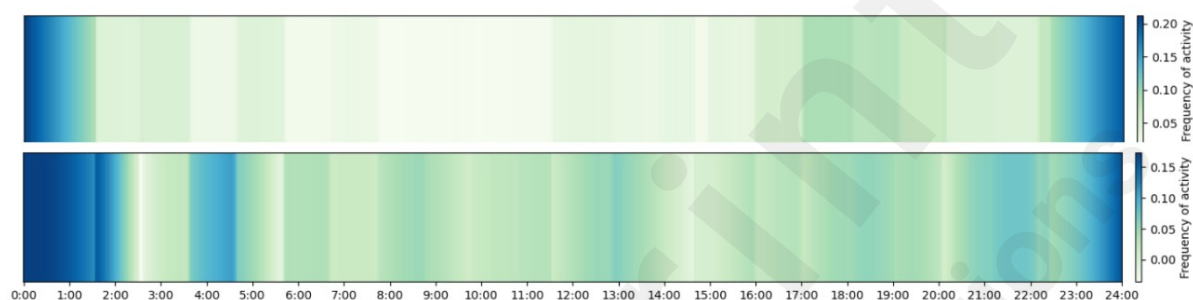
Model	$\downarrow ERDE_5$	$\downarrow ERDE_{50}$	$\uparrow F1_{ERD}$
MIND	0.080	0.069	0.671
-Sleep	0.094	0.078	0.633
-Text	0.153	0.122	0.430
-LLM	0.103	0.093	0.655
-History	0.106	0.078	0.668
-Fusion	0.084	0.071	0.669

Table 4 presents the results of our ablation experiments. It can be observed that removing the Sleep Dynamics has a significant impact on the model; furthermore, the Sleep Dynamics alone struggles to

achieve favorable performance. Replacing the LLM component in the model with existing methods exerts a substantial influence on the experimental results. Removing user historical features has a more pronounced effect on $ERDE_5$, which is more sensitive to earliness, but a relatively minor impact on $ERDE_{50}$, which has a higher tolerance for earliness. Removing the cross-attention and expert gating mechanisms only has a slight impact on the overall performance of the model.

Case Study

Figure 4. Visualization of Sleep–Wake Rhythms in Depressed vs. Control Users



We visualize sleep–wake rhythms heatmaps for depressed and control users. Control users typically show near-zero activity between 6:00–11:00, indicating sufficient and regular sleep. In contrast, depressed users display irregular activity across all hours, often engaging in late-night posting, with loosely defined sleep windows (e.g., 5:00–12:00).

Table 5 shows examples of LLM scoring and analysis:

These examples highlight MIND’s interpretability: high-risk posts (e.g., self-harm) are prioritized, while latent depressive tendencies are also captured, providing actionable insights for clinical support.

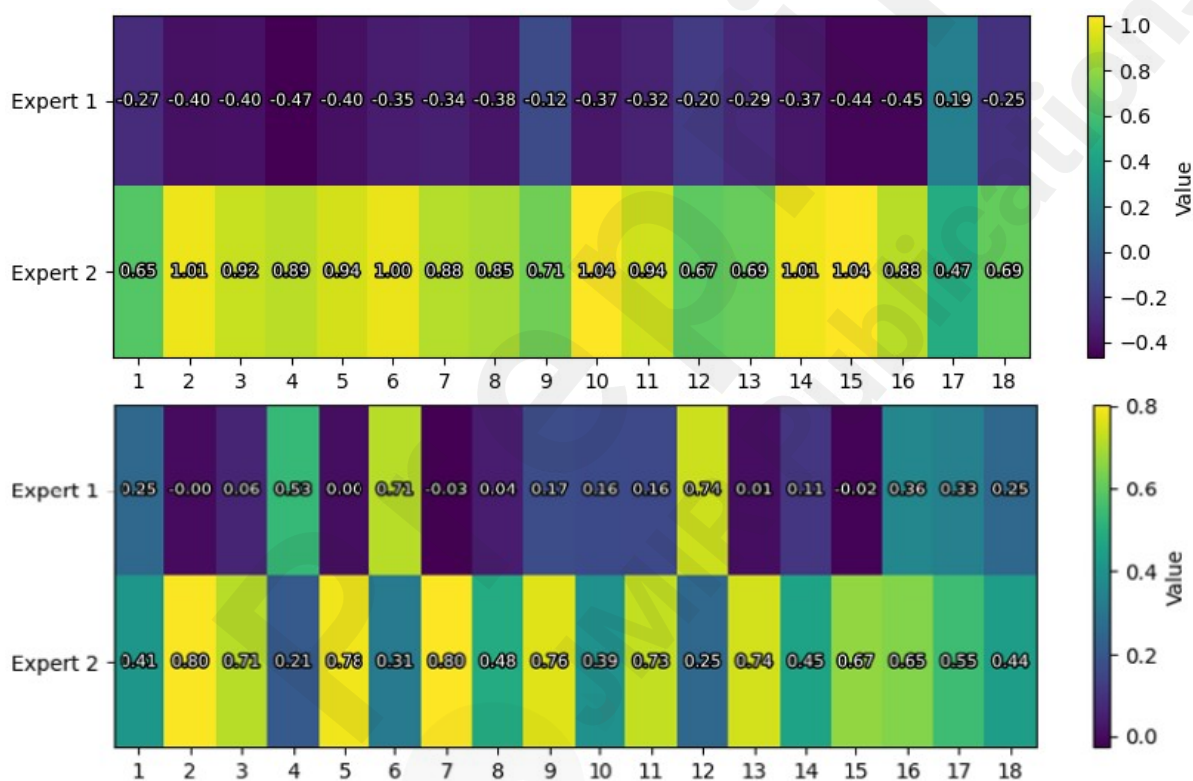
Table 5. Examples of Rank LLM Scores and Analysis LLM Depression Factor Analysis

Score	User post	Depression Factor
5	Today is a bad day, maybe I should cut my wrist	Potential self-harm ideation
2	My job is unstable. It’s fine now, but what about the future.	Potential issues with stress or anxiety related to work hours.
1	I tried to express this meaning with out directly accusing OP of abusive behavior.	Potential issues may include fear of confrontation or difficulty in addressing conflict directly.
0	I always thought this was why boys urinate standing up.	None

Contribution Analysis

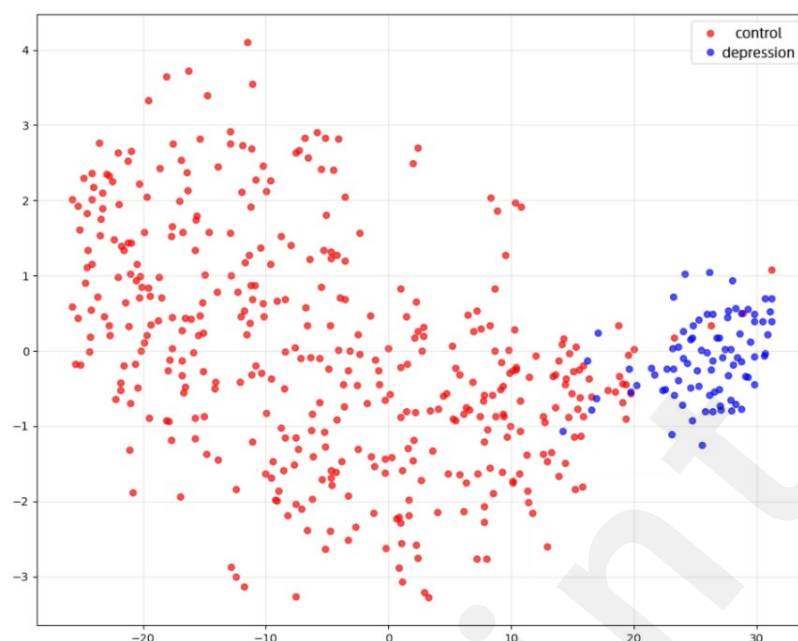
We visualized the contribution degree of different features to the final classification. Column 1 corresponds to the user's sleep-wake rhythm features, Column 18 corresponds to the historical features, and Columns 2–17 correspond to the texts with the strongest depressive state among the texts posted by users. "Expert" represents the contribution degree of the expert model in Moe to the result. A higher contribution to the final discrimination result is indicated by a larger number at the corresponding position and a color that leans more towards yellow, while a lower contribution is shown by a smaller number and a color that leans more towards dark blue. As shown in Figure 5, Expert 1 is sensitive to the features of depressed users, whereas Expert 2 is more sensitive to those of non-depressed users. Although sleep-wake rhythm features and historical features have a relatively stable contribution ratio to the discrimination, the text features of users contribute the most to the final result.

Figure 5. Heatmap of Feature Contribution Analysis



Reduction-Based Validation

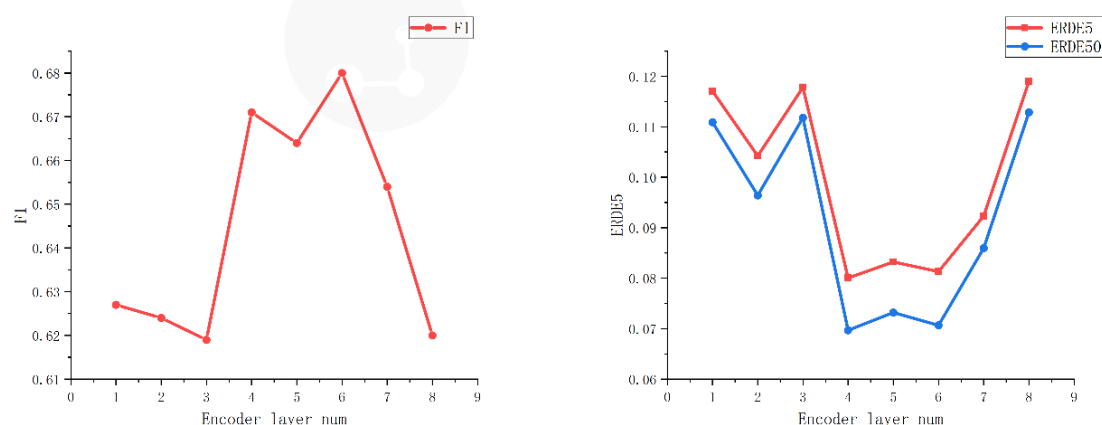
Figure 6. Dimensionality Reduction Visualization of the Distinction Between Depressed and Control Users



This subsection presents the visualization of user features mapped in a two-dimensional space, where the features have been processed using our proposed method. For dimensionality reduction, we adopted t-distributed Stochastic Neighbor Embedding (t-SNE)[35] and Principal Component Analysis(PCA)[36]. Specifically, features of depressed users are displayed in blue, while those of non-depressed users are shown in red. As illustrated in Figure 6, the features of depressed users are clustered in a distinct region. Meanwhile, the features of most non-depressed users are far apart from those of depressed users; in contrast, the difference between the features of a small number of non-depressed users and those of depressed users is not significant.

Hyperparameter Analysis

Figure 7. Performance Changes of the Model in Terms of F1 and ERDE as Parameters Vary.



The most computationally intensive component of the entire model is the encoder layer. To strike a

balance between computational efficiency and accuracy, we conducted experiments on the selection of the number of encoder layers. As shown in the Figure 7. Performance Changes of the Model in Terms of F1 and ERDE as Parameters Vary., when the number of Encoder Layer layers is between 3 and 5, the ERDE values reach the minimum, while the F1 score also peaks in a similar range. Within this range, increasing the number of layers yields limited improvement in model performance; thus, we selected 4 layers, which offer the fastest computation speed.

Discussion

Overall Analysis

The previous section presented the performance of MIND on the eRisk2017 dataset, along with results from ablation studies, case analyses, contribution visualization, and reduction-based validation. Overall, our findings demonstrate that the proposed approach consistently outperforms existing methods across evaluation metrics, with the exception of higher inference time costs. In this section, we provide a detailed analysis of each module's strengths and limitations.

Sleep Dynamics Analysis

In the ablation experiments, removing the SD module resulted in a 17.5% increase in the $ERDE_5$ and a 13.0% increase in the $ERDE_{50}$, while $F1_{ERD}$ dropped by 5.6%. These results indicate that SD's primary contribution lies in enhancing earliness, with a relatively smaller impact on accuracy.

Although the introduction of sleep-wake rhythms modeling provides a novel modality for ERD, the SD module alone yields limited performance, confirming that text features remain the dominant modality, while SD serves as an auxiliary enhancement. This observation aligns with the Contribution Analysis, where sleep-wake rhythms provided stable contributions but were not the most decisive factor compared to the text feature.

Case studies showed that sleep-wake rhythm features clearly distinguished depressed from non-depressed users. Depressed users exhibited irregular, insufficient sleep-wake rhythms, while non-depressed users displayed consistent sleep cycles. Such visualization not only aids classification but also offers clinicians interpretable evidence for interventions. However, the approach has limitations: since sleep-wake rhythms are inferred from aggregated posting timestamps, the model only captures coarse-grained, long-term sleep habits. While effective for users with persistent depressive tendencies, it is less sensitive to short-term depressive episodes, underscoring the need for finer-grained integration with contextual textual analysis.

LLM Depression Profiler Analysis

Replacing the LDP module with template matching in ablation studies caused the $ERDE_5$ to increase by 28.7% and $ERDE_{50}$ by 34.7%, with a 2.3% decrease in $F1_{ERD}$. This demonstrates that the LLM

component substantially improves early detection capability, though gains in accuracy are more modest.

One of the most significant benefits of LDP is its improvement in interpretability. Unlike static filtering, the LLM not only identifies depressive posts but also traces underlying risk factors (e.g., hopelessness, suicidal ideation). This provides actionable insights into users' depressive triggers and key posts.

However, performance was found to be sensitive to the choice of base model. During experiments, different LLMs yielded varying results, with GPT-4o showing the most balanced performance. This suggests that model selection and optimization remain critical directions for future research.

Moreover, reliance on external APIs and longer inference time limits scalability. Deploying locally optimized LLM could address both efficiency and privacy concerns. Historical features, while less impactful overall, contributed significantly at earlier time steps, indicating their importance for very early detection.

Text-Sleep Fusion Analysis

The Text-Sleep Fusion (TSF) module was designed to align and integrate multimodal features. Ablation experiments confirmed its role: removing TSF increased the ERD metric by 9.16% and earliness by 7.2%, although F1 remained stable (fluctuations <1%).

This demonstrates that TSF consistently enhances early warning performance while maintaining accuracy. Its lightweight fusion process effectively balances multimodal collaboration without incurring significant overhead.

Further analysis revealed that when only text features were available, TSF offered no clear advantage over a simple Transformer encoder. This suggests that its benefit primarily arises from the cross-modal alignment between text and sleep-wake rhythm features. Improving TSF's generalization ability to text-only scenarios is a potential avenue for future improvement.

Conclusions

In this study, we proposed a novel method for early risk detection of depression on social media, integrating sleep-wake rhythms modeling and LLM-based text filtering and analysis. Unlike prior work, which exclusively relies on text features for depressive state analysis, our approach incorporates posting timestamps to infer users' sleep-wake rhythms and leverages LLM to improve upon traditional template-matching methods, thereby constructing a more accurate profile of depressive texts.

Experiments on the *eRisk2017* benchmark demonstrate that MIND not only achieves superior performance across multiple evaluation metrics, despite higher inference costs, but also provides enhanced interpretability (via sleep-wake rhythm feature visualization and LLM-based tracing of

latent depressive factors). The sleep–wake rhythm features and the LLM’s identification of latent depressive factors allow the model to trace back depressive risk signals, offering strong support for clinical intervention and treatment.

Overall, MIND establishes a new paradigm for multimodal ERD by combining sleep–wake rhythms and advanced language modeling, breaking the limitation of prior text-only methods—with high practical value for real-world applications in mental health monitoring.

Acknowledgements

This study is supported by the National Natural Science Foundation of China (No. 62163033), Talent Innovation and Entrepreneurship Project of Lanzhou, China (No. 2021-RC-49), at the Gansu Provincial Department of Education: Industrial Support Program Projects (No. 2025CYZC-010). The funding body played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

We acknowledge the support of artificial intelligence tools in completing this thesis. Within the research methodology, LLMs were adopted as a functional component to facilitate the study’s technical implementation. For the thesis text, GPT provided partial linguistic polishing assistance. Notably, the author retains full responsibility for the thesis’s academic integrity, including all research content, analytical processes, and conclusions, which have undergone independent validation.

Conflicts of Interest

none declared

Data Availability

Our research on depression risk may raise ethical concerns. The data used in this study come from a publicly available dataset shared by other researchers. To protect user privacy, all social media data underwent a strict anonymization process conducted by the dataset providers before usage.

We followed all relevant ethical guidelines and legal requirements to ensure that no privacy violations occurred during the research process. The proposed model is not intended as a diagnostic tool but rather as a risk estimation framework to support monitoring and early intervention.

All textual examples shown in this paper were paraphrased by the authors to remove sensitive information and do not represent original user posts.

Authors' Contributions

The author independently designed the research, developed the methodology, conducted experiments, analyzed data, and drafted the entire thesis.

The corresponding author provided critical guidance on research design, thesis structure, and result interpretation.

A fellow student offered auxiliary support such as partial data collection and preprocessing.

Abbreviations

TRD :Traditional Risk Detection,

ERD :Early Risk Detection

MIND:Monitoring of Individual Nighttime Dynamics via LLM Analysis

SD:SleepDynamics

LDP:LLM Depression Profiler

MoE: Mixture of Experts

References

- [1] <https://github.com/YuBicheng/MIND>
- [2] James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A, others. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990--2017: a systematic analysis for the global burden of disease study 2017. *The lancet*; 2018(392):1789--1858. doi:10.1016/S0140-6736(18)32279-7
- [3] Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, et al. (2013) Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLOS Medicine* 10(11): e1001547. doi:10.1371/journal.pmed.1001547
- [4] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. Predicting depression via social media. *Proceedings of the International AAAI Conference on Web and Social Media*;2021: 7(1), 128-137. doi: 10.1609/icwsm.v7i1.14432
- [5] Parapar J, Martin-Rodilla P, Losada DE, Crestani F. Overview of risk 2022: early risk prediction on the internet. *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy, September 5–8, 2022, Proceedings*; 2022:233–256. doi:10.1007/978-3-031-13643-6_18

- [6] Parapar J, Martin-Rodilla P, Losada DE, Crestani F. Overview of erisk 2023: early risk prediction on the internet. *Experimental IR Meets Multilinguality, Multimodality, and Interaction*; 2023:294--315 . doi:10.1007/978-3-031-42448-9_2
- [7] Parapar J, Martin-Rodilla P, Losada DE, Crestani F. Overview of erisk 2024: early risk prediction on the internet. *Experimental IR Meets Multilinguality, Multimodality, and Interaction*; 2024:73--92. doi:10.1007/978-3-031-71908-0_4
- [8] Marco JH, Alonso S, Andani J. Early intervention with cognitive behavioral therapy reduces sick leave duration in people with adjustment, anxiety and depressive disorders. *Journal of Mental Health*; 2020(29):247--255. doi:10.1080/09638237.2018.1521937
- [9] Elman, J.L. Finding structure in time. *Cognitive Science*;1990:14, 179-211. doi:10.1207/s15516709cog1402_1
- [10] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*;2014:1746--1751.doi:10.3115/v1/D14-1181
- [11]Zhang E, Masoud N, Bandegi M,Lull J,Malhan RK.Step attention: sequential pedestrian trajectory prediction.*IEEE Sensors Journal*;2022;vol. 22, no. 8, pp. 8071-8083. doi:10.1109/JSEN.2022.3158271
- [12] Hartvigsen T, Sen C, Kong X, Rundensteiner E. Adaptive-halting policy network for early classification. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; 2019:101--110. doi:10.1145/3292500.3330974
- [13] Bichen Wang, Yuzhe Zi, Yanyan Zhao, Pengfei Deng, and Bing Qin. 2024; ESDM: Early Sensing Depression Model in Social Media Streams. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 6288--6298, Torino, Italia. ELRA and ICCL.
- [14] Trotszek M, Koitka S, Friedrich CM. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. *IEEE Transactions on Knowledge and Data Engineering*; 2020:588--601. doi:10.1109/tkde.2018.2885515
- [15] Zhang Z, Chen S, Wu M, Zhu KQ. Psychiatric scale guided risky post screening for early detection of depression. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*; 2022:5220--5226. doi:10.24963/ijcai.2022/725
- [16] Sadeque F, Xu D, Bethard S. Measuring the latency of depression detection in social media. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*; 2018. doi:10.1145/3159652.3159725

- [17] Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*; 2018(6). doi:10.1007/s13755-018-0046-0
- [18] Yang L, Jiang D, He L, Pei E, Oveneke MC, Sahli H. Decision tree based depression classification from audio video and language information. *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*; 2016:89–96. doi:10.1145/2988257.2988269
- [19] An M, Wang J, Li S, Zhou G. Multimodal topic-enriched auxiliary learning for depression detection. *Proceedings of the 28th International Conference on Computational Linguistics*; 2020:1078–1089. doi:10.18653/v1/2020.coling-main.94
- [20] Chen Z, Deng J, Zhou J, Wu J, Qian T, Huang M. Depression detection in clinical interviews with llm-empowered structural element graph. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*; 2024:8181--8194. doi:10.18653/v1/2024.naacl-long.452
- [21] Nutt D, Wilson S, Paterson L. Sleep disorders as core symptoms of depression. *Dialogues in Clinical Neuroscience*; 2008(10):329--336. doi:10.31887/DCNS.2008.10.3/dnutt
- [22] Franzen PL, Buysse DJ. Sleep disturbances and depression: risk relationships for subsequent depression and therapeutic implications. *Dialogues in Clinical Neuroscience*; 2008(10):473-481.doi:10.31887/DCNS.2008.10.4/plfranzen
- [23] Riemann D, Berger M, Voderholzer U. Sleep and depression — results from psychobiological studies: an overview. *Biological Psychology*; 2001(57):67-103. doi:10.1016/S0301-0511(01)00090-4
- [24] Yasugaki S, Okamura H, Kaneko A, Hayashi Y. Bidirectional relationship between sleep and depression. *Neuroscience Research*; 2025(211):57-64.doi:10.1016/j.neures.2023.04.006
- [25] Akre, Samir, Cohen, Zachary D. Comparing self reported and physiological sleep quality from consumer devices to depression and neurocognitive performance. *npj Digital Medicine*; 2025:92.doi:10.1038/s41746-025-01493-6
- [26] Fei-Fei L, Fergus R, Perona P. A bayesian approach to unsupervised one-shot learning of object categories. *Proceedings Ninth IEEE International Conference on Computer Vision (ICCV)*; 2003:1134--1141. doi:10.1109/ICCV.2003.1238429
- [27] Palatucci M, Hinton GE, Pomerleau D, Mitchell TM. Zero-shot learning with semantic output codes. *Advances in Neural Information Processing Systems 22 (NeurIPS)*; 2009:1410--1418 doi:10.5555/2984093.2984252
- [28] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P,

- Sastry G, Aspell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33 (NeurIPS); 2020:1877--1901 ISBN:9781713829546
- [29] Ashish V, Noam Sr, Niki P, Jakob U, Llion J, Aidan N. G, Łukasz K, Illia P. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*; 2017:6000–6010. doi:10.5555/3295222.3295349
- [30] Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE. Adaptive mixtures of local experts. *Neural Computation*; 1991:79–87. doi:10.1162/neco.1991.3.1.79
- [31] Jacob D, Ming-Wei C, Kenton L, Kristina T. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; 2019:4171–4186. doi:10.18653/v1/N19-1423
- [32] Losada DE, Crestani F. A test collection for research on depression and language use. *Experimental IR Meets Multilinguality, Multimodality, and Interaction*; 2016:28--39 . doi:10.1007/978-3-319-44564-9_3
- [33] Koltcov S, Koltsova O, Nikolenko S. Latent dirichlet allocation. *Proceedings of the 2014 ACM conference on Web science*; 2014. doi:10.1145/2615569.2615680
- [34] Burdisso SG, Errecalde M, Montes-y-Gómez M. T-ss3: a text classifier with dynamic n-grams for early risk detection over text streams. *Pattern Recognition Letters*; 2020:130–137. doi:10.1016/j.patrec.2020.07.001
- [35] Laurens vdM , Geoffrey H. Visualizing data using t-SNE. *Journal of Machine Learning Research*; 2008:2579--2605.
- [36] Felipe LG, Gustavo RF, Henrique FDA, Filipi NS, Cesar HC, Diego RA, Luciano DFC. Principal component analysis: a natural approach to data exploration. *ACM Comput*; 2022:Surv. 54, 4, 34. doi:10.1145/3447755

Supplementary Files

Figures

Task Schematic Diagram. The left side illustrates the TRD task, which predicts a user's depression status using all available user information. In contrast, the right side represents the ERD task, which aims to predict depression status by utilizing early-stage information.

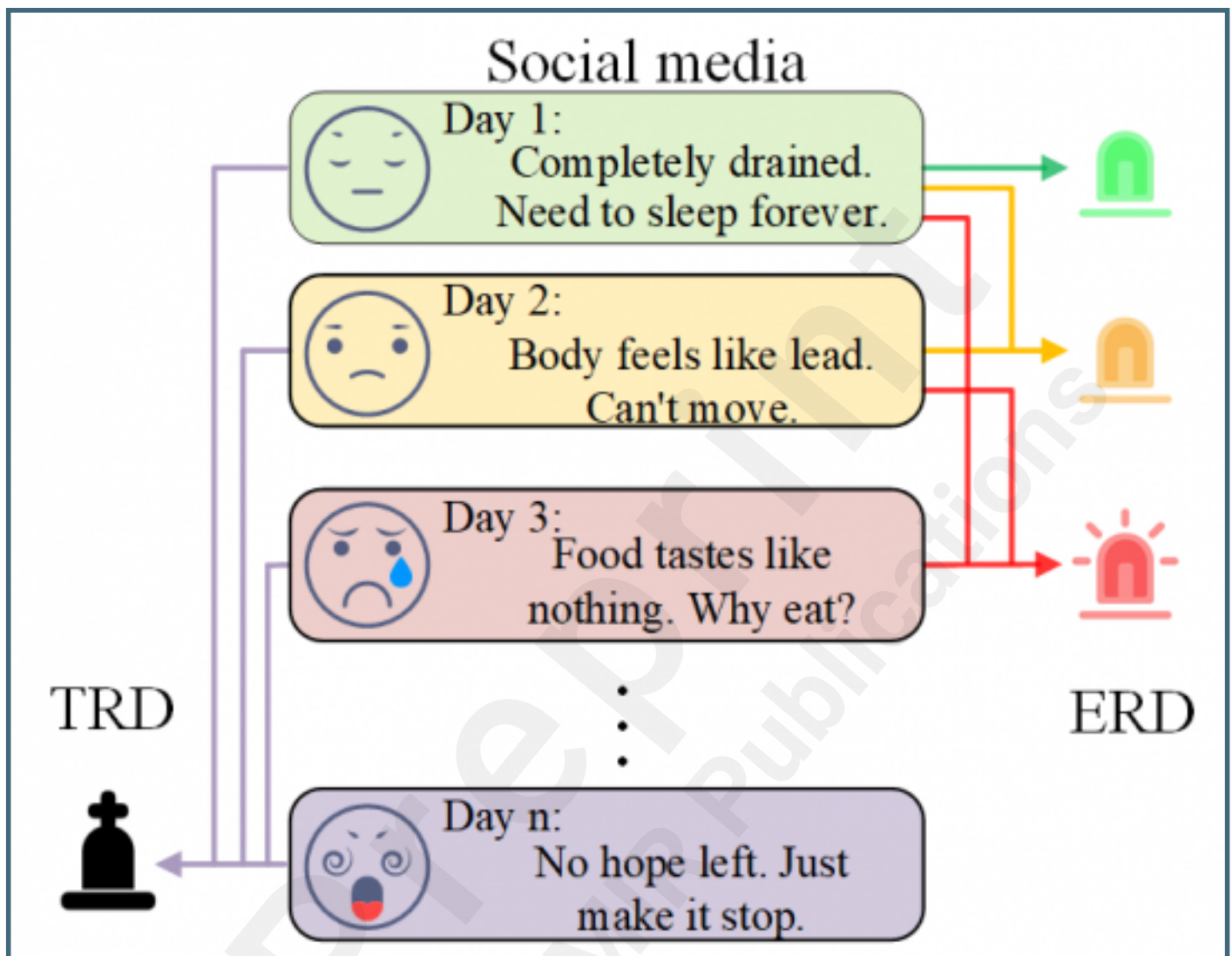
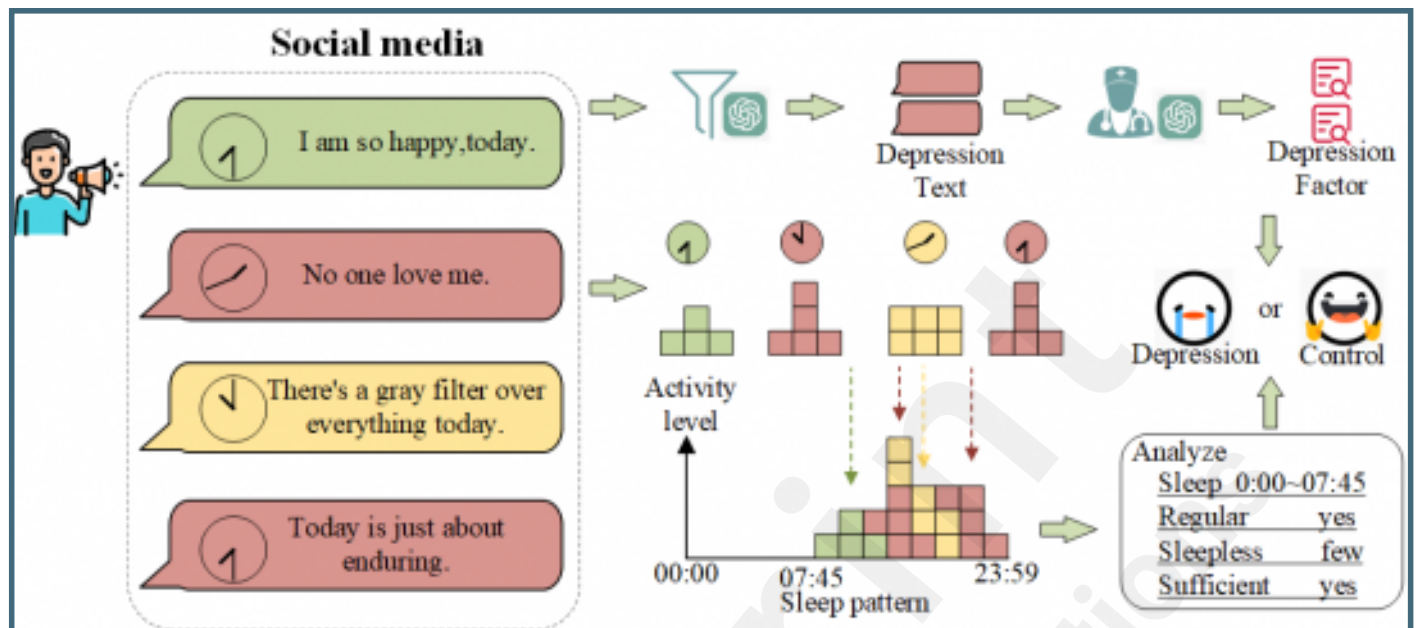


Illustration of innovation points. This figure outlines how user data are transformed into sleep pattern features and user depression profile features.



Overview of the MIND Model.

