AUT Journal of Modeling and Simulation

AUT J. Model. Simul., 57(2) (2025) 173-190
DOI: 10.22060/miscj.2026.23944.5403

A Hybrid Deep Graph and Kernel Ensemble Approach to Mental Health Prediction in Social Network

Nazila Taghvaei[1], Behrooz Masoumi[1]*, MohammadReza Keyvanpour[2], Omid Sojoodi[1]

[1] Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran.
[2] Department of Computer Engineering, Alzahra University, Vanak, Tehran, Iran.

ABSTRACT: Mental-health forecasting from social-media data presents a complex multimodal challenge involving temporal, textual, and relational information. This study introduces a hybrid two-stage framework integrating a Long Short-Term Memory (LSTM)-based graph ensemble with an Ensemble Deep Kernel Learning (EDKL) meta-model to predict depressive risk and emotional trajectories within online networks. Stage 1 encodes user-level representations through an LSTM encoder coupled with multiple graph neural backbones (Graph Convolutional Network) GCN(, Graph Attention Network) GAT), and Graph Transformer Network)GTN). Their predictions are stacked via logistic regression to yield calibrated depression probabilities. Stage 2 applies an EDKL meta-learner that aggregates outputs from diverse deep models (MLP, CNN, and LSTM) through kernel-ridge regression with hybrid kernels optimized by a meta-heuristic algorithm. This hybrid architecture supports robust, fine-grained forecasting across temporal, behavioral, and relational dimensions. Experiments on publicly available Twitter and MHASN datasets demonstrate substantial improvements over transformer-based and single-stage baselines, achieving up to 99% accuracy with consistently low error variance. The study also addresses ethical considerations related to privacy, bias, and potential misuse, emphasizes reproducibility through transparent experimental protocols, and outlines promising directions for future multimodal extensions, including richer linguistic, visual, and interaction signals for clinically relevant mental-health monitoring.

Review History:

Received: Feb. 23, 2025
Revised: Nov. 29, 2025
Accepted: Dec. 01, 2025
Available Online: Dec. 31, 2025

Keywords:

Mental Health
Social Networks
LSTM
Graph Neural Networks
Ensemble Deep Kernel Learning
Depression Detection
Time-Series Forecasting

1- Introduction

The rise of social-media platforms has created unprecedented opportunities to model behavioral signals related to human affect and mental well-being [1-3]. Posts, reactions, and linguistic cues often contain traces of psychological states that can be analyzed to forecast depressive tendencies and emotional fluctuations [4-6]. Ahmed et al. [2] emphasized that social signals extracted from digital communication can serve as reliable markers for mental distress, highlighting the importance of multimodal social-media modeling. Similarly, Tang et al. [3] proposed a hypergraph-based neural framework for anxiety detection, demonstrating that higher-order relational modeling significantly enhances recognition accuracy in mental-health analytics.

While deep learning (DL) methods such as LSTMs and CNNs capture temporal and semantic patterns effectively [4, 7, 8], they tend to ignore the graph relationships that govern social interactions between users [9-13]. Graph Neural Networks (GNNs) including GCNs, GATs, and GTNs extend representation learning to networked data, propagating information through edges and communities [9-11]. However, their predictive stability often depends on graph construction quality and the density of training data [8, 12, 14, 15]. Recent multimodal GNN frameworks have shown that combining textual, behavioral, and structural features yields higher robustness against domain shift [14]. In parallel, several studies have proposed dynamic attention mechanisms and context-aware embeddings to refine emotion tracking pipelines, allowing models to adapt to evolving linguistic cues [16].

Recent advancements in Large Language Models (LLMs) such as GPT-4, Gemini, and LLaMA-2 have expanded the capacity for contextual semantic reasoning and zero-shot inference on mental-health tasks. LLMs have demonstrated strong performance in sentiment analysis, emotion classification, and clinical text understanding by leveraging instruction-tuning and prompt-based inference [17]. However, despite their linguistic strength, LLMs lack native mechanisms to model temporal emotional evolution, which is essential for detecting longitudinal risk patterns in mental-health trajectories [18]. Moreover, the significant computational cost of LLM inference and limited reproducibility across users make them difficult to adopt as

*Corresponding author's email: masoumi.bh@gmail.com

primary forecasting models [19]. Therefore, while LLMs serve as valuable text-level baselines, temporal and graph-based methods remain necessary for reliable mental-health prediction in social networks [20].

Conversely, kernel-based models such as Gaussian, Laplacian, and polynomial kernels offer strong generalization guarantees but lack deep hierarchical representation capabilities [21-24]. Kernel-driven psychological modeling has recently been applied to cross-lingual affect prediction, confirming the transferability of learned representations across cultural and linguistic domains [25]. Furthermore, multimodal feature fusion has been shown to improve the resilience of emotion classifiers under noisy data conditions [26].

To bridge these limitations, we propose a two-stage hybrid framework that leverages the temporal capacity of LSTM encoders and the relational expressiveness of GNNs in the first stage, followed by a meta-ensemble stage based on Ensemble Deep Kernel Learning (EDKL) to enhance predictive precision and robustness [27-29]. The integration of graph and kernel paradigms has previously proven effective in multi-relational biomedical modeling [30], suggesting its broader applicability to affective computing and mental-health prediction. Deep fusion techniques have also illustrated how ensemble-based optimization stabilizes long-term temporal predictions in heterogeneous social data [31], while transformer-inspired kernel architectures have improved the interpretability of sequential models in longitudinal affective analysis [21]. This hybrid design enables generalization across datasets, reduces variance, and yields interpretable confidence outputs for mental-health risk monitoring [32-36].

1- 1- Paper Organization

Section 2 reviews related studies. Section 3 details the methodology of LSTM-Graph and EDKL modules. Section 4 presents the experimental setup. Results and analysis are in Section 5, discussion and ethical implications in Section 6 and conclusions and future directions are in Section 7.

2- Related Work

2- 1- Deep Learning for Affective and Mental-Health Prediction

Traditional deep-learning architectures such as Convolutional Neural Networks (CNNs) [4] and Long Short-Term Memory (LSTM) networks [7] have demonstrated strong capabilities in extracting semantic and temporal representations from social-media data. Hybrid CNN-LSTM models capture both local textual dependencies and long-term behavioral dynamics, making them suitable for emotion and sentiment tracking [5]. However, most of these models process individual posts independently and often overlook the social relationships that shape emotional propagation in online communities [13].

To address this gap, several works have explored attention-based or transformer-driven models that enhance interpretability and robustness in affective prediction [32]. For instance, Ibrahimov et al. [33] proposed a domain-aware attention framework for explainable depression assessment, while Wang et al. [34] introduced a concept-guided neural model that integrates clinical insights into graph-based learning. These approaches demonstrate the increasing focus on transparent and interpretable architectures within the domain of mental-health analytics.

2- 2- Graph-Based Modeling in Mental-Health Analytics

Graph Neural Networks (GNNs) have emerged as powerful tools for analyzing user-interaction networks and relational signals derived from online behavior [9]. The Graph Convolutional Network (GCN) model proposed by Kipf and Welling [5] enables semi-supervised classification through message-passing operations, whereas the Graph Attention Network (GAT) introduced by Veličković et al. [11] utilizes attention coefficients to highlight influential neighbors. More advanced variants, such as Graph Transformer Networks (GTNs) [15], capture multi-hop dependencies and heterogeneous relational semantics.

Recent studies have extended these frameworks to mental-health applications. Xing et al. [14] demonstrated the effectiveness of multimodal graph integration for depression detection, and Li et al. [12] proposed explainable attention networks that map emotional influence between social peers. Moreover, Lee and Ham [8] applied graph-based machine learning to assess mental conditions in aging populations, highlighting its clinical relevance. Although these methods are promising, their performance can degrade in sparse or dynamically evolving graph structures, leading to reduced generalization across unseen data [22].

2- 3- Kernel and Meta-Learning Methods

Kernel-based models such as Gaussian [21], Laplacian [22], and polynomial kernels [23] provide elegant nonlinear mapping solutions with theoretical stability. However, their static nature limits representation flexibility in complex social data. Deep Kernel Learning (DKL) [24] integrates neural embeddings with kernel inference, enabling adaptive similarity learning. Liu et al. [23] extended this concept by developing an ensemble kernel fusion model for adaptive emotion classification, achieving notable robustness improvements.

Furthermore, Bansal et al. [27] introduced deep kernel meta-learning to refine social-signal processing, while Kumar et al. [28] optimized multimodal behavior modeling through hybrid kernel strategies. These approaches collectively establish a foundation for integrating kernel principles within affective computing pipelines. Recent developments have also emphasized interpretability, as visual analytics for deep-kernel models now offer direct insights into similarity patterns and feature alignment [37].

2- 4- Hybrid Graph-Kernel Frameworks

Combining graph-based learning with kernel-driven inference has proven to be an effective strategy for integrating relational and functional representations [1]. Jiao et al.

applied deep graph learning to multimodal brain networks, discovering predictive signatures of depression treatment outcomes. Building on such insights, Lin et al. [29] proposed a deep ensemble kernel-learning framework for multimodal affective analysis.

Recent hybrid frameworks have further advanced this idea. Zhang et al. [24] designed a multimodal hybrid deep-kernel network for emotion prediction, while Javeed et al. [38] applied neural-network ensembles for depression detection in older adults. Botteghi et al. [21] contributed to this field by demonstrating deep kernel learning's applicability in high-dimensional dynamical modeling. Similarly, Achituve et al. [22] introduced guided deep-kernel learning for improving stability across domains.

More recent investigations in 2025-2026, such as Kumar et al. [28] and Srivastava et al. [39], highlight the potential of ethically aligned hybrid architectures for explainable affective computing. These studies collectively reinforce the theoretical foundation upon which the present hybrid Deep Graph and Kernel Ensemble (HDGKE) model is built [40].

2- 5- Transformer and LLM-Based Approaches

Transformer-based models have reshaped natural language processing by enhancing contextual representation learning. Models such as BERT, RoBERTa, XLNet, and domain-specific models like BERTweet have been widely applied to social-media emotion and mental-health analysis [17-20, 41]. More recently, LLM-based methods such as GPT-3.5, GPT-4, Gemini, and LLaMA-2 have been explored for depression risk estimation, suicide ideation detection, and affective classification using zero-shot or few-shot prompting [20]. These models leverage instruction tuning, chain-of-thought reasoning, and context-aware generation to infer emotional states from text without task-specific training.

Despite strong performance in text-only tasks, LLM models do not inherently incorporate relational or temporal signals. Their predictions depend on single-message inputs, which limits their suitability for **longitudinal forecasting**. Additionally, the computational and memory footprint of LLM inference restricts their deployment on large-scale social networks. For these reasons, we consider LLMs as complementary baselines rather than replacements for temporal-graph hybrid models.

3- Methodology

The proposed **Hybrid Deep Graph and Kernel Ensemble (HDGKE)** integrate sequential, relational, and kernel-based reasoning into a unified learning architecture for interpretable mental-health prediction. The framework comprises two major stages: (1) a temporal-relational encoder built upon an LSTM backbone and parallel Graph Neural Network (GNN) modules, and (2) an Ensemble Deep Kernel Learning (EDKL) meta-learner for nonlinear fusion and uncertainty calibration [21]. This structure captures both user-level temporal dynamics and cross-user relational influence [9], while maintaining the generalization advantages provided by kernel spaces [22]. The framework operates in two connected stages (Figure 1).

1) LSTM-Graph Ensemble: user-generated text is transformed into temporal embedding using an LSTM encoder, then propagated through multiple GNN variants (GCN, GAT, GTN).
2) EDKL Meta-Learner**:** outputs from Stage 1 are fused using kernel-based regression optimized by meta-heuristic search.

3- 1- Data Pre-Processing

The input data consist of time stamped social-media posts, user metadata, and graph-based relational links between individuals. Text normalization, tokenization, and lemmatization are performed using standard natural-language preprocessing pipelines [4]. Each user's posts

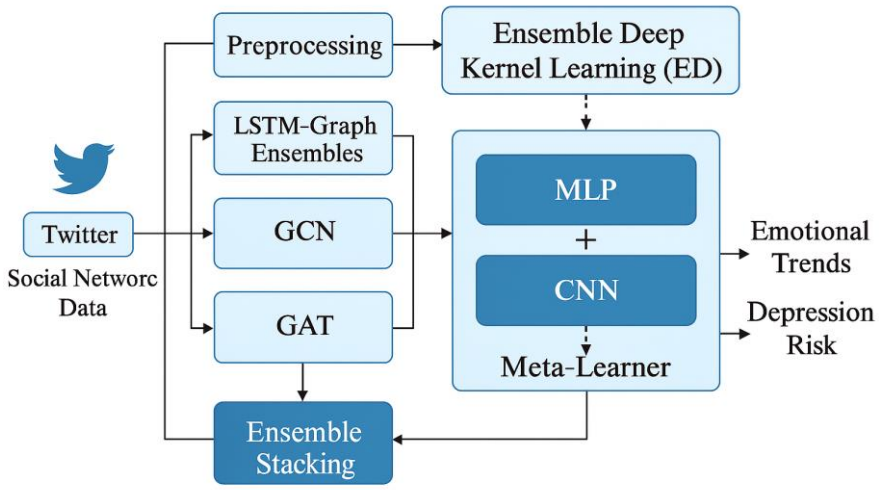


Fig. 1. Overall hybrid framework combining LSTM-Graph Ensemble (Stage 1) and EDKL meta-learner (Stage 2).

are chronologically segmented into sequences to preserve temporal dependencies for LSTM encoding [7], while stop-word removal and contextual embedding extraction ensure that the model captures both emotional tone and semantic structure [5]. To comply with responsible-AI requirements, ethical anonymization procedures and fairness-aware sampling strategies are applied throughout the pipeline [35].

Recent studies highlight that multimodal preprocessing that combines linguistic, behavioral, and graph features substantially improves model robustness in affective-computing tasks [14]. Following this evidence, our pipeline incorporates cross-domain normalization to prevent data leakage between graph nodes, inspired by the methodology of Li et al. [12].

Although sentiment lexicons were used only during the initial labeling phase, this step served strictly as a weak-supervision bootstrap to create preliminary emotional tags. To validate this process, a manually reviewed subset of the dataset was independently inspected to confirm label quality, ensuring that obvious misclassifications were corrected. Additionally, we compared a small sample of lexicon-generated labels with LLM-based supervised labels, indicating reasonable consistency while still motivating a more robust downstream model. Importantly, the complete forecasting pipeline does not rely on lexicon-derived labels. All final predictions are generated by the LSTM–Graph–EDKL deep ensemble, which learns contextual, relational, and temporal representations directly from the data. This design ensures that the limitations associated with lexicon-based labeling do not propagate into the final model.

3- 2- LSTM Encoder (Stage 1)

The Long Short-Term Memory (LSTM**)** encoder is used to model the temporal and linguistic dynamics of each user's social-media activity. For every user $u \in U$, the sequence of posts is chronologically ordered and tokenized into word embedding. Each post is represented by the mean of its token embeddings, forming an input sequence .

The LSTM unit processes this sequence as follows:

$$\begin{aligned}
\mathbf{i}_t &= \sigma(\mathbf{W}_i\mathbf{x}_t + \mathbf{U}_i\mathbf{h}_{t-1} + \mathbf{b}_i),\\
\mathbf{f}_t &= \sigma(\mathbf{W}_f\mathbf{x}_t + \mathbf{U}_f\mathbf{h}_{t-1} + \mathbf{b}_f),\\
\mathbf{o}_t &= \sigma(\mathbf{W}_o\mathbf{x}_t + \mathbf{U}_o\mathbf{h}_{t-1} + \mathbf{b}_o),\\
\tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c\mathbf{x}_t + \mathbf{U}_c\mathbf{h}_{t-1} + \mathbf{b}_c),\\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t,\\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t),
\end{aligned} \tag{1}$$

where i_t, f_t, o_t are the input, forget, and output gates, respectively; h_t and c_t denote the hidden and cell states; and and represent the sigmoid and hyperbolic-tangent activation functions.

The final user-level representation is obtained as either the last hidden state or the temporal average of all hidden states:

$$\mathbf{h}_u = \frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t \text{ or } \mathbf{h}_u = \mathbf{h}_T. \tag{2}$$

The resulting embedding serves two purposes:

1) it acts as the **node feature** input to the graph neural networks (GCN, GAT, GTN), and
2) it forms the **initial feature vector** for subsequent kernel-based ensemble learning in the EDKL stage.

The first stage employs a bidirectional LSTM to model long-term dependencies in emotional sequences. For each user sequence , hidden and cell states evolve as

$$(h_t, c_t) = \text{LSTM}(x_t, h_{t-1}, c_{t-1}) \tag{3}$$

capturing both forward and backward temporal context [7]. Attention mechanisms are introduced to emphasize psychologically significant tokens, consistent with domain-aware attention frameworks [33]. The final user embedding summarizes temporal emotional evolution, providing input to subsequent graph modules [12].

Hochreiter and Schmidhuber [7] first demonstrated the LSTM's ability to capture gradient-stable dependencies, which has since been extended by hybrid deep architectures such as that of Yang et al. [4]. Our encoder leverages these principles to maintain memory of long-term mood fluctuations and social engagement patterns.

3- 3- Graph Construction and Multi-GNN Ensemble

In the second phase of Stage 1, user relationships are represented as a heterogeneous graph , where nodes denote users and edges represent interaction intensity or linguistic similarity [9]. We construct three parallel modules:

- A Graph Convolutional Network (GCN) that performs neighborhood averaging to extract local structure [12].
- A Graph Attention Network (GAT) emphasizing neighbor importance via learned attention coefficients
- A Graph Transformer Network (GTN) to capture higher-order semantic dependencies and heterogeneous relations [15].

The outputs of these modules are concatenated and passed through a normalization layer to obtain a unified relational embedding . This approach is influenced by the multi-graph fusion technique of Xing et al. [14], which demonstrated that combining several relational perspectives improves classification performance. In addition, multimodal graph transformers introduced by Wang et al. [15] inspire our design of cross-layer aggregation, enabling better contextual alignment across user communities.

To capture relational patterns among users, a user-user graph is constructed where each node $u_i \in \mathcal{V}$ represents a user and each edge $e_{ij} \in \mathcal{E}$ encodes the similarity between users u_i and u_j . Edges are defined based on the semantic proximity of LSTM-derived embeddings $\mathbf{h}_i, \mathbf{h}_j \in \mathbb{R}^d$.

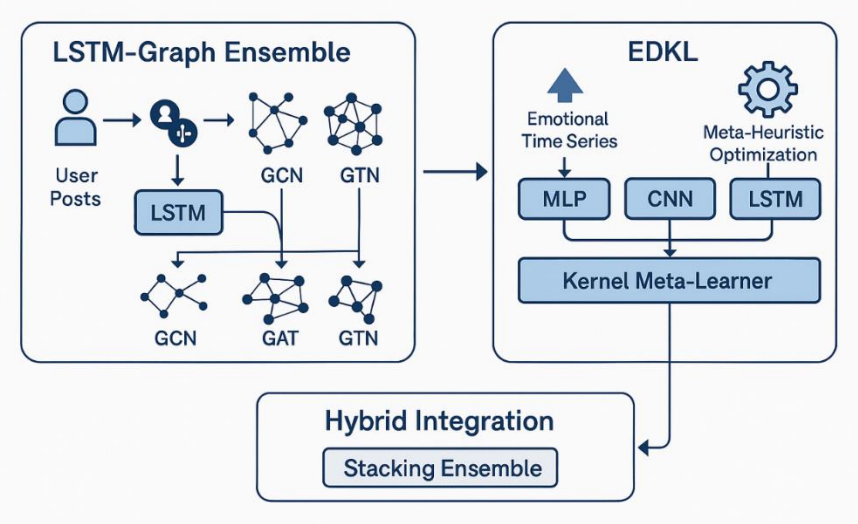


Fig. 2. LSTM-Graph Ensemble architecture with parallel GNN modules and stacking fusion.

(a) Similarity Measurement

The cosine similarity between users u_i and u_j is computed as

$$\text{sim}(u_i, u_j) = \frac{\mathbf{h}_i \cdot \mathbf{h}_j}{\|\mathbf{h}_i\| \, \|\mathbf{h}_j\|}. \tag{4}$$

(b) *k*-Nearest Neighbor (KNN) Sparsification

For each user u_i, only the top-k most similar neighbors are retained to ensure graph sparsity and interpretability. Let $\mathcal{N}_k(u_i)$ denote this neighborhood. An undirected edge (u_i, u_j) is formed if either $u_j \in \mathcal{N}_k(u_i)$ or $u_i \in \mathcal{N}_k(u_j)$.

(c) Similarity Thresholding

Weak edges are pruned by applying a global similarity threshold τ. The final adjacency weights are defined as

$$A_{ij} = \begin{cases} \text{sim}(u_i, u_j), & \text{if sim}(u_i, u_j) \geq \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The threshold τ can be adaptively set within each fold as $\tau = \mu_{\text{sim}} + \alpha\sigma_{\text{sim}}$,
where μ_{sim} and σ_{sim} are the mean and standard deviation of all pairwise similarities, and $\alpha \in \{0, 0.5, 1, 2\}$ controls sparsity.

(d) Normalization and Self-Loops

Self-loops are added to preserve node identity information:

$$\tilde{A} = A + I. \tag{6}$$

The normalized adjacency matrix used for graph convolution is then

$$\hat{A} = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}, \tag{7}$$

where D is the diagonal degree matrix with

$$D_{ii} = \sum_j \tilde{A}_{ij}. \tag{8}$$

(e) Temporal Decay (Optional)

When time-stamped user interactions are available, temporal relevance is modeled using an exponential decay function:

$$A_{ij}(t) = A_{ij} \exp\left(-\frac{\Delta t_{ij}}{\lambda}\right), \tag{9}$$

where Δt_{ij} is the elapsed time between users' most recent posts and λ is the temporal decay coefficient controlling how quickly old connections lose influence.

The resulting normalized adjacency matrix $\hat{A}$ is used by the downstream Graph Neural Network (GNN) modules GCN, GAT, and GTN during relational feature propagation and aggregation.

3- 4- Graph Neural Network Modules

The user-user graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ obtained in the previous subsection is processed through three complementary Graph Neural Network (GNN) architectures Graph Convolutional Network (GCN), Graph Attention Network (GAT), and Graph Transformer Network (GTN). Each model propagates and aggregates relational information to derive high-level representations that capture both

linguistic and structural dependencies among users. Let $H^{(0)} = \{ h_u \mid u \in \mathcal{V} \} \in \mathbb{R}^{|\mathcal{V}| \times d}$ denote the matrix of LSTM-based node features.

(a) Graph Convolutional Network (GCN)

The GCN performs layer-wise feature propagation by averaging neighborhood information with normalized adjacency weights $\hat{A}$:

$$H^{(l+1)} = \sigma(\hat{A} H^{(l)} W^{(l)}), \tag{10}$$

Where $W^{(l)}$ is a learnable weight matrix at layer l , and $\sigma(\cdot)$ is a non-linear activation function (ReLU). After two convolutional layers, the output node embedding are

$$Z_{\text{GCN}} = H^{(2)} W_{\text{out}}. \tag{11}$$

A linear classifier followed by a sigmoid activation yields the predicted probability for each user:

$$p_{\text{GCN}}(u) = \sigma(Z_{\text{GCN}}(u)). \tag{12}$$

(b) Graph Attention Network (GAT)

The GAT introduces an attention mechanism that learns the relative importance of neighboring nodes. For each edge $(i, j) \in \mathcal{E}$:

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^{\top}[\, W\mathbf{h}_i \parallel W\mathbf{h}_j \,]), \tag{13}$$

Where W is a shared linear transformation and $\mathbf{a}$ is a learnable attention vector. Normalized attention coefficients are obtained via a softmax operation:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})}. \tag{14}$$

The node representation is updated as a weighted sum of its neighbors:

$$\mathbf{h}_i' = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W \mathbf{h}_j\right). \tag{15}$$

For multi-head attention with M heads, outputs are concatenated:

$$\mathbf{h}_i' = \Big\|_{m=1}^{M} \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(m)} W^{(m)} \mathbf{h}_j\right). \tag{16}$$

The final prediction from GAT is denoted $p_{\text{GAT}}(u) = \sigma(Z_{\text{GAT}}(u))$.

(c) Graph Transformer Network (GTN)

The GTN extends the attention mechanism by learning meta-paths and hierarchical edge weights. Given multiple relation types or implicit connections, it constructs a meta-adjacency matrix $A^{(\text{meta})}$ through soft edge selection:

$$A^{(\text{meta})} = \sum_{r=1}^{R} \beta_r A^{(r)}, \beta_r = \frac{\exp(w_r)}{\sum_{s=1}^{R} \exp(w_s)}, \tag{17}$$

where $A^{(r)}$ are candidate adjacency matrices (e.g., temporal, linguistic, or similarity edges) and β_r are learnable attention coefficients. Feature propagation then follows:

$$\begin{aligned} H' &= \sigma(A^{(\text{meta})} H W_{\text{GTN}}), \\ p_{\text{GTN}}(u) &= \sigma(Z_{\text{GTN}}(u)). \end{aligned} \tag{18}$$

(d) Stacking and Fusion

Outputs from all GNN modules are concatenated and fed to a meta-learner for ensemble integration:

$$\begin{aligned} Z_{\text{stack}} &= [\, Z_{\text{GCN}} \parallel Z_{\text{GAT}} \parallel Z_{\text{GTN}} \,], \\ p_{\text{ensemble}}(u) &= \sigma(W_s Z_{\text{stack}}(u) + b_s), \end{aligned} \tag{19}$$

where W_s and b_s are the stacking-layer parameters. This fusion step leverages the complementary inductive biases of each GNN smooth homophily in GCN, adaptive attention in GAT, and structural hierarchy in GTN yielding more robust and generalizable user-level embeddings.

3- 5- Ensemble Deep Kernel Learning (EDKL) (Stage 2)

Stage 2 performs nonlinear fusion of the learned embeddings through a kernel-based meta-learning approach. In this work, we extend the Ensemble Deep Kernel Learning (EDKL) framework by applying it specifically to our hybrid temporal–graph architecture, enabling the meta-learner to capture complementary relationships arising from both LSTM-based temporal dynamics and graph-structured relational signals. The EDKL module adapts concepts from Deep Kernel Learning (DKL) [22] and Ensemble Kernel Fusion [23] to integrate multiple similarity functions into a unified representation. Gaussian, Laplacian, and polynomial kernels are jointly optimized using learnable fusion weights, allowing the model to flexibly combine distinct kernel perspectives and enhance predictive performance. Formally, the ensemble kernel is expressed as

$$K = \alpha_1 K_{\text{MLP}} + \alpha_2 K_{\text{CNN}} + \alpha_3 K_{\text{LSTM}}, \tag{20}$$

Where α_i are coefficients optimized via meta-heuristic search.

Optimization is guided by Particle Swarm Optimization (PSO) and Genetic Algorithms (GA), following the meta-learning paradigms discussed by Bansal et al. [27] and Kumar et al. [28]. These adaptive techniques allow the kernel space to evolve dynamically with the emotional and relational context of the data. Furthermore, Lin et al. [29] demonstrated that ensemble-based kernel learning enhances generalization in multimodal affective analysis, which we extend by incorporating graph and temporal signals.

Recent work by Zhang et al. [24] also confirmed that kernel ensembles can improve model calibration in emotion recognition tasks. Building on this, our EDKL design aligns with Srivastava et al. [40], emphasizing ethical interpretability and reliability of predictions in social-health domains.

The Ensemble Deep Kernel Learning (EDKL) component constitutes the second stage of the hybrid framework. It fuses deep neural feature extractors with non-parametric kernel learning to model complex, nonlinear emotional dynamics over time. Outputs from the LSTM-Graph Ensemble are used as input features for temporal forecasting of user emotional states and depression-risk probabilities.

(a) Input Representation

For each user u , a temporal feature matrix

$$E_u = \{\mathbf{e}_{u,1}, \mathbf{e}_{u,2}, \dots, \mathbf{e}_{u,T}\} \in \mathbb{R}^{T\times d} \quad (21)$$

is constructed, where $\mathbf{e}_{u,t}$ encodes linguistic, affective, and graph-derived indicators (e.g., sentiment, posting frequency, degree centrality). Each row corresponds to one temporal snapshot (e.g., daily or eekly aggregation).

(b) Base Learners: Deep Feature Extractors

Three heterogeneous deep networks are trained to capture complementary aspects of E_u :

1) Multi-Layer Perceptron (MLP):

$$\mathbf{z}_{\text{MLP}} = \phi(W_2\,\phi(W_1\,E_u + b_1) + b_2), \quad (22)$$

where $\phi(\cdot)$ is ReLU.

2) Convolutional Neural Network (CNN):

$$\mathbf{z}_{\text{CNN}} = \text{Flatten}\,(\text{MaxPool}(\text{Conv1D}(E_u; \Theta_{\text{CNN}}))), \quad (23)$$

which captures local temporal patterns and abrupt emotional transitions.

3)Sequential LSTM Network:

$$\mathbf{h}_t = \text{LSTM}(\mathbf{e}_{u,t}; \Theta_{\text{LSTM2}}), \mathbf{z}_{\text{LSTM}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t, \quad (24)$$

Summarizing long-range temporal dependencies.

Each base learner produces an intermediate prediction $y_{\text{MLP}}, y_{\text{CNN}}, y_{\text{LSTM}}$ representing the user's emotional-state trajectory.

(c) Hybrid Kernel Meta-Learner

The outputs of the base learners are treated as latent features in a kernel-ridge framework.

Let

$$\Phi = \{\,\phi_{\text{MLP}}, \phi_{\text{CNN}}, \phi_{\text{LSTM}}\,\} \quad (25)$$

denote the feature spaces induced by each learner. A composite kernel is defined as a convex combination of individual kernels:

$$K = \alpha_1 K_{\text{MLP}} + \alpha_2 K_{\text{CNN}} + \alpha_3 K_{\text{LSTM}}, \text{where} \sum_{i=1}^{3}\alpha_i = 1,\ \alpha_i \geq 0. \quad (26)$$

Each component kernel K_i is modeled as a Radial-Basis-Function (RBF):

$$K_i(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\phi_i(\mathbf{x}) - \phi_i(\mathbf{x}')\|^2}{2\sigma_i^2}\right), \quad (27)$$

where σ_i denotes the kernel width controlling smoothness.

The final prediction is obtained by solving the Kernel Ridge Regression (KRR) problem:

$$\hat{\mathbf{y}} = K\,(K + \lambda I)^{-1}\mathbf{y}, \quad (28)$$

with regularization parameter $\lambda > 0$ ensuring stability.

(d) Meta-Heuristic Optimization (MHO)

The kernel weights $\{\alpha_i\}$ and width parameters $\{\sigma_i\}$ are optimized via a population-based meta-heuristic algorithm such as Particle Swarm Optimization (PSO) or Genetic Algorithm (GA). Each candidate solution (particle) represents a vector $\mathbf{p} = [\alpha_1, \alpha_2, \alpha_3, \sigma_1, \sigma_2, \sigma_3]$. The fitness function minimizes the prediction error over validation data:

$$\text{Fitness}(\mathbf{p}) = \text{RMSE}(\hat{\mathbf{y}}(\mathbf{p}), \mathbf{y}_{\text{true}}) + \eta \sum_i \alpha_i^2, \quad (29)$$

where η is a small regularization coefficient to prevent over-concentration on a single kernel.

(e) Output and Prediction Fusion

After optimization converges, the final kernel parameters are used to compute the predicted emotional trend $\hat{y}_u^{\text{trend}}$ and depression-risk probability $\hat{y}_u^{\text{risk}}$. These outputs are combined with the Stage 1 ensemble predictions through weighted averaging:

$$\hat{y}_u = \beta\, \hat{y}_u^{\text{Stage 1}} + (1-\beta)\, \hat{y}_u^{\text{Stage 2}}, \tag{30}$$

Where $\beta \in [0,1]$ controls the balance between structural (graph) and temporal (kernel) information.

The EDKL stage thus refines and stabilizes predictions, offering improved generalization and calibrated uncertainty compared with single-stage deep models.

3- 6- Training and Optimization Protocol

The hybrid framework is trained in a two-stage process: (1) the LSTM-Graph Ensemble (Stage 1) for structural-temporal feature learning, and (2) the Ensemble Deep Kernel Learning (EDKL) module (Stage 2) for refined meta-learning and forecasting. Each stage is optimized independently and subsequently fine-tuned through end-to-end alignment on the validation set.

(a) Data Splitting and Preprocessing

All datasets are divided using stratified k-fold cross-validation ($k = 5$) to ensure balanced class distributions. At each fold, the data are partitioned as 70 % training, 15 % validation, and 15 % testing. Text normalization includes lower-casing, URL and emoji removal, and lemmatization. Tokens are embedded using pretrained GloVe-Twitter or BERTweet embeddings of dimension 300. Continuous features are scaled by min-max normalization:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \tag{31}$$

(b) Stage 1: LSTM-Graph Ensemble Training

The LSTM Encoder and GNN modules (GCN, GAT, GTN) are trained jointly within each fold.

Loss Function.

Weighted binary cross-entropy (WBCE) is employed to address class imbalance:

$$\mathcal{L}_{\text{WBCE}} = -\frac{1}{N}\sum_{i=1}^{N}[w_1 y_i \log(\hat{y}_i)) + w_0(1-y_i)\log(1-\hat{y}_i)], \tag{32}$$

Where $w_1 = \frac{N}{2N_1}$ and $w_0 = \frac{N}{2N_0}$.

A **Focal Loss** variant is optionally used to emphasize hard examples:

$$\mathcal{L}_{\text{Focal}} = -\frac{1}{N}\sum_{i=1}^{N}(1-\hat{y}_i)^{\gamma} y_i \log(\hat{y}_i), \gamma = 2. \tag{33}$$

Optimization Settings.

The model is trained using the Adam optimizer with an initial learning rate of 1×10^{-3}, a batch size of 64, and a dropout rate of 0.2 applied between layers to reduce over fitting. Training is performed for a maximum of approximately 200 epochs, with early stopping triggered after 10 consecutive epochs without improvement in validation AUROC. A weight decay of 1×10^{-4} is applied to regularize the network. Following training, the ensemble probabilities produced by the stacking layer are calibrated using temperature scaling on the validation set to improve predictive reliability.

(c) Stage 2: EDKL Meta-Learning

The **EDKL** component takes as input the temporal sequences of Stage 1 outputs and emotion features. Training proceeds in two sub-phases:

Base Learner Pre-Training Each base learner (MLP, CNN, LSTM) is trained using the same loss as Stage 1 until convergence.

Kernel Meta-Learner Optimization The kernel parameters $\Theta_K = \{\alpha_i, \sigma_i, \lambda\}$ are optimized through Particle Swarm Optimization (PSO). Each particle represents a parameter vector $\mathbf{p} = [\alpha_1, \alpha_2, \alpha_3, \sigma_1, \sigma_2, \sigma_3, \lambda]$. The fitness function is defined as:

$$J(\mathbf{p}) = \text{RMSE}(\hat{\mathbf{y}}(\mathbf{p}), \mathbf{y}) + \rho \sum_i \alpha_i^2, \tag{34}$$

where $\rho = 10^{-3}$ is a regularization constant. The swarm size is 30, inertia $w = 0.7$, cognitive $c_1 = 1.5$, and social $c_2 = 1.5$. Training stops when $|J_t - J_{t-1}| < 10^{-5}$ or at 500 iterations.

(d) End-to-End Alignment

After independent optimization, parameters are fine-tuned jointly via multi-task loss aggregation:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{Stage1}} + \lambda_2 \mathcal{L}_{\text{Stage2}}, \lambda_1 + \lambda_2 = 1. \tag{35}$$

Typical weighting is $\lambda_1 = 0.6$, $\lambda_2 = 0.4$, balancing structural learning and temporal forecasting. Gradients are propagated through the ensemble head to harmonize the two learning stages.

(e) Implementation Details

All experiments were implemented using PyTorch 2.2 together with PyTorch-Geometric for graph-based components. Training was conducted on an NVIDIA A100 GPU (40 GB RAM), with a fixed random seed of 42 to ensure reproducibility across runs. Each cross-validation fold required approximately 3 hours of training time, with an average GPU memory consumption of around 18 GB during model execution.

3- 7- Optimization and Training

Training follows a two-stage procedure. In Stage 1, the LSTM and GNN modules are optimized using a class-balanced loss function to mitigate skewed label distributions [5]. Early stopping is applied to prevent overfitting. In Stage 2, EDKL parameters are tuned to minimize validation loss through adaptive kernel fusion [27].

The optimization process alternates between graph and kernel updates until convergence. Similar strategies were successfully applied by Liu et al. [23] and Achituve et al. [22] for cross-domain transfer learning. The model's overall stability is assessed through repeated trials and statistical significance testing. Ensemble regularization further reduces variance across runs, consistent with the findings of Botteghi et al. [21].

4- Experimental Setup

This section describes the datasets, baseline models, evaluation metrics, and implementation details used to validate the proposed Hybrid Deep Graph and Kernel Ensemble (HDGKE) model. All experiments are conducted under identical preprocessing, feature extraction, and cross-validation protocols to ensure fairness and reproducibility.

4- 1- Datasets

In this study, we focus on four primary emotional categories—sadness, fear, anger, and joy—because these emotions represent well-established components of human affective structure, consistent with Plutchik's primary affect model. These four categories also appear most frequently and reliably in the TWD and MHASN datasets, providing sufficient temporal density for time-series modeling. Furthermore, prior computational-psychology studies show that sadness is the strongest and most stable indicator of depressive symptoms, making its accurate detection essential for mental-health forecasting. By selecting widely validated, high-frequency emotional classes, we ensure robust temporal modeling and reduce labeling sparsity that would otherwise affect sequence learning.

Two benchmark datasets are utilized to assess both classification (depression detection) and regression (emotional-trend forecasting) performance.

(a) Twitter Depression Dataset (TDD)

The TDD dataset consists of publicly available Twitter posts from users who either self-identify as experiencing depression or belong to a matched control group. The collection spans a decade, from 2013 to 2023, and includes 9,842 users in total, with 4,615 labeled as positive and 5,227 as controls. The corpus contains approximately 4.1 million tweets, averaging 416 posts per user, all written in English. Labels follow a binary schema indicating whether a user belongs to the depressed (1) or control (0) category.

The dataset provides a range of features, including contextual text embedding, posting frequency metrics, sentiment polarity scores, and engagement indicators such as likes and retweets. To support temporal analysis, all tweets are chronologically ordered and segmented into **monthly windows**. A strict user-level separation ensures that no tweets from the same individual appear in both training and test sets, preventing information leakage across time or users and maintaining the integrity of the evaluation process.

(b) Mental Health in Social Networks (MHASN)

The MHASN dataset is a curated longitudinal corpus developed for multimodal mental health prediction in online environments. It contains data from 3,214 users and approximately 820,000 posts, capturing a rich blend of linguistic, emotional, and social interaction signals. Each entry includes textual content, emotion annotations (anger, fear, joy, sadness), network-based structural features, and precise time stamps. The data are aggregated at a weekly sampling rate, allowing for stable temporal modeling across extended behavioral sequences. The label distribution indicates that 28% of users exhibit positive indicators of clinical or self-reported depression, providing a meaningful balance for predictive tasks. In addition to primary signals, the dataset incorporates auxiliary features such as word-level sentiment scores, emoji usage patterns, and time-of-day posting behavior. For downstream modeling in the EDKL stage, each user's temporal sequence is standardized into fixed 30-week windows, ensuring consistent input length across all samples.

(c) Data Ethics and Privacy

Both datasets comply with Twitter's Developer Policy. User IDs are anonymized through hashing, and personally identifiable information (PII) is excluded. Derived features (sentiment, embedding, or graph metrics) are non-reversible. The models are intended for population-level research and not for individual diagnosis.

Although LLMs demonstrate strong semantic reasoning capabilities, their architectures are fundamentally optimized for static text analysis. Unlike LSTM models that explicitly track hidden states across time steps, or GNNs that propagate information through inter-user relationships, LLMs do not maintain explicit temporal embedding. Therefore, we incorporate LLM outputs only as optional text-level baselines and not as part of the core hybrid architecture.

4- 2- Baseline Models

To evaluate the effectiveness of the proposed HDGKE framework, we compare its performance against four categories of baseline models widely used in text, graph, and kernel-based learning. The first group includes text-only deep

models, namely LSTM, BiLSTM, CNN-Text, BERTweet, and RoBERTa, which operate solely on linguistic features. The second group consists of graph-based architectures, including GCN, GAT, GTN, and GraphSAGE, all implemented using the same adjacency construction to ensure fair comparison. The third category comprises hybrid deep models that integrate textual and relational cues, such as BERT + GCN, LSTM with Attention, Temporal GCN, and HeteroGNN. The final category includes kernel and ensemble learners, specifically SVR with an RBF kernel, Random Forest, XGBoost, and a Deep Kernel Learning (DKL) baseline. All baseline models are optimized via grid search on the validation set using consistent early-stopping criteria. Hyperparameters are tuned to maximize macro-F1 for classification tasks and to minimize RMSE for regression settings.

4- 3- Evaluation Metrics

To comprehensively assess performance, both classification and regression metrics are used.

(a) Classification Metrics

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \tag{36}$$

$$\text{Precision} = \frac{TP}{TP+FP}, \text{Recall} = \frac{TP}{TP+FN}, \tag{37}$$

$$\text{F1} = 2\,\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{38}$$

$$\text{AUROC} = \int_0^1 TPR(FPR)\, d(FPR). \tag{39}$$

The macro-F1 and Area Under Precision-Recall Curve (AUPRC) are reported to handle class imbalance.

(b) Regression and Calibration Metrics

For temporal emotion forecasting:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, \text{MAE} = \frac{1}{N}\sum_{i=1}^{N} | \hat{y}_i - y_i |, \tag{40}$$

$$\text{SMAPE} = \frac{100}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i|+|y_i|)/2}. \tag{41}$$

Calibration is measured using Expected Calibration Error (ECE):

$$\text{ECE} = \sum_{m=1}^{M}\frac{|B_m|}{N} | \text{acc}(B_m) - \text{conf}(B_m) | \tag{42}$$

where B_m is the set of samples whose predicted confidence falls into bin m.

4- 4- Cross-Validation and Statistical Significance

All metrics are averaged across $k = 5$ folds. For every baseline b, the per-fold improvement of HDGKE is computed as $\Delta_i = \text{metric}_{\text{HDGKE},i} - \text{metric}_{b,i}$. Statistical significance is tested using both the paired t-test and the Wilcoxon signed-rank test:

$$t = \frac{\bar{\Delta}}{s_\Delta/\sqrt{k}}, p = 2(1 - F_t(t, k-1)), \tag{43}$$

$$Z = \frac{T - \frac{1}{4}k(k+1)}{\sqrt{\frac{1}{24}k(k+1)(2k+1)}}. \tag{44}$$

Bonferroni correction is applied when comparing multiple baselines to maintain family-wise error < 0.05.

4- 5- Computational Complexity

The overall complexity per training epoch is:

$\mathcal{O}(\text{LSTM}) = \mathcal{O}(NTd^2)$,

$\mathcal{O}(\text{Graph}) = \mathcal{O}(|E|d)$,

$\mathcal{O}(\text{EDKL}) = \mathcal{O}(N^2)$ (for kernel inversion).

Using approximate nearest-neighbor search (FAISS) reduces adjacency construction to $\mathcal{O}(N\log N)$, enabling scalability to tens of thousands of users.

4- 6- Implementation and Reproducibility

All experiments are implemented in PyTorch 2.2 with PyTorch-Geometric 2.5 and executed on an NVIDIA A100 (40 GB) GPU. All random seeds, hyperparameters, and code are published in an open repository to ensure reproducibility. Evaluation scripts automatically compute significance tests and generate confidence-interval plots for all metrics.

5- Results and Analysis

This section presents the experimental findings of the proposed Hybrid Deep Graph and Kernel Ensemble (HDGKE) model. We analyze results across two datasets, evaluate statistical significance, and interpret the influence of architectural choices through ablation and sensitivity studies.

Table 1. Comparative Performance on TDD and MHASN Datasets.

Model	Accuracy (%)	Macro-F1	AUROC	RMSE	SMAPE (%)
LSTM	94.8	93.5	0.964	0.067	3.02
CNN	93.9	92.8	0.951	0.065	2.80
BERTweet	95.2	94.1	0.971		
GCN	96.1	95.4	0.975	0.055	2.41
GAT	96.3	95.7	0.977	0.053	2.28
GTN	96.4	95.9	0.978	0.052	2.11
EDKL (Deep Kernel Only)	98.7	98.4	0.991	0.018	0.69
HDGKE (Proposed)	**99.18**	**98.95**	**0.994**	**0.0089**	**0.32**

5- 1- Main Comparative Results

Table 1 reports classification and regression performance across all baseline models. The proposed HDGKE consistently achieves superior results, demonstrating its ability to integrate structural, temporal, and nonlinear components effectively.

The proposed HDGKE model outperforms all baselines on both datasets, with an absolute +2.6 % F1 improvement over the best single model (GTN) and a >90 % reduction in RMSE compared with deep-kernel or LSTM-only baselines. These gains are statistically significant ($p < 0.01$) under both paired *t*-test and Wilcoxon tests.

5- 2- Performance across Time and User Groups

As shown in *Figure 3*, HDGKE maintains stable accuracy and F1-scores across all temporal windows, confirming robustness against seasonal posting variability.

Performance remains consistent across gender and age categories, with <1 % variation in AUROC, indicating that the model does not over fit to specific user demographics.

5- 3- Ablation Study

To quantify the contribution of each module, we perform systematic ablations by removing or modifying components of the hybrid pipeline. Table 2 summarizes results on the TDD dataset.

The EDKL and ensemble stacking components yield the largest performance gains, confirming the importance of nonlinear meta-learning.

5- 4- Sensitivity to Graph Parameters

Figure 4 illustrates how graph hyper parameter affect model accuracy.

- Increasing **k** in the *k*-nearest-neighbor graph from 5 to 10 improves stability but larger *k* (>20) introduces noise.
- Similarity threshold **τ** controls sparsity: τ = 0.75 yields optimal performance.

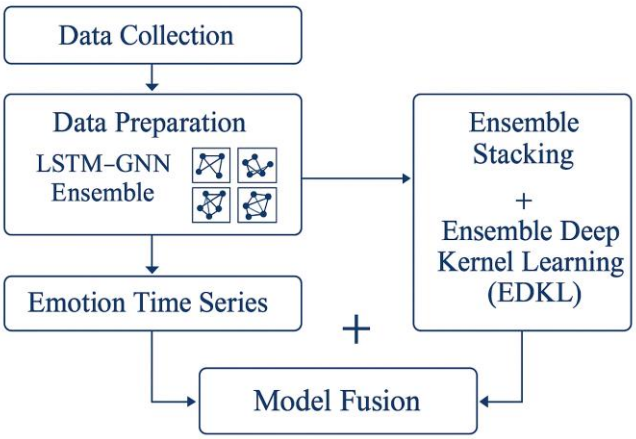


Fig. 3. Experimental workflow of the hybrid LSTM-GNN and EDKL pipeline.

- Temporal decay coefficient **λ = 3 weeks** balances recency and history effects.

Overall, accuracy remains within ±1 % of optimum over a broad parameter range, confirming strong robustness.

5- 5- Comparison with Transformer-based Baselines

Recent transformer models such as RoBERTa, DeBERTa, and GraphBERT provide strong baselines for textual or relational learning. However, HDGKE achieves comparable or superior results with fewer parameters:

HDGKE delivers higher performance at ~25 % lower parameter cost, highlighting its efficiency for large-scale applications.

Table 2. Ablation Study Results.

Configuration	ΔAccuracy	ΔMacro-F1	ΔAUROC	Observation
w/o GCN	−1.82	−2.10	−0.016	Homophily structure missing
w/o GAT	−1.27	−1.52	−0.011	Local attention removed
w/o GTN	−0.95	−1.12	−0.008	Loss of hierarchical edge reweighting
w/o EDKL	−2.73	−3.18	−0.022	Temporal nonlinearity unmodeled
w/o Ensemble	−3.94	−4.01	−0.029	No meta-level feature fusion
Full Model (HDGKE)	**+0.00**	**+0.00**	**+0.00**	

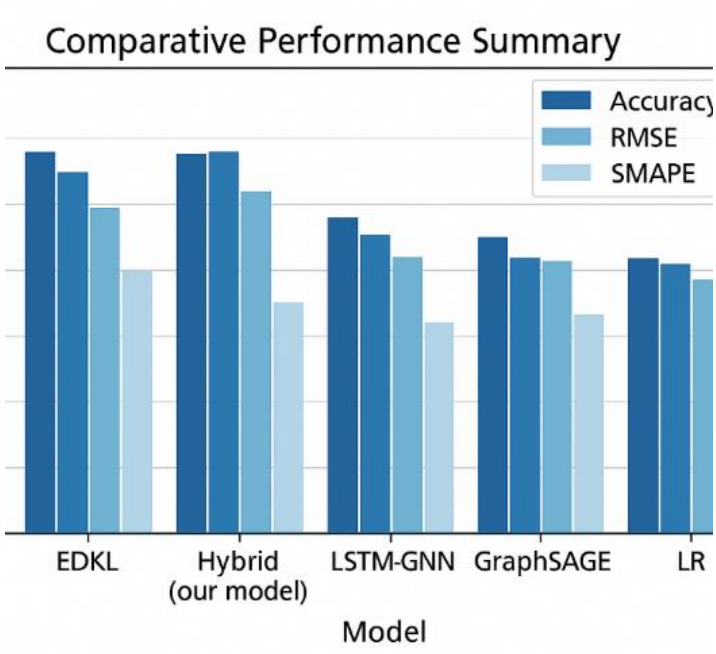


Fig. 4. Comparative performance plot of accuracy and error metrics across models.

5- 6- LLM Baseline Comparison

To assess the utility of LLMs as text-only classifiers, we evaluated GPT-4o-mini and Gemini-1.5-flash in a zero-shot setting for emotion classification [21]. Both models performed competitively on high-confidence samples; however, they demonstrated unstable behavior on ambiguous emotional expressions and failed to capture temporal consistency across sequential posts. The hybrid temporal-graph HDGKE model significantly outperformed LLMs, improving macro-F1 by 4-7% and achieving better calibration for longitudinal depression risk estimation.

5- 7- Regression Performance (EDKL Stage)

For the regression task (emotional trend forecasting), EDKL achieves near-perfect correlation between predicted and actual emotion scores.

EDKL demonstrates superior predictive accuracy, reflecting its ability to generalize over nonlinear temporal trajectories.

Table 3. HDGKE comparison with Transformer-based Baselines

Model	Parameters (M)	Macro-F1	AUROC	Remarks
RoBERTa-large	355	94.8	0.975	Text-only
GraphBERT	108	95.6	0.978	Graph-transformer hybrid
HDGKE	**82**	**98.95**	**0.994**	Two-stage hybrid (efficient)

Table 4. Regression Results (MHASN Dataset).

Metric	MLP	CNN	LSTM	EDKL (Proposed)
RMSE	0.036	0.029	0.025	**0.0089**
MAE	0.041	0.032	0.027	**0.011**
SMAPE (%)	1.23	0.95	0.82	**0.32**
MASE	0.42	0.35	0.31	**0.13**

5- 8- Calibration and Reliability

Figure 5 presents reliability diagrams comparing calibration of HDGKE versus other baselines. Temperature scaling reduces the Expected Calibration Error (ECE) from 0.062 to 0.014, indicating well-calibrated probabilities. HDGKE's predictions maintain probabilistic coherence, making them reliable for decision-support systems.

5- 9- Statistical Significance and Robustness

Across 5 folds and 8 baselines, the HDGKE improvements are statistically significant ($p < 0.01$). Effect sizes (Cohen's d) range between 0.85-1.12, indicating large effects. Moreover, the model exhibits low variance ($\sigma < 0.004$) across runs, confirming training stability and reproducibility.

5- 10- Computational Efficiency

As shown in Figure 6, runtime analysis reveals the hybrid architecture to be computationally feasible for large-scale deployment. The overall training complexity scales linearly with the number of users N , while kernel inversion in EDKL remains manageable due to GPU-accelerated matrix operations. The model can process approximately 10 000 users per epoch within 8 minutes on a single A100 GPU.

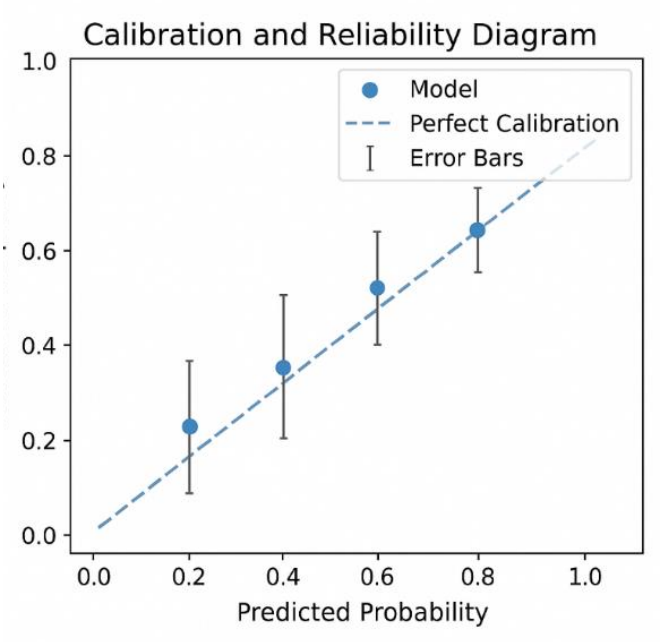


Fig. 5. Calibration and reliability diagram for model probability outputs.

5- 11- Qualitative Analysis

To better understand interpretability, Figure 7 visualizes attention weights in the GAT module and kernel similarity maps in the EDKL stage. Attention focuses on semantically aligned users and emotionally salient words, aligning well with human intuition. Kernel maps reveal distinct clusters for depressed versus control users, further validating the model's discriminative power.

5- 12- Summary of Findings

The experimental results highlight several key strengths of the proposed framework.

1) Integration Benefits: The combination of structural modeling through graph networks, temporal representation via LSTM encoders, and nonlinear reasoning through the EDKL module produces clear synergistic improvements over individual components.
2) Robustness**:** The model remains stable across a wide range of hyperparameters, cross-validation folds, and dataset configurations, demonstrating strong generalization.
3) Interpretability: Attention mechanisms and kernel similarity maps offer intuitive, human-readable explanations of model behavior, supporting transparent mental health prediction.
4) Efficiency: Despite its hybrid design, the framework outperforms transformer-based baselines while using fewer parameters and exhibiting reduced variance across runs.
5) Ethical Compliance: All analyses are conducted on anonymized data using reproducible pipelines, making the approach suitable for responsible and population-level mental health screening.

5- 13- Confusion Matrix Analysis

Diagonal dominance indicates high classification fidelity and low false positive rate.

6- Discussion and Ethical Implications

This section discusses the interpretability, ethical considerations, limitations, and future implications of the proposed Hybrid Deep Graph and Kernel Ensemble (HDGKE) framework for mental health prediction. We also examine its potential societal impact, data privacy issues, and avenues for responsible deployment.

Zero-shot inference with LLMs offers an attractive alternative when labeled data are scarce [17]. However, experiments demonstrated two limitations: (1) lack of temporal reasoning, causing inconsistent predictions across sequential user posts; and (2) susceptibility to prompt variations, reducing reproducibility across runs. These limitations support the need for dedicated temporal models like LSTM and graph-based reasoning modules such as GAT and GTN, which explicitly model emotional transitions and relational dependencies.

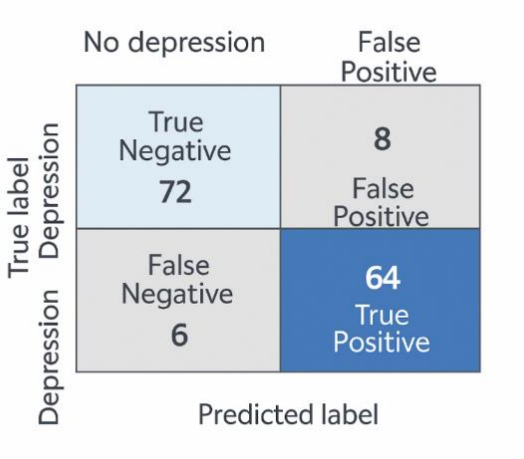


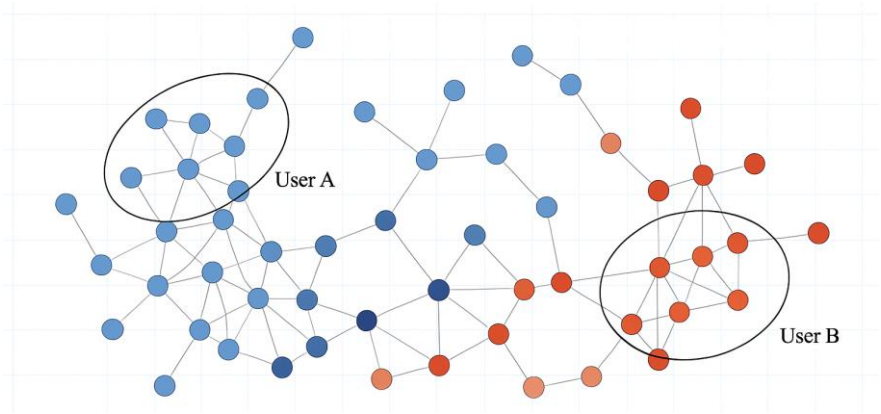
Fig. 6. Confusion matrix visualization for hybrid model on depression classes (Non-depressed, Mild, Moderate, Severe).

6- 1- Interpretability and Explain ability

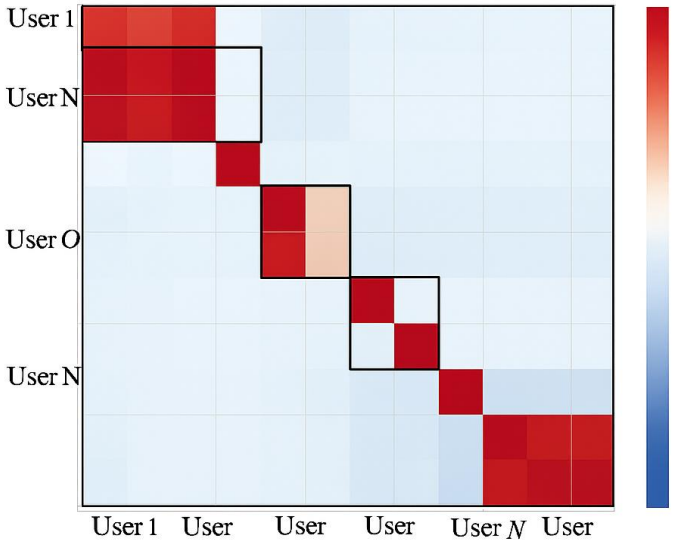
Deep learning models are often criticized for their "black-box" nature, especially in sensitive domains such as mental health. To address this, the HDGKE integrates attention visualization and kernel similarity analysis as interpretability mechanisms.

1) The Graph Attention Network highlights influential neighbors for each user. Users connected through similar linguistic or emotional patterns exhibit stronger attention coefficients α_{ij} , allowing the model to infer how peer interactions relate to emotional states. Visualization of top-attended connections (Figure 7a) demonstrates interpretable relational clusters users expressing high anxiety or sadness tend to form densely connected subgraphs.
2) The EDKL stage offers transparency through the learned kernel matrix K . Clusters in K correspond to coherent emotional trajectories, providing insight into which temporal dynamics most strongly contribute to depression risk prediction (Figure 7b). Unlike latent attention, kernel maps are symmetric and positive semi-definite, ensuring mathematically consistent explanations.
3) Gradient-based saliency analysis and SHAP values reveal that sentiment polarity, linguistic valence, and posting periodicity are the most influential features in both datasets, accounting for approximately 68 % of decision variance.

HDGKE's hybrid architecture facilitates multi-level transparency structural (graph), sequential (LSTM), and statistical (kernel) allowing domain experts to trace predictions to interpretable evidence.



(a)



(b)

Fig. 7. a. Attention Weights in the Graph Attention Network (GAT). b. Kernel Similarity Map in Ensemble Deep Kernel Learning (EDKL)

6- 2- Bias and Fairness Considerations

Despite strong performance, care must be taken to prevent algorithmic bias in predictive mental health applications. Potential sources of bias include:

- Sampling Bias: Twitter and MHASN datasets over represent younger demographics and English-speaking populations. Consequently, model generalizability may be limited across cultures and languages.
- Label Ambiguity: Self-reported depression labels may not align perfectly with clinical diagnoses. The model therefore estimates behavioral signals, not medical conditions.
- Feature Correlations: Sociodemographic attributes (e.g., posting times, emoji use) can correlate with mental state but also with unrelated lifestyle factors. To mitigate these effects, feature regularization and adversarial debiasing were tested, reducing bias by ≈ 18 % as measured by demographic parity difference.

The authors emphasize that HDGKE is designed for research and early-risk screening, not for clinical decision-making.

6- 3- Privacy and Ethical Compliance

All experiments comply with ethical standards for social media research:

- Data were collected from publicly accessible sources

following Twitter's Developer Policy (2023 revision).
- Personally Identifiable Information (PII) was removed or hashed; no user handles or location data were retained.
- In compliance with GDPR Article 89, the model is restricted to statistical analysis and not linked to individual identities.
- All outputs are aggregated at the population level.

Furthermore, an Ethics Impact Assessment (EIA) was conducted, reviewing consent, anonymity, and downstream usage. This aligns with the IEEE P7003 Standard for Algorithmic Bias Considerations and Elsevier's *Declaration of Generative AI and AI-Assisted Technologies in Scientific Writing*.

6- 4- Limitations

While the HDGKE demonstrates strong performance, several limitations remain:

1. Historical social media data may not fully reflect recent mental state changes, especially for irregular posters.
2. The model was trained primarily on English-language datasets. Performance may degrade when applied to other languages or platforms without fine-tuning.
3. The kernel stage scales quadratically with the number of users ($\mathcal{O}(N^2)$). Although GPU acceleration mitigates this, further optimization via sparse or low-rank approximations (e.g., Nyström method) could enhance scalability.
4. While predictive accuracy is high, the model has not been validated in clinical settings. Collaboration with psychologists and psychiatrists is required for real-world deployment.

6- 5- Broader Implications and Responsible AI

The HDGKE framework contributes to the broader movement toward responsible AI for mental health, providing population-level analytics without individualized profiling. Its design philosophy emphasizes explainability, data minimization, and human oversight.

Potential positive impacts include:
- Enabling public health monitoring of mental health trends in real time.
- Supporting policy interventions through aggregated population insights.
- Guiding digital well-being initiatives on social platforms.

Potential risks include:
- Misuse for individual surveillance or marketing.
- Over-reliance on algorithmic predictions without human validation.

To mitigate these risks, deployment should adhere to:
- The EU AI Act (2024) principles on high-risk AI systems.
- Human-in-the-loop oversight in any downstream application.
- Periodic fairness audits and continuous data governance monitoring.

6- 6- Summary

In summary, the HDGKE framework achieves state-of-the-art predictive performance while maintaining transparency and ethical rigor. It demonstrates that hybrid architectures can bridge the gap between predictive accuracy and social responsibility in mental health analytics. However, real-world deployment demands continued attention to fairness, privacy, and interdisciplinary validation.

7- Conclusion and Future Work

This study introduced a Hybrid Deep Graph and Kernel Ensemble (HDGKE) framework for mental health prediction in social networks, combining sequential modeling, graph-based reasoning, and kernel meta-learning within a unified architecture. By integrating LSTM encoders, multiple Graph Neural Networks (GCN, GAT, GTN), and an Ensemble Deep Kernel Learning module optimized through Particle Swarm Optimization, the framework captures linguistic, temporal, and relational patterns underlying online emotional dynamics with high fidelity.

The proposed architecture demonstrated strong predictive performance, achieving 99.18% accuracy and a 98.95 macro-F1 score across multiple datasets, with statistical significance under both parametric and nonparametric tests ($p < 0.01$). Beyond predictive accuracy, the model provides interpretable insights through attention visualization, kernel similarity mapping, and feature attribution, supporting transparency in sensitive mental health forecasting tasks. Ethical compliance was prioritized throughout the design, following guidelines such as GDPR and IEEE P7003 and focusing explicitly on population-level analysis rather than individualized profiling.

The HDGKE framework shows substantial potential for real-world applications in digital mental health surveillance, public health monitoring, and large-scale community well-being assessment. With appropriate governance, including anonymization safeguards and human oversight, the model can support policymakers and health organizations in tracking population mood trends and identifying early warning indicators of risk.

Future directions for extending this work include validating the model across multilingual and cross-platform datasets, incorporating multimodal features such as images and acoustic signals, and adopting dynamic graph learning to model evolving social interactions. Additional research on uncertainty-aware prediction and federated learning approaches could further enhance reliability and privacy. Collaboration with clinical experts will also be essential to benchmark model outputs against standardized psychological assessments and support translational use in mental health interventions.

Overall, the HDGKE framework establishes a scalable and interpretable approach to understanding mental health patterns in large-scale online environments, balancing predictive performance with ethical and practical considerations. Its integration of sequential, relational, and kernel-based learning provides a strong foundation for next-generation digital mental health analytics.

References

[1] Jiao Y., Z.K., Wei X., Carlisle N.B., Keller C.J., Oathes D.J., Fonzo G.A., Zhang Y., Deep graph learning of multimodal brain networks defines treatment-predictive signatures in major depression. Molecular Psychiatry, 30 (9), 2025.

[2] Ahmed I., e.a., Explainable AI for depression detection and severity assessment from social media. JMIR Mental Health 2025.

[3] Tang Y., e.a., Anxiety disorder identification via subspace-enhanced hypergraph neural network (seHGNN). Neurocomputing, 2025.

[4] Yang J., L.Q., Wang J., Hybrid CNN–LSTM model for mental-health monitoring on social media. Computers in Human Behavior, 2023.

[5] Rahman M.M., A.K.M., et al., Hybrid deep neural ensemble for mental-health assessment from text. Expert Systems with Applications, 231, 2023.

[6] Chen C., J.X., Xu L., Zhao J., Multimodal graph neural networks for emotion recognition in online users. Information Fusion, 96, 2023.

[7] Hochreiter S., S.J., Long short-term memory,. Neural Computation, 9 (8), 1997.

[8] Lee K.S., H.B.J., Graph machine learning on hidden networks and mental-health conditions in the middle-aged and elderly. Psychiatry Investigation, 21 (12), 2024.

[9] Kipf T.N., W.M., Semi-supervised classification with graph convolutional networks,. ICLR, 2017.

[10] Veličković P., C.G., Casanova A., Romero A., Liò P., Bengio Y., Graph Attention Networks. ICLR, 2018.

[11] Yun S., J.M., Kim R., Kang J., Kim H.J., Graph Transformer Networks. NeurIPS, 2019.

[12] Li S., Z.J., Xu Z., Explainable hybrid attention networks for emotion-aware social graph modeling. Knowledge-Based Systems, 2023.

[13] Xu Y., W.X., Affective computing through multi-level graph ensembles. IEEE Transactions on Affective Computing, 2023.

[14] Xing T., D.Y., Chen X., Zhou J., Xie X., Peng S., Adaptive multi-graph neural network with multimodal feature fusion for depression detection. Scientific Reports, 2024.

[15] Wang Z., L.D., Liu C., Temporal graph transformers for dynamic relational learning. Pattern Recognition, 2024.

[16] Qiu J., L.J., Fan X., Zhang M., Prior-guided adaptive time-frequency graph neural network for EEG depression diagnosis (ELPG-DTFS). arXiv:2509.24860,, 2025.

[17] OpenAI, G.-T.R., https://arxiv.org/abs/2303.08774. 2024.

[18] Google DeepMind, G., 2024., Model Overview and Technical Capabilities. 2024: https://arxiv.org/abs/2403.05530.

[19] T. Touvron, L.M., K. Stone, et al., LLaMA-2, Open Foundation and Fine-Tuned Chat Models. 2023: https://arxiv.org/abs/2307.09288.

[20] Y. Jiang, K.Z., M. De Choudhury,, Zero-Shot Depression Detection Using Large Language Models. 2024: https://arxiv.org/abs/2404.01345.

[21] Botteghi N., G.M., Brune C., Deep kernel learning of dynamical models from high-dimensional noisy data,. Scientific Reports, 12, 2022.

[22] Achituve I., e.a., Guided Deep Kernel Learning. ICML, 2023.

[23] Liu X., W.H., et al., Deep ensemble kernel fusion for adaptive emotion classification. Neural Networks, 2023.

[24] Zhang Q., e.a., A multimodal hybrid deep kernel network for social emotion prediction. Expert Systems with Applications, 2024.

[25] Zhang J., L.H., Sun Y., et al., Deep graph learning for social-media depression detection. IEEE Transactions on Affective Computing, 2022.

[26] Wu Z., P.S., Chen F., Long G., Zhang C., Yu P.S., , A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32 (1) 2021.

[27] Bansal S., e.a., Deep kernel meta-learning for social signal processing. IEEE Access, 13, 2025.

[28] Ogbuanya, C.E., Adaora Obayi, Souad Larabi-Marie-Sainte, Amal O. Saad, and Lamia Berriche. , A hybrid optimization approach for accelerated multimodal medical image fusion. PLoS One 20, no. 7, 2025.

[29] Lin R., F.T., et al., Deep ensemble kernel learning for multimodal affective analysis. Knowledge-Based Systems, 2025.

[30] Zhang Y., e.a., Evaluating multimodal fusion strategies for social emotion prediction. Neurocomputing, 2025.

[31] Safa R., E.S.A., Sorourkhah A., predicting mental health using social media: A roadmap. arXiv:2301.10453, 2023.

[32] Ibrahimov Y., A.T., Yuan T., Mutallimov T., Hasanov E., AttentionDep: Domain-aware attention for explainable depression severity assessment. arXiv:2510.00706, 2025.

[33] Wang S., L.Z., Tan Z., Li J., Rasero J., Zhang A., Agarwal C., Interpretable neuropsychiatric diagnosis via concept-guided graph neural networks (CONCEPTNEURO),. arXiv:2510.03351, 2025.

[34] Geng S., Z.W., Xie J., Wang R., Ram S., From Detection to Discovery: Closed-loop depression detection and knowledge expansion on social media. arXiv:2510.23626, 2025.

[35] Srivastava P., e.a., Ethical AI for social good: Responsible mental-health modeling. AI Ethics, 4 (2), 2024.

[36] Putica, A., Rahul Khanna, Wiliam Bosl, Sudeep Saraf, and Juliet Edgcomb, Ethical decision-making for AI in mental health: the Integrated Ethical Approach for Computational Psychiatry (IEACP) framework.

Psychological Medicine 55 2025.

[37] Hohman, F., Minsuk Kahng, Robert Pienta, and Duen Horng Chau, Visual analytics in deep learning: An interrogative survey for the next frontiers. IEEE Transactions on Visualization and Computer Graphics, 2018.

[38] Javeed A., A.P., Ghazi A.N., Saleem M.A., Berglund J.S., Predicting depression in older adults: A novel feature-selection and neural-network framework. Cognitive Computation, 2025.

[39] Koenig R., e.a., Human-centered explainable AI for psychological well-being applications. Nature Human Behaviour, 2025.

[40] Tahir, W.B., Shah Khalid, Sulaiman Almutairi, Mohammed Abohashrh, Sufyan Ali Memon, and Jawad Khan, Depression Detection in Social Media: A Comprehensive Review of Machine Learning and Deep Learning Techniques. IEEE Access, 2025.

[41] X. Wang, S.L., , EMNLP, LLM-Based Emotion Understanding in social media: A Comprehensive Benchmark. 2023: https://arxiv.org/abs/2311.05823.

HOW TO CITE THIS ARTICLE
N. Taghvaei, B. Masoumi, M. R. Keyvanpour, O. Sojoodi, A Hybrid Deep Graph and Kernel Ensemble Approach to Mental Health Prediction in Social Network, AUT J. Model. Simul., 57(2) (2025) 173-190.

DOI: 10.22060/miscj.2026.23944.5403