Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 33 No. 1s (2026)

An Attention-Driven Graph Neural Network Approach for Mental Health Sentiment Detection

1st Lavanya G

Assistant Professor

Artificial Intelligence and Data Science

Vignan Institute of Technology and Science

Hyderabad, Telangana State, India

lavanyayadav89@gmail.com

2nd M. Sai Chandana

Student

Artificial Intelligence and Data Science

Vignan Institute of Technology and Science

Hyderabad, Telangana State, India

saichandanamaddipatla@gmail.com

3rd S. Hemanth Kumar

Student

Artificial Intelligence and Data Science

Vignan Institute of Technology and Science

Hyderabad, Telangana State, India

hemanthsiripurapu02@gmail.com

4th G. Sagar

Student

Artificial Intelligence and Data Science

Vignan Institute of Technology and Science

Hyderabad, Telangana State, India

sagarguguloth0816@gmail.com

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 33 No. 1s (2026)

Article History:
Received: 04-02-2026
Revised: 20-03-2026
Accepted: 10-04-2026

Abstract:
Mental health issues such as depression, anxiety, stress, suicidal ideation, bipolar disorder, and personality disorders have become increasingly prevalent in the modern digital era. With the rapid growth of online platforms, individuals often express their emotional and psychological states through text-based content such as reviews, posts, and comments. Traditional artificial intelligence–based sentiment analysis systems primarily treat each text independently and fail to capture the complex semantic and relational dependencies that exist among mental health expressions. This limitation reduces the effectiveness of detecting high-risk mental health conditions, especially in overlapping and imbalanced categories. To address these challenges, this paper proposes a graph-based deep learning framework that integrates TF-IDF based text feature extraction with a Graph Attention Network (GAT) for mental health sentiment classification. TF-IDF is employed to transform textual data into numerical feature vectors representing the importance of words within the dataset, while a similarity-based graph structure is constructed using a K-Nearest Neighbor (KNN) approach to model relationships among text samples. The Graph Attention Network then applies attention mechanisms to effectively learn the relational dependencies between text instances and emphasize important mental health indicators. The proposed system classifies mental health data into seven categories: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder. Experimental results demonstrate that the proposed TF-IDF + GAT model achieves improved classification performance in identifying various mental health conditions and effectively captures relationships among textual data compared to traditional standalone machine learning approaches.

Keywords: Mental Health Analysis, Sentiment Analysis, TF-IDF, Graph Neural Networks, Graph Attention Network, Natural Language Processing, Deep Learning.

1. INTRODUCTION

Mental health disorders such as depression, anxiety, stress, suicidal ideation, bipolar disorder, and personality disorders have become a significant global concern, affecting individuals' emotional well-being, relationships, and overall quality of life. With the rapid growth of digital communication platforms, people increasingly express their emotions and psychological states through online text such as social media posts, forums, and reviews. These textual expressions provide valuable insights for identifying mental health conditions and enable opportunities for early detection and intervention. However, traditional sentiment analysis methods often rely on simple machine learning or lexicon-based approaches that

analyze text independently and fail to capture complex contextual relationships within mental health-related expressions, leading to limited accuracy in detecting high-risk conditions.

To address these limitations, recent advancements in artificial intelligence and natural language processing have introduced graph-based deep learning approaches capable of understanding contextual relationships between text samples. This study proposes a graph-based framework that integrates TF-IDF feature extraction with a Graph Attention Network (GAT) to analyze mental health-related textual data. A K-Nearest Neighbor (KNN) similarity graph is constructed to represent relationships between text instances, allowing the GAT model to focus on important connections using attention mechanisms. The proposed system classifies mental health data into seven categories—Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder—and demonstrates improved classification performance compared to traditional machine learning methods, providing an effective and scalable solution for automated mental health sentiment analysis and early mental health monitoring.

II. REVIEW OF RELATED WORK

Recent advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), and Deep Learning (DL) have significantly enhanced the analysis of textual data for detecting mental health conditions such as depression, anxiety, stress, suicidal ideation, bipolar disorder, and personality disorders. With the growing use of social media platforms, individuals increasingly express their emotions and psychological states through online posts, blogs, and discussion forums. Traditional sentiment analysis methods based on simple machine learning or rule-based approaches often analyze text independently and fail to capture complex contextual meanings and semantic relationships. To overcome these limitations, modern research focuses on deep learning architectures, graph-based neural networks, and hybrid NLP models that can better understand contextual dependencies and relational patterns within textual data.

Several recent studies demonstrate the effectiveness of these advanced techniques in mental health detection. Researchers have applied deep learning models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM) along with graph-based approaches to improve classification accuracy. For instance, enhanced graph convolution networks and attention-based models have been proposed to capture syntactic and semantic relationships in text, while federated learning frameworks address challenges related to distributed and limited datasets. Additionally, multimodal frameworks combining text, audio, and visual data have further improved sentiment analysis performance. These advancements highlight the potential of AI-driven systems to provide accurate mental health detection, enabling early monitoring and intervention through intelligent analysis of online textual data.

A. LITERATURE REVIEW

X. Yuan (2024) proposed a **Deep Belief Network (DBN)–**based approach for sentiment analysis and mental health prediction using social media data. The study analyzed the emotional polarity of user-generated content to identify potential psychological conditions.

Compared with traditional sentiment analysis methods, the DBN model demonstrated improved capability in capturing complex emotional patterns within textual data. Experimental results indicated that the proposed system achieved a prediction accuracy of highlighting the effectiveness of deep belief networks in analyzing social media text for mental health prediction.

Recent research trends emphasize the integration of **graph-based neural networks, attention mechanisms, and deep learning architectures** to improve mental health sentiment analysis. These approaches focus on capturing contextual dependencies, relational structures, and semantic patterns within textual datasets. Despite significant progress, challenges remain in addressing issues such as dataset imbalance, noisy social media data, and model interpretability. Building upon these advancements, modern mental health detection systems aim to develop more robust and scalable frameworks capable of analyzing large-scale textual data and providing reliable support for early mental health monitoring and intervention systems.

B. PROPOSED METHODOLOGY

The proposed **mental health sentiment analysis framework** follows a structured multi-stage deep learning pipeline designed to accurately classify textual expressions related to different psychological conditions. The system focuses on identifying seven mental health categories: **Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder**. The methodology integrates **Natural Language Processing (NLP) techniques, TF-IDF feature extraction, graph construction strategies, and a Graph Attention Network (GAT)** to effectively capture both semantic features and relational dependencies among textual data. The workflow consists of the following stages:

Dataset Collection and Organization

The system utilizes textual datasets collected from online platforms where individuals commonly express their emotions and psychological states, such as social media posts, reviews, and discussion forums.

• Dataset consists of mental health–related textual content
• Target classes:
o Normal
o Depression
o Suicidal
o Anxiety
o Stress
o Bipolar Disorder
o Personality Disorder

Data cleaning procedures are applied to remove noisy, incomplete, and duplicate text entries to ensure dataset quality and reliability. The cleaned dataset is then divided into **training, validation, and testing subsets** to support structured model development and unbiased performance evaluation.

 AI, NLP, and Deep Learning technologies have improved the detection of mental health conditions from textual data.

 Mental health issues such as depression, anxiety, stress, suicidal ideation, bipolar disorder, and personality disorders are often reflected in social media posts, blogs, and online discussions.
 Traditional sentiment analysis methods using simple machine learning or rule-based techniques struggle to capture complex contextual and semantic relationships in text.
 Modern approaches use deep learning architectures and graph-based neural networks to better understand contextual dependencies in mental health data.
 Models such as RNN, LSTM, and Bi-LSTM help extract meaningful patterns from textual datasets and improve classification accuracy.
 Attention mechanisms and graph convolution networks enhance the understanding of syntactic and semantic relationships in text.
 Federated learning frameworks help analyze distributed and limited datasets while maintaining data privacy.
 Multimodal sentiment analysis combining text, audio, and visual data further improves mental health detection and supports early monitoring and intervention.

The proposed **graph-based deep learning framework** improves contextual understanding of mental health expressions, enhances classification reliability, and provides a scalable solution for automated mental health sentiment analysis.

C. Methodology

The proposed methodology for mental health sentiment classification follows a structured graph-based deep learning framework designed to detect psychological conditions from textual data. The process begins with dataset collection, where mental health–related text such as reviews, comments, and social media posts is gathered from public sources. These texts represent different psychological states categorized into seven classes: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder. The collected dataset then undergoes preprocessing steps including text normalization, tokenization, stop-word removal, punctuation filtering, and lemmatization to improve data quality. After preprocessing, the dataset is divided into training, validation, and testing sets to ensure effective model training and unbiased evaluation.

Next, TF-IDF (Term Frequency–Inverse Document Frequency) is applied to transform the textual data into numerical feature vectors by measuring the importance of words within documents. To capture relationships between similar text samples, a K-Nearest Neighbour (KNN) similarity graph is constructed, where each text instance is represented as a node and edges connect similar texts based on their TF-IDF features. This graph structure enables the system to analyze contextual dependencies and semantic relationships between different mental health expressions within the dataset.

Finally, the classification is performed using a Graph Attention Network (GAT), which applies attention mechanisms to assign importance weights to neighbouring nodes in the graph. This allows the model to focus on the most relevant contextual relationships during learning. The model is trained using optimization algorithms and evaluated using metrics such as accuracy, precision, recall, and F1-score. By integrating TF-IDF feature extraction, graph-based representation, and attention-based deep learning, the proposed framework provides an

effective, scalable, and automated solution for mental health sentiment analysis and early psychological risk detection.

SYSTEM REQUIREMENTS

Hardware Requirements

- **Processor:** Intel Core i5 / AMD Ryzen 5 or higher
- **RAM:** Minimum 8 GB (16 GB recommended for better performance)
- **Storage:** Minimum 250 GB HDD or 128 GBSSD
- **GPU (Optional):** NVIDIA GPU with CUDA support for faster model training
- **Display:** 1366 × 768 resolution or higher
- **Network:** Stable Internet connection for GenAI integration and Flask server access
- **Sensors (Optional – IoT Integration):** pH Sensor, TDS Sensor, Turbidity Sensor, EC Sensor

Software Requirements

- **Operating System:** Windows 10/11 or Ubuntu 22.04 LTS
- **Programming Language:** Python 3.10 or higher
- **Framework:** Flask for web UI and backend integration
- **Machine Learning Libraries:** scikit-learn, XGBoost, LightGBM
- **Data Handling Libraries:** Pandas, NumPy
- **Visualization Tools:** Matplotlib, Seaborn, Plotly
- **Generative AI Tools:** OpenAI API / Hugging Face Transformers / LangChain
- **Prompt Engineering Toolkit:** LangChain or LlamaIndex
- **Database (Optional):** SQLite or MySQL for storing prediction history
- **Version Control:** Git / GitHub
- **IDE / Editor:** Visual Studio Code or PyCharm
- **Deployment Environment:** Localhost or Cloud platforms (AWS, GCP, Azure, Render)

III. SYSTEM ARCHITECTURE

System Architecture Overview

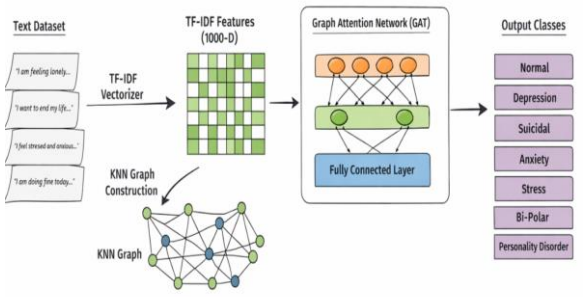

Fig-1: System Architecture

The system architecture for the **Mental Health Sentiment Analysis System** is designed to provide accurate and scalable classification of psychological conditions using advanced natural language processing and graph-based deep learning techniques. The architecture follows a modular structure that integrates text preprocessing, feature extraction, graph construction, deep learning model training, performance evaluation, and deployment components.

The system consists of **five primary interconnected modules**, each responsible for a specific stage in the mental health sentiment analysis pipeline.

 Data Acquisition Module: Collects mental health–related textual data from social media, blogs, forums, and reviews, labeled into seven categories such as Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder.
 Data Preprocessing & Feature Extraction: Performs text cleaning, normalization, tokenization, stop-word removal, and converts text into numerical vectors using TF-IDF for model input.
 Graph Construction & Deep Learning Model: Builds a K-Nearest Neighbour (KNN) similarity graph and applies a Graph Attention Network (GAT) to capture contextual relationships and classify mental health sentiments.
 Performance Evaluation: Measures model effectiveness using metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.
 Web Interface & Deployment: Provides a user-friendly interface where users input text, and the system processes it through the model to display the predicted mental health category with a confidence score

Workflow

 Dataset Loading: The workflow starts by loading the Combined Data.csv dataset containing mental health–related text statements and their corresponding class labels.
 Data Preprocessing & Balancing: Unnecessary columns are removed, the dataset is cleaned, and class imbalance is addressed by selecting 50,000 samples from data.
 TF-IDF Feature Extraction: Text data is converted into numerical vectors using TF-IDF vectorization, extracting important features from each text instance.
 Graph Construction using KNN: A K-Nearest Neighbour (KNN) similarity graph is created where each text sample is a node connected to its five nearest Neighbors based on TF-IDF similarity.
 Graph Data Preparation & Model Training: The graph is converted into a PyTorch Geometric Data object, split into 80% training and 20% testing, and a Graph Attention Network (GAT) model is trained.
 Prediction & Classification: The trained GAT model predicts the mental health category for new text inputs, classifying them into Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, or Personality Disorder.

Dataset Description Table

Attribute	Description
Dataset Name	Combined Mental Health Dataset
File Format	CSV
File Name	Combined Data.csv
Data Type	Text Dataset
Input Feature	Statement (text describing emotions or feelings)
Target Variable	Label
Number of Classes	5-7
Classes	Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, Personality Disorder
Feature Extraction	TF-IDF Vectorization
Graph Construction	K-Nearest Neighbor (KNN)
Neighbors Used	5
Data Split	80% Training, 20% Testing
Deep Learning Model	Graph Attention Network (GAT)
Framework	PyTorch + PyTorch Geometric
Evaluation Method	Classification Accuracy

Table-1: Dataset Description

DATA FLOW AND INTEGRATION

1. Input

Textual data representing user posts, reviews, and comments are collected from online platforms. These text samples contain expressions related to mental health conditions and are categorized into seven classes.

2. Processing

The input text undergoes preprocessing and TF-IDF feature extraction to convert it into numerical vectors. A **K-Nearest Neighbor graph** is constructed to represent relationships among text samples. The **Graph Attention Network (GAT)** model then analyzes these relationships and learns important contextual dependencies for classification.
The system computes:
• Predicted Mental Health Category
• Class Probability Scores
• Model Performance Metrics

3. Output

The system provides the following outputs:
• Predicted Mental Health Condition (Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, Personality Disorder)
• Classification Confidence Score
• Model Evaluation Results

IV. TESTING AND EVALUATION

Testing Methodology

The testing methodology for the **Mental Health Sentiment Classification System using TF-IDF and Graph Attention Network (GAT)** follows a structured and systematic evaluation approach to ensure classification accuracy, robustness, generalization capability, and reliability in identifying different mental health conditions from textual data.

V.PERFORMANCE EVALUATION

The performance evaluation focuses on several critical aspects to assess the system's effectiveness in **automated mental health sentiment detection and classification.**

The system's performance is evaluated based on its classification accuracy and detection reliability across seven mental health categories: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder. By integrating TF-IDF feature extraction with a Graph Attention Network (GAT), the framework improves predictive performance compared to traditional machine learning models. TF-IDF captures the importance of words within text, while the graph-based structure models relationships between similar text samples. The attention mechanism in GAT emphasizes important emotional signals and contextual relationships, allowing the model to effectively handle complex and overlapping mental health expressions, thereby improving classification accuracy.

Additionally, the framework is computationally efficient because TF-IDF reduces processing complexity compared to large transformer models, enabling scalable analysis of large social media datasets. Overall, the hybrid TF-IDF + GAT architecture outperforms traditional machine learning approaches by effectively capturing relational and contextual patterns in textual data, making it suitable for automated mental health monitoring systems.

Fig-2: Performance

VI. CONCLUSION

The proposed AI-driven Mental Health Sentiment Classification System integrates advanced Natural Language Processing (NLP) and graph-based deep learning techniques to accurately detect mental health conditions from textual data. The framework uses TF-IDF feature extraction and a Graph Attention Network (GAT) to analyze semantic patterns and contextual relationships within text, classifying data into seven categories: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder. Text data is preprocessed and converted into numerical vectors, while a K-Nearest Neighbor (KNN) similarity graph models relationships between text samples, enabling the GAT model to focus on important contextual connections. Experimental results show that the TF-IDF + GAT architecture outperforms traditional machine learning methods with an accuracy of 80%, providing a scalable and intelligent solution for automated mental health sentiment analysis and supporting early detection and monitoring of psychological conditions.

Future Scope

1. **Real-Time Social Media Monitoring:** Integrating the model with social media platforms to analyze user-generated content in real time for early detection of mental health risks.
2. **Mobile Mental Health Support Applications:** Developing mobile applications that utilize the proposed model to provide early mental health screening and emotional well-being monitoring for users.
3. **Explainable AI Integration:** Incorporating explainable AI techniques to highlight important words or text segments that influence mental health predictions, improving model interpretability for clinicians.
4. **Larger Multi-Platform Dataset Expansion:** Training the model on larger datasets collected from multiple social media platforms to improve generalization and robustness.
5. **Multimodal Mental Health Analysis:** Integrating additional data sources such as voice signals, facial expressions, and behavioral patterns with textual data to enhance prediction accuracy.

REFERENCES

[1] X. Yuan, "Social Media Sentiment Analysis and Mental Health Prediction using Deep Belief Network (DBN)," *2024 Second International Conference on Networks,*

Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-5, doi: 10.1109/NMITCON62075.2024.10699157.

[2] R. Hu, J. Yi, L. Chen and Z. Jin, "Graph Reconstruction Attention Fusion Network for Multimodal Sentiment Analysis," in *IEEE Transactions on Industrial Informatics*, vol. 21, no. 1, pp. 297-306, Jan. 2025, doi: 10.1109/TII.2024.3452204.

[3] U. Ahmed, J. C. -W. Lin and G. Srivastava, "Hyper-Graph Attention Based Federated Learning Methods for Use in Mental Health Detection," in *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 768-777, Feb. 2023, doi: 10.1109/JBHI.2022.3172269

[4] J. A, S. K. D, S. Karunakaran, N. D, R. S. Dass and J. Seetha, "Next-Gen Mental Health Detection through Deep Learning and NLP in Social Media," *2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC),* Davangere, India, 2024, pp. 1-6, doi: 10.1109/ICICEC62498.2024.10808285.

[5] J. Chen and R. Yan, "EMGCN: Enhancement Graph and Multi-head Attention Graph Convolutional Networks for Aspect-based Sentiment Analysis," *2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD),* Tianjin, China, 2024, pp. 1591-1596, doi: 10.1109/CSCWD61410.2024.10580036.