TPM Vol. 32, No. S9, 2025
ISSN: 1972-6325
https://www.tpmap.org/

FQ-GNN: FIREFLY-OPTIMIZED QUANTUM GRAPH NEURAL NETWORKS FOR CONTEXTUAL DEPRESSION DETECTION FROM SOCIAL MEDIA TEXTS

MS. R. GEETHA
RESEARCH SCHOLAR, DEPARTMENT OF COMPUTER SCIENCE, RATHINAM COLLEGE OF ARTS AND SCIENCE, COIMBATORE, EMAIL: gthram92@gmail.com

DR. D. VIMAL KUMAR M.C.A., M.PHIL., PH.D.
ASSOCIATE PROFESSOR & HEAD, DEPARTMENT OF COMPUTER SCIENCE, RATHINAM COLLEGE OF ARTS AND SCIENCE, COIMBATORE, EMAIL: drvimalcs@gmail.com, vimalkumar.cs@rathinam.in

Abstract
FQ-GNN is aphase of research work that presents a novel framework related to depression detection from social media textswith the use of a quantum graph neural network improvedthrough a bio-inspired Firefly algorithm. This proposed approach constructs a heterogeneous graph where social media users, posts, and interactions are modeled as edges and nodes weighted bytemporal decay and semantic similarity. Node-level embeddings are derived through a transformer-based semantic encoder integrated with temporal context filtering, catching both interaction chronology and textual content. Then, these embeddings are managed through quantum-aware graph convolutional layers to extract rich semantic and structural features. To professionally explore the high-dimensional parameter space and maximizethe multi-objective performance metrics, a firefly-driven quantum hyper-parameter optimizer is employed. Lastly, adaptive fusion and neighbor-aware refinement combine structural and semantic information to make accurate depression calculations. Extensive evaluations establish that FQ-GNN significantly outperforms conventional baselines, providing contextually-aware detection of depressive behavior in online social interactions and robustness.
Keywords: Quantum Graph Neural Networks, Firefly Optimization, Depression Detection, Social Media Analysis, Semantic-Contextual Embeddings, Bio-Inspired Hyper-Parameter Optimization

I. INTRODUCTION

One of the most widespread health challenges is Depression, which affectsover a million people across the world and also leads to significant psychological, economic, and social consequences [1].It reduces productivity, damages cognitive functions, disturbs interpersonal relationships, and if it is untreated over a long period of time, it can alsolead to severe mental health crises [1], [5]. Conventional diagnostic procedures, like psychometric assessments, clinical interviews, and self-reported questionnaires, often suffer from delays, subjectivity, and low scalability, making it tough to reach larger populations efficiently [5], [7], [8].

Platforms like Twitter, Facebook, and Reddit are some of the social platforms that are growing rapidly, and these platforms provide a unique opportunity to monitor mental health indicators in real-time. These platforms host a vast amount of user-generated content, where often people express their emotions, thoughts, and personal experiences [2], [3]. Studyingthese textual pattern, it offers the potential for firstexposure of depressive tendencies andpermitstaking action to advance a medical disorder a little earlier, before its symptoms get worse. Social media analysis suggests scalable solutions that tie traditional clinical methods and offer insights into population-level mental health trends [6], [4].

Regardless of these opportunities, the inherent characteristics of the data cause anautomaticdetection of depression from social media text. The posts are often short, informal,noisy, and contextually ambiguous, which reduces traditional machine learning performance of the model[6], [8]. Also, features of the texts are sparse, high-dimensional, and non-linear, which increases the risk of overfitting and makes thesimplification across datasets a difficult task [12], [15].

This highlights the need for robust, scalable, and well-organized frameworks combining with semantic understanding, optimization of the feature, and advanced sorting techniques for exact depression prediction to avoid the above-stated problems.

Thoughprevious studies have advanced social media-based depression detection,still there are a fewsignificant challenges,like

- Data noise and sparsity: Difficult to capture meaningful patterns from short posts that include slang, emojis, or misspellings [6], [8].

- Feature selection and hyperparameter optimization: Settings of most of themodelsare fixed, and it useshandmade features. Thisleads to suboptimal performance and poor generalization [12].
- HighComputational complexity: Deep learning models like LSTMs, CNNs, or hybrid architectures needimportant computational resources. This restrictsthe real-time application [13], [11], [17].
- Limited semantic understanding: Conventional vectorization methods fail to capture nuanced contextual and emotional information that is present in the social media text [13], [17], [19].
- Robustness under noisy conditions: Noise, not posting constantly, and platform-specific biases adversely affect prediction accuracy [6], [12], [21].

Addressing these problems, a novel, unified approach that combines semantic–contextual embeddings,quantum graph neural networks,and bio-inspired optimization algorithmsto achieverobust, correct, and scalable detection of depression from texts of social media.

Research Objectives

The objectives of the proposed FQ-GNN framework are:

- To develop a Firefly-based Quantum Optimization (FQO) framework for hyperparameter tuning and feature selection, and to enhance both the accuracy and robustness under noisy conditions [21].
- Develop a Quantum Graph Neural Network (Q-GNN) to extract high-order semantic and structural features from social media text that are represented as a semantic–temporal graph [13], [17], [19].
- Assimilate semantic embeddings, such as BERT and RoBERTa, to encode both contextual and syntactic information, which allows the model to understand nuanced linguistic and emotional cues [13], [17], [19].
- To perform a complete performance evaluation using depression prediction accuracy, precision, recall, f1 score, specificity and depression prediction time in social media datasets [9], [11], [15].

The aim of this proposed work is to fill the gaps in current approaches, ensuring the robustness, scalability, and efficient prediction of depression while addressing challenges inherent to social media data. The following section presents the related works.

II. RELATED WORKS

A. Early Social Media Approaches

Social media has been documented as a valuable source of information for monitoring mental health.Initialresearch demonstrated that the linguistic patterns and posting behavior could be leveraged for the detection of depression. De Choudhury et al. [2] displayed that the temporal changes in language and social activity onTwitter,which is a social media platform, indicate depressive tendencies. Jalonen [3] emphasized social media as a critical outlet for giving expression to negative emotions, providing a rich dataset for sentiment and affect analysis. Guntuku et al. [4] studied numerous DL and ML approaches, highlighting the importance of both feature engineering and temporal modeling for effective depression detection. Evans-Lacko et al. [5] debated socio-economic disparities affecting contact to mental health services, further motivating scalable and automated solutions.

B. LearningMethods based on Machine

Traditional machine learning approaches applied algorithms such as Random Forests,Support Vector Machines (SVMs), and clustering techniques to social media data. Cacheda et al. [6] confirmed early depression detection using social network analysis integrated with Random Forest classifiers, highlighting the significance of relational features. Adek et al. [7] and Ahmed et al. [8] used clustering techniques for short-text data, improving the feature representation and dimensionality reduction.

C. Deep Learning Approaches

The acceptance of deep learning architectures significantly improved predictive capabilities. A hybrid CNN-BiLSTM model for depression detection from the tweets was proposed by Kour and Gupta [9]. This effectively captures both sequential and spatial patterns. An embedded LSTM architecture was developed by Singh et al. [10], whileDEPTWEET was introduced by Kabir et al. [11], which classifies depression severity from text data. Ren et al. [12] incorporated multi-criteria decision-making with discrete Z-numbers to improve early diagnosis.

D. Semantic Embeddings

Transformer-based embeddings, whichinclude BERT and RoBERTa, have proven effective in capturing contextual and semantic nuances. Kurniadi and Paramita [13] enhanced the detection of depression in short texts ofsocial media by using these embeddings. Kanahuati-Ceballos et al. [15] leveraged embeddings to optimize LSTM, RNN, and Forest's Random models.MOGAM was introduced by Cha et al. [16], which is a multimodal objectoriented graph attention model, and Bendebane et al. [19] fine-tuned BERT for multi-labeled Twitter datasets, signifying the importance of capturing textual data's deep semantic relationships.

E. Graph Neural andOptimization Networks

Current research highlights integrating bio-inspired algorithms for optimization with graph neural networks to improveperformance and robustness. The proposed FQ-GNN framework applies Firefly-based Quantum Optimization (FQO) to tune hyperparameters and select discriminative features [21]. Quantum Graph Neural Network (Q-GNN) captures high-order structural and semantic patterns in social media text [13], [17], [19]. The combination of both addresses the noise, high-dimensionality, and low-SNR conditions, resulting in scalable, robust, and efficient depression detection.

While earlier studies focused on semantic embeddings or ML/DL models individually, in the upcoming section FQ-GNN framework introduces a unified approach that combines FQO optimization, Q-GNN feature learning, and semantic embeddings, ensuring practical applicability for large-scale social media depression detection. This methodimproves accuracy, offers robustness against noise, and offers scalability for real-time deployment across different social media platforms.

III. PROPOSED WORK

This is the continuation of the earlier research work and can be seen in [22]. Early detection of depressive tendencies from social-media posts requires a representation that can preserve (i) the semantic nature of text, (ii) the structural relations among users, posts, and topics, and (iii) the temporal evolution of interactions. Instead of treating each post as an isolated sample, the proposed framework begins by constructing a heterogeneous, semantically weighted, temporally aware graph.
This graph later serves as the computational substrate for the Firefly-Enhanced Adaptive Quantum Graph Neural Network (FFA-QGNN) classifier.
Definition of Heterogeneous Social-Media Graph
We denote the processed dataset as

$$\mathcal{D} = \{(v_x, q_y, u_{yz}, w_{yz})\} \quad \ldots(1)$$

where $v_x$denotes the x-th user,
$q_y$denotes the y-th social-media post,
u_{yz}denotes the time stamp of the k-th appearance of post $\mathrm{q_y}$, and
$w_{yz} \in \{0,1\}$denotes the ground-truth depression label associated with that post.
A heterogeneous graph is then defined as

$$\mathrm{H} = (\mathrm{W}, \mathrm{F}, \mathrm{Y}) \quad \ldots(2)$$

where Wdenotes the complete set of nodes (including users, posts, hashtags, and latent topics), $\mathrm{F} \subseteq \mathrm{W} \times \mathrm{W}$ denotes the set of typed edges that represent interactions or semantic affinities among nodes, and $\mathrm{Y} \in \mathbb{R}^{|\mathrm{W}| \times \mathrm{d_0}}$denotes the matrix of $\mathrm{d_0}$- dimensional initial feature vectors attached to all nodes.
where $|\ \mathrm{V}\ |$is the total number of nodes and $\mathrm{d_0}$is the raw feature dimensionality.
Multi-Relational Adjacency Integration
Since the network contains multiple relation types—such as user–post, post–topic, user–user (friendship or follow) and post–post (reply/retweet)—a single adjacency cannot fully describe the graph. Therefore, we integrate all relation-specific adjacencies into a unified one as

$$B = \frac{\sum_{\mathrm{x} \in \mathcal{X}} \omega_{\mathrm{x}} B^{(\mathrm{x})}}{\sum_{\mathrm{x} \in \mathcal{X}} \omega_{\mathrm{x}}} \quad \ldots(3)$$

where $\mathcal{X}$denotes the set of relation types,
$\mathrm{A}^{(\mathrm{x})} \in \{0,1\}^{|\mathrm{W}| \times |\mathrm{W}|}$denotes the binary adjacency matrix corresponding to relation type x, and
$\omega_{\mathrm{x}} > 0$denotes a learnable scalar weight that reflects the contribution of relation rto depression-related information flow.
where the denominator normalizes the weighted sum to keep Ain a comparable numeric range.
Semantic Similarity of Text-Bearing Nodes
The raw connectivity may not capture the semantic affinity among posts or users.
To address this, each text-bearing node (posts, user profiles, topics) is encoded first by a transformer-based language model (e.g., DeBERTa or RoBERTa), producing an embedding vector $\mathrm{e_i} \in \mathbb{R}^{\mathrm{d_e}}$for node i. We then define the normalized semantic similarity between nodes iand jas

$$T_{\mathrm{ij}} = \frac{\langle f_{\mathrm{i}}, f_{\mathrm{j}} \rangle}{\| f_i \|_2 \ \ \| f_{\mathrm{j}} \|_2 + \varepsilon} \quad \ldots(4)$$

where $\langle \cdot, \cdot \rangle$denotes the inner product,
$\|\cdot\|_2$denotes the Euclidean norm,
and $\varepsilon > 0$is a small constant to prevent division by zero.
where $T_{\mathrm{ij}} \in [0,1]$provides a soft semantic linkage even between otherwise unconnected nodes.
Temporal Decay Weight for Recency Sensitivity
Since depressive tendencies often manifest in recent linguistic and behavioral patterns, edges corresponding to older interactions must contribute less to the model.
We introduce a continuous temporal decay as

$$\phi_{xy} = \exp\left[-\lambda\,(V_{\mathrm{cur}} - V_{xy})\right] \quad \ldots(5)$$

where $\mathrm{V_{cur}}$denotes the current reference time,
V_{xy}denotes the time stamp of the interaction between nodes x and y,
and $\lambda > 0$is a tunable decay-rate constant controlling how fast the influence of older links diminishes.
where $\phi_{xy} \in (0,1]$naturally emphasizes recent interactions.

Unified Weighted Adjacency

By combining the purely structural relation from (3), the semantic affinity from (4), and the temporal factor from (5), the final adjacency used in the learning stage is defined as

$$\widetilde{\mathrm{B}}_{xy} = \mathrm{B}_{xy} \cdot T_{xy} \cdot \phi_{xy} \quad \ldots(6)$$

where $\mathrm{B}_{xy}$is the element of the integrated adjacency from (3),

$T_{xy}$is the semantic similarity from (4), and

$\phi_{xy}$is the temporal decay weight from (5).

where $\widetilde{\mathrm{B}}_{xy}$fuses heterogeneous information into a single graph operator suitable for subsequent GNN processing.

Feature Normalization

To avoid training instability caused by heterogeneity of raw features, each node feature vector is normalized as

$$\hat{y}_x = \frac{y_x - \mu_g}{\sigma_g + \delta} \quad \ldots(7)$$

where $y_x$is the raw feature vector of node i,

$\sigma_g$and $\mu_g$are respectively standardand mean deviation of each feature dimension computed over all nodes, and $\delta > 0$is a small constant for numerical stability.

where $\hat{y}_x$becomes the standardized input feature vector ready for embedding refinement.

Semantic–Contextual Embedding Layer

While Stage 1 builds a heterogeneous graph $\mathrm{H} = (\mathrm{W}, \mathrm{F}, \mathrm{Y})$endowed with the unified weighted adjacency $\tilde{B}$and normalized initial features $\hat{y}_x$, the raw features $\hat{y}_x$mostly capture surface-level statistics such as word counts or profile meta-data.

Accurate depression prediction requires deep semantic understanding of textual content together with contextualization by user–post–topic relations.

Stage 2 introduces a Semantic–Contextual Embedding (SCE) layer, which converts raw node features into task-aware latent embeddings that integrate linguistic, behavioral, and structural cues.

Transformer-Based Textual Encoder

For each text-bearing node (post, user biography, topic tag) we first compute a context-sensitive token-level representation using a pre-trained transformer (e.g., RoBERTa or DeBERTa).

Given a sequence of $\mathrm{L_i}$tokens for node i, the transformer produces hidden states $\mathrm{h_i^{(\ell)}}$at each layer ℓ.

We define the initial semantic embedding for node ias

$$x_j^{(0)} = \frac{1}{I_j} \sum_{\mathrm{k}=1}^{I_j} \mathrm{h}_{j,\mathrm{k}}^{(I_{\mathrm{tr}})} \quad \ldots(8)$$

where $\mathrm{h}_{j,\mathrm{k}}^{(I_{\mathrm{tr}})}$denotes the hidden vector of the k-th token at the top transformer layer I_{tr},

and $x_j^{(0)} \in \mathbb{R}^x$is the averaged embedding vector for node j.

where the averaging ensures a fixed-dimensional representation irrespective of token length.

Sentiment–Polarity Augmentation

Because emotional valence is a crucial signal in depression detection, we augment the transformer embedding with a normalized sentiment-polarity scalar $a_{\mathrm{x}} \in [-1,1]$obtained from a specialized affective lexicon or sentiment model.

The sentiment-augmented embedding is defined as

$$x_j^{(1)} = [x_j^{(0)}; \alpha_a a_{\mathrm{x}}] \quad \ldots(9)$$

where $[\,\cdot\,;\,\cdot\,]$denotes vector concatenation

and $\alpha_a > 0$is a scaling factor balancing the influence of sentiment with that of the semantic vector.

where the augmentation enriches the representation with explicit emotional tone.

Contextual Neighborhood Aggregation

Merely averaging neighbor features may dilute discriminative cues.

To capture context-dependent importance of neighboring nodes, we adopt an attention-weighted neighborhood aggregation.

For each node i, its context-aware embedding is updated as

$$x_j^{(2)} = \sigma\left(A_p\, x_j^{(1)} + \sum_{\mathrm{k} \in \mathcal{N}(j)} \alpha_{jk} A_q\, x_k^{(1)} \right) \quad \ldots(10)$$

where $A_p$and $A_q$are trainable weight matrices for self- and neighbor-contributions respectively,

$\sigma(\cdot)$is a non-linear activation function (e.g., ReLU),

$\mathcal{N}(\mathrm{j})$denotes the neighborhood of node iin the graph,

and $\alpha_{jk}$is the attention coefficient measuring the relevance of neighbor k to node j.

where the second term adaptively emphasizes semantically or temporally related neighbors.

Computation of Attention Coefficients

The attention coefficient $\alpha_{jk}$is computed using a semantic–structural compatibility function:

TPM Vol. 32, No. S9, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

$$\alpha_{xy} = \frac{\exp\left(\psi^{\top}\tanh\left(X_b x_j^{(1)} \parallel X_b z_k^{(1)} \parallel \tilde{B}_{jk}\right)\right)}{\sum_{l\in\mathcal{N}(j)} \exp\left(\psi^{\top}\tanh\left(X_b z_i^{(1)} \parallel X_b z_l^{(1)} \parallel \tilde{B}_{jl}\right)\right)} \quad \ldots(11)$$

where $X_b$is a trainable projection matrix,
$\psi$is a trainable attention vector,
$\tanh(\cdot)$introduces non-linearity,
$\parallel$denotes vector concatenation,
and $\tilde{B}_{jk}$is the unified weighted adjacency entry from Stage 1 (6).
where the denominator normalizes the coefficients so that $\sum_{k\in\mathcal{N}(k)} \alpha_{jk} = 1$.

Dropout-Regularized Embedding

To mitigate over-fitting in the high-dimensional feature space, a dropout mask is applied:

$$x_j^{(3)} = \mathrm{Dropout}_{p_d}\left(x_j^{(1)}\right) \quad \ldots(12)$$

where $\mathrm{Dropout}_{p_d}(\cdot)$randomly zeros each element of the input vector with probability $p_d \in (0,1)$during training.
where $x_j^{(3)}$serves as the robust semantic–contextual embedding used as an input to the succeeding QGNN stage.

Quantum-Enhanced Graph Neural Model

Stages 1 and 2 provided, respectively, a heterogeneous weighted graph Gand the semantic–contextual node embeddings $x_j^{(3)}$.

Stage 3 introduces the Quantum-Enhanced Graph Neural Network (Q-GNN) layer that leverages graph propagation enriched with quantum-inspired linear-unitary transformations.

This layer extracts higher-order relational dependencies—crucial for revealing latent depressive patterns that emerge from the interaction among posts, users, and topical contexts.

Quantum State Initialization

In order to model uncertainty and superposed neighborhood influence, each node iis conceptually treated as a quantum state whose initial amplitude vector is derived from the Stage 2 embedding:

$$\psi_i^{(0)} = \frac{x_j^{(3)}}{\parallel x_j^{(3)} \parallel_2 + \epsilon} \quad \ldots(13)$$

where $x_j^{(3)}$is the dropout-regularized semantic–contextual embedding from (12),
$\|\cdot\|_2$is the Euclidean norm,
and $\epsilon > 0$is a small stabilizer.
where the normalization constrains the state vector to lie approximately on the unit hypersphere, making it suitable for quantum-inspired evolution.

Quantum Propagation Rule

Let Hdenote a learnable Hermitian-like propagation operator defined on the graph.
The one-step propagation of node states is given by a unitary-inspired rule:

$$\psi_j^{(\ell+1)} = \sigma\left(V^{(\ell)}\psi_j^{(\ell)} + \sum_{k\in\mathcal{N}(j)} \beta_{jk}^{(\ell)} W^{(\ell)} \psi_k^{(\ell)}\right) \quad \ldots(14)$$

where $V^{(\ell)}$and $W^{(\ell)}$are learnable linear operators for self- and neighbor-influence at layer ℓ,
$\beta_{ij}^{(\ell)}$is the normalized quantum-inspired neighborhood weight,
$\mathcal{N}(j)$is the set of neighbors of node i,
and $\sigma(\cdot)$is a nonlinear activation (e.g., ELU).
Where the formulation mimics local quantum diffusion but remains fully differentiable for end-to-end learning.

Quantum-Inspired Neighborhood Weight

The neighborhood weight is defined using a phase-modulated relevance score:

$$\beta_{jk}^{(\ell)} = \frac{\exp(\Re\{\theta_{jk}^{(\ell)}\})}{\sum_{l\in\mathcal{N}(j)} \exp\left(\Re\{\theta_{jl}^{(\ell)}\}\right)} \quad \ldots(15)$$

where $\Re\{\cdot\}$denotes the real part,
and $\theta_{jk}^{(\ell)}$is the complex-valued compatibility defined as

$$\theta_{jk}^{(\ell)} = (\psi_j^{(\ell)})^{H} X_r^{(\ell)} \psi_j^{(\ell)} + \eta\, \tilde{A}_{jk} \quad \ldots(16)$$

where $(\cdot)^{H}$denotes the Hermitian transpose,
$X_r^{(\ell)}$is a complex-valued bilinear interaction matrix at layer ℓ,
$\eta > 0$is a scaling parameter blending structural adjacency,
and $\tilde{A}_{jk}$is the unified weighted adjacency from Stage 1 (6).
Where (15) and (16) jointly determine the phase-aware attention that governs neighbor influence in the propagation rule (14).

Layer-wise Quantum Feature Update

The output feature representation after L_q quantum-GNN layers is computed as

$$h_j = \psi_j^{(M_r)} \quad \ldots(17)$$

where M_r is the number of propagation layers and

h_j becomes the final graph-aware quantum feature vector of node j.

Where these high-level features now encode both semantic–contextual and relational-structural signals.

Residual Fusion with Original Semantic Embedding

To mitigate over-smoothing and preserve core semantic signals from Stage 2, a residual fusion is performed:

$$\tilde{h}_j = \gamma\, h_j + (1-\gamma)\, y_k^{(3)} \quad \ldots(18)$$

where $\gamma \in [0,1]$ is a fusion hyper-parameter controlling the contribution of the Q-GNN output h_j relative to the original embedding $y_k^{(3)}$.

Where this fusion ensures that discriminative lexical information is not lost while gaining graph-level relational context.

Bio-Inspired Optimization with Enhanced Deep Classifier

While the Q-GNN in Stage 3 yields rich semantic-relational embeddings $\tilde{h}_i$, the choice of classifier architecture, learning-rate schedule, regularization coefficients, and embedding fusion weights strongly affects the final detection accuracy and robustness under sparse, noisy data.

Stage 4 therefore introduces a Firefly-Driven Quantum-Aware Optimization (FQO) framework to co-optimize the deep classification head and its hyperparameters.

The approach hybridizes the global exploration strength of Firefly Algorithm (FA) with a stability-guided local exploitation rule tailored for high-dimensional neural architectures.

Classifier Architecture

The classifier $f_\Theta(\cdot)$ is a two-branch deep head:

- a feed-forward branch to capture dense non-linear decision boundaries, and
- a temporal branch based on a light-weight Bi-GRU to exploit posting-sequence regularities.

Given a node embedding $\tilde{h}_j$ from (18), the classifier output for node is

$$\hat{x}_j = f_\Theta(\tilde{h}_j) = \mathrm{Softmax}(A_p \phi(\tilde{h}_j) + c_d) \quad \ldots(19)$$

where $\phi(\cdot)$ denotes the joint feature extracted by the two-branch head,

A_p and c_d are the final classification weights and bias,

and $\hat{x}_j \in \mathbb{R}^D$ is the estimated probability vector over Ctarget classes (e.g., depressive vs. non-depressive).

Where the hybrid architecture is more expressive than purely feed-forward or purely recurrent alternatives, yet remains tractable for large-scale social data.

Joint Optimization's Objective Function

The overall optimization must balance classification fidelity, model compactness, and robustness. Hence, the multi-objective loss is formulated as

$$\mathcal{J}(\Theta, \lambda) = \mathcal{L}_{CE} + \lambda_1 \parallel \Theta \parallel_2^2 + \lambda_2\, \mathcal{L}_{Qdiff} + \lambda_3\, \mathcal{L}_{SNR} \quad \ldots(20)$$

where

- $\mathcal{L}_{CE}$ is the cross-entropy loss over training samples,
- $\parallel \Theta \parallel_2^2$ is the weight-decay term promoting compactness,
- $\mathcal{L}_{Qdiff}$ is a quantum-state discrepancy penalty encouraging consistency between Stage 3 states $\psi_i^{(L_q)}$ and fused features $\tilde{h}_j$,
- $\mathcal{L}_{SNR}$ is a robustness loss that penalizes performance degradation under synthetic low-SNR perturbations, and
- $\lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ are the regularization weights to be tuned.

where the choice of λ and key architectural hyperparameters (e.g., hidden sizes, fusion factor γ in (18), Bi-GRU depth) defines a high-dimensional search space.

Firefly-Driven Quantum-Aware Optimization (FQO)

Firefly Encoding

Each candidate firefly y_j represents a vector of:

$$y_j = [\, \gamma, d_{BiGRU}, h_f, \alpha_{lr}, \lambda_1, \lambda_2, \lambda_3 \,] \quad \ldots(21)$$

where

γ is the residual-fusion weight in (18),

d_{BiGRU} is the number of recurrent layers,

h_f is the hidden-layer width of the feed-forward branch,

α_{lr} is the initial learning rate,

and the λ terms are as in (20).

where this unified encoding enables the FA to simultaneously adjust both structural and optimization parameters.

Attractiveness Function

The attractiveness β_{ij} of a firefly j with respect to firefly k is defined as

$$\beta_{ij} = \beta_0 \exp(-\xi \parallel x_j - x_i \parallel_2^2) \quad \ldots(22)$$

where $\beta_0 > 0$is the base attractiveness and $\xi > 0$is the light-absorption coefficient.
where a lower Euclidean distance in hyper-parameter space indicates greater similarity and thus stronger attraction.
Quantum-Aware Random Walk
To increase exploration while preserving stability in late iterations, a quantum-aware random perturbation is injected:

$$x_k^{i+1} = x_k^i + \beta_{kj}(x_j^i - x_k^i) + \zeta\, Q(0, \sigma_p^2) \quad \ldots(23)$$

where
$Q(0, \sigma_p^2)$denotes a zero-mean random vector drawn from a Levy-like quantum distribution, $\zeta > 0$is a scaling factor, and the last term diversifies the search to avoid premature convergence.
where the quantum-inspired perturbation enhances the ability to escape local optima when tuning deep-model hyperparameters.
Fitness Evaluation
The fitness of each candidate is evaluated by

$$F(p_i) = \omega_1\, Acc_{val} - \omega_2\, CompCost - \omega_3\, Var_{SNR} \quad \ldots(24)$$

where $Acc_{val}$is the validation accuracy of the classifier trained with candidate p_i,CompCost is the normalized computational cost,Var_{SNR}measures prediction variance under noisy conditions, and $\omega_{1,2,3}$are preference weights.
Where the maximization of $F(p_i)$leads to a balanced trade-off between predictive power, efficiency, and robustness.
Convergence Criterion
The optimization proceeds until either the maximum iteration $T_{max}$is reached or the relative improvement of the best fitness falls below a tolerance $\epsilon_{conv}$for r consecutive iterations:

$$\left| \frac{F_{best}^{i} - F_{best}^{i-r}}{F_{best}^{i-r}} \right| < \epsilon_{conv} \quad \ldots(25)$$

where this dual criterion avoids unnecessary computation once satisfactory hyper-parameters are obtained.
Adaptive Feature Fusion and Decision Refinement
While the FQO-tuned classifier in Stage 4 optimizes the hyper-parameters of the decision head, further performance gains and interpretability can be achieved by explicitly modelling the interaction between semantic features $s_i$and structural–temporal features $g_i$of each user i.
Stage 5 introduces an Adaptive Feature Fusion (AFF) layer followed by a Decision Refinement (DR) block, which together provide a more balanced and robust representation for final classification.
Split of Feature Modalities
Let the output of the Stage 3 Q-GNN for node ibe decomposed into two complementary parts:

$$\tilde{h}_j = [\, s_j \,\|\, g_j \,] \quad \ldots(26)$$

where
$s_j \in \mathbb{R}^{d_s}$denotes the purely semantic embedding aggregated from text and topic similarity, $g_j \in \mathbb{R}^{d_g}$denotes the structural–temporal embedding induced by graph relations and posting dynamics, and $[\cdot\|\cdot]$is the concatenation operator.
Attention-Weighted Modality Balancing
An adaptive attention gate computes per-sample modality weights:

$$\alpha_j = \sigma(X_b[s_j \,\|\, g_j] + c_b) \quad \ldots(27)$$

where
$X_b$and $c_b$are learnable parameters of the gate,
$\sigma(\cdot)$is the element-wise sigmoid,
and $\alpha_j \in [0,1]^{d_s+d_g}$modulates the relative importance of semantic vs. structural components for each dimension.
Fused Representation
The adaptively fused feature vector is then expressed as

$$u_j = \alpha_j \odot s_j + (1 - \alpha_j) \odot g_j \quad \ldots(28)$$

where
$\odot$denotes the element-wise (Hadamard) product.
where the fusion dynamically balances textual cues (e.g., sentiment, depressive expressions) with structural cues (e.g., interaction pattern, posting intervals), mitigating over-reliance on either modality.
Residual Calibration
To retain information possibly suppressed by the fusion, a residual calibration term is added:

$$r_j = \delta\,(\mathrm{LayerNorm}([\, s_j \,\|\, g_j]) - u_j) \quad \ldots(29)$$

where
$\delta \in (0,1)$is a learnable scalar controlling the residual strength,
and $\mathrm{LayerNorm}(\cdot)$is the standard layer-normalization operator.
The final enriched representation for decision making is

$$v_j = u_j + r_j \quad \ldots(30)$$

Decision Refinement Block

The vector v_j is passed to a light refinement network consisting of a gated linear unit (GLU) and a smoothing filter:

$$p_j = \mathrm{GLU}(X_p v_j + b_p) \quad \ldots(31)$$

$$z_j = (1-\rho)\, p_J + \rho\, \mathrm{Avg}_{K \in \mathcal{N}(j)} p_j \quad \ldots(32)$$

where
X_p, b_p are parameters of the GLU,
$\rho \in [0,1]$ is a smoothing coefficient,
and $\mathcal{N}(j)$ denotes the set of graph-neighbours of node j.
Where the GLU selectively activates informative components, while the neighbour-aware averaging propagates local consensus to reduce label noise.
Final Prediction
The refined feature z_j is finally classified by

$$\hat{y}_j = \mathrm{Softmax}(X_o z_j + c_o) \quad \ldots(33)$$

where X_o, c_o are output weights and bias.
Integrated Workflow and Convergence Analysis
The previous stages describe five cooperating algorithmic modules.
his now joins them into a single end-to-end workflow and analyses the convergence properties of FQO-QGNN–AFF system.
End-to-End Workflow
1. Graph Construction:
Social-media records are converted into a heterogeneous graph $H = (W, F, Y)$, with edges weighted by semantic similarity and temporal decay.
2. Semantic–Contextual Embedding:
A transformer encoder and temporal context filter yield node-level representations $h_j^{(0)}$ that incorporate raw text and interaction chronology.
3. Quantum GNN Feature Learning:
The embeddings pass through multiple quantum-aware graph convolutional layers to produce $\tilde{h}_j$ as in (18).
4. Bio-Inspired Hyper-Parameter Optimization:
A firefly-driven quantum-aware optimizer searches the high-dimensional space of model hyper-parameters to maximize the multi-objective fitness (24).
5. Adaptive Fusion and Decision Refinement (Stage 5):
Semantic s_i and structural g_i components are balanced via (27)–(30); the fused representation is refined with neighbour-aware smoothing (31)–(32), and the final classifier (33) predicts depression status.
The complete pipeline therefore realises

$$\hat{y}_j = \mathcal{F}_\Phi(H, Y) \quad \ldots(34)$$

where $\mathcal{F}_\Phi(\cdot)$ denotes the integrated mapping defined by all learnable parameters Φ of Stages 2–5.
where this formulation makes the entire framework an explicit compositional operator on the input graph and its features.
Unified Objective Function
All trainable parts of the pipeline jointly minimise

$$\mathcal{J}_{\text{total}} = \mathcal{J}(\Theta, \lambda) + \mu_1 \mathcal{L}_{\text{fuse}} + \mu_2 \mathcal{L}_{\text{smooth}} \quad \ldots(35)$$

where
$\mathcal{J}(\Theta, \lambda)$ is the Stage 4 objective (20),
$\mathcal{L}_{\text{fuse}} = \| u_j - [s_j \| g_j] \|_2^2$ encourages consistent fusion,
$\mathcal{L}_{\text{smooth}} = \sum_{(j,k)\in E} \| z_j - z_k \|_2^2$ encourages local decision smoothness,
and $\mu_1, \mu_2 > 0$ are trade-off coefficients.
where this compound loss ties together the contributions of all stages under a single learning principle.
Convergence of Inner Training Loop
For fixed hyper-parameters supplied by the FQO outer loop, the network parameters Θ are trained by stochastic gradient descent with an adaptive step size η_t satisfying

$$\sum_{j=1}^{\infty} \eta_j = \infty, \sum_{j=1}^{\infty} \eta_j^2 < \infty \quad \ldots(36)$$

Under standard smoothness and bounded-variance assumptions, the expected gradient norm satisfies

$$\lim_{j\to\infty} \mathbb{E}[\| \nabla_\Theta \mathcal{J}_{\text{total}}^{(j)} \|] = 0 \quad \ldots(37)$$

indicating convergence to a stationary point of the objective's total .
where this guarantees stability of the dynamic'straining within each outer iteration of the bio-inspired search.
Convergence of Outer Bio-Inspired Loop

Let $F^{(i)}$denote the best fitness value at the i^{th}iteration of the Firefly optimizer. Using the attractiveness (22) and quantum perturbation (23), and assuming the fitness landscape is bounded above by F_{max}, it follows from the standard FA dynamics that

$$F_{max} - F^{(i+1)} \leq (1 - \beta_{min})[F_{max} - F^{(i)}] \quad \ldots(38)$$

where $\beta_{min} \in (0,1)$is the minimum non-zero attractiveness among all firefly pairs.
Iterating (38) yields

$$F_{max} - F^{(i)} \leq (1 - \beta_{min})^{i}[F_{max} - F^{(0)}] \quad \ldots(39)$$

so that the best fitness improves geometrically toward the optimum as $t \to \infty$.
where this establishes the global convergence of the bio-inspired search under mild assumptions.
Integrated Convergence Criterion
A practical stopping rule couples both loops:

$$\| \nabla_{\Theta} \mathcal{J}_{total} \| \leq \epsilon_1, | F^{(i)} - F^{(i-1)} | \leq \epsilon_2 \quad \ldots(40)$$

for small tolerances $\epsilon_1, \epsilon_2 > 0$.
where training terminates when the inner learner is close to a stationary solution and the outer optimizer shows negligible further improvement.

Table – 1. Symbols and Description

Symbol	Description
u_j	User entity (i)
v_j	Post entity (i)
t_j	Timestamp of post (i)
y_j	Ground-truth depression label
$A^{(x)}$	Adjacency matrix for relation type (x)
S_{jk}	Semantic similarity between nodes (j) and (k)
T_{jk}	Temporal decay weight
A_{jk}	Unified weighted adjacency element
$x_j, \tilde{x}_j$	Raw and normalized node feature vectors
e_j	Transformer embedding for node (i)
s_j	Sentiment polarity scalar
h_j, h_j'	Contextual embedding and dropout output
ψ_j	Quantum state vector of node (i)
A_p, A_q	Weight matrices (self and neighbor)
ω_{jk}	Quantum neighborhood weight
λ	Residual fusion coefficient
θ	Firefly hyperparameter vector
β_{jk}	Firefly attractiveness
(ρ)	Residual calibration scalar
w_j	Attention weight for modality fusion
$u_j, v_j, r_j,$	Intermediate fusion and refined feature vectors
W_r, b_r	Final classifier weights and bias
γ	Smoothing coefficient
$F(\theta)$	Fitness of hyperparameter configuration
J_{total}	Overall objective function
ϵ_1, ϵ_2	Convergence tolerances

IV. Experimental Setup and Performance Evaluation

On social media datasets for contextual depression detection, this proposed FQ-GNN framework was evaluated. Posts and interactions were pre-processed and changed into a heterogeneous graph H = (W, F, Y). In that nodes represent users, posts, and interactions, and edges encode temporal decay and semantic resemblance. Using a transformer encoder joint with a temporal context filter to capture bothchronological interactions and textual semantics, the Node-level embeddings were generated. These embeddings were then circulated through quantum-aware graph convolutional layers to extract rich structural and semantic representations.
There are four hyperparameters: the number of quantum layers, learning rate, embedding dimensions, and fusion weights. These were optimized using a firefly-driven quantum-aware optimizer, which maximizes multi-objective fitness by precision, balancing prediction accuracy and recall. The final adaptive fusion and neighbor-aware refinement stages formed fused node representations for the classification of depression.
All experiments were conducted with varying numbers of tweet samples, ranging from 2,000 - 20,000. These are equated against three baselines: DOPR-GCDBN [24], CBA [23], and BERT-CNN [22]. Evaluation was executed on standard performance metrics.Each and every experiment was repeated multiple times to confirmthe statistical significance.

TPM Vol. 32, No. S9, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

A. Performance Evaluation Metrics
The metrics that follow were used to estimatethe performance of the model:
Accuracy (ACC): Measures the total correctness of depression predictions.
Precision (P): Ratio of properly predicted depressive instances to all predicted depressive instances, indicating the reliability of positive predictions.
Recall (R): Ratio of correctly predicted depressive instances to all actual depressive instances that reflects sensitivity.
Specificity (SP): Ability of the model to properly identify the non-depressive instances.
Depression Prediction Time (DPT): Average computational time (in seconds) essential to categorize a batch of tweets.

V. RESULTS AND DISCUSSIONS

FQ-GNN framework's performance for depression prediction is estimated in Precision, Prediction Accuracy,Recall,Specificity, and Prediction Time under varying tweet sample sizes from 2,000 - 20,000. As mentioned earlier, baseline methods: DOPR-GCDBN [24], CBA [23], and BERT-CNN [22] are included in the comparative analysis.

A. Depression Prediction Accuracy
For all the methods, the depression prediction accuracy is represented in Figure 1 and 1Table 2. The proposed FQ-GNN consistently outperforms existing works across all tweet sample sizes. For example, at 20,000 tweets, FQ-GNN achieves 98.50%, while DOPR-GCDBN records 96.74%, CBA, and BERT-CNN record 93.00% and 90.22%, respectively. This shows an improvement of 1.76% over [24], 5.81% and 8.28% over [23] and [22], respectively. Overall, the accuracy gain ranges from ~1.5–2% against [24] and ~5–8% against [22][23], representing the robust predictive capability of the framework with growing data size.

B. Recall, Precision,and Specificity
Table 3 and Figures 2–4 depict the precision, recall, and specificity metrics. For precision, FQ-GNN achieves 0.986 at 20,000 tweets, improving over 0.964 ([24]), 0.928 ([23]), and 0.913 ([22]), translating to ~2.2%, 6.5%, and 7.9% improvement, respectively. In the same way, recall rises to 0.995, while specificity attains 0.988, reflecting consistent superiority of FQ-GNN across all sample sizes. The percentage developments are more pronounced at larger dataset sizes, highlighting the scalability of the model.

C. Depression Prediction Time
Table 4 and Figure 5 report the depression prediction time in seconds. At 20,000 tweets, FQ-GNN attains 104.2 s, which is ~7.3%, 17.5%, 29.3% faster than [24], [23], and [22], respectively, representing that the proposed method is not only more accurate but also computationally efficient. The decrease in prediction time establishes the optimization efficiency embedded in FQ-GNN for large-scale tweet datasets.

FQ-GNN reliably outperforms baseline methods across all four metrics and tweet sample sizes.The development in accuracy, recall, precision, and specificity ranges from 1.5 to 2% compared to [24] and 5 to 8% compared to [22][23].The time of prediction is significantly decreased compared to all baselines, highlighting both efficiency and accuracy.Bigger datasets strengthen the benefits of FQ-GNN, assuring its robustness and scalability for depression discovery in social media analytics.These results validate that the proposed FQ-GNN framework attains superior predictive performance while sustaining low computational overhead, making it highly suitable for real-time depression monitoring applications.

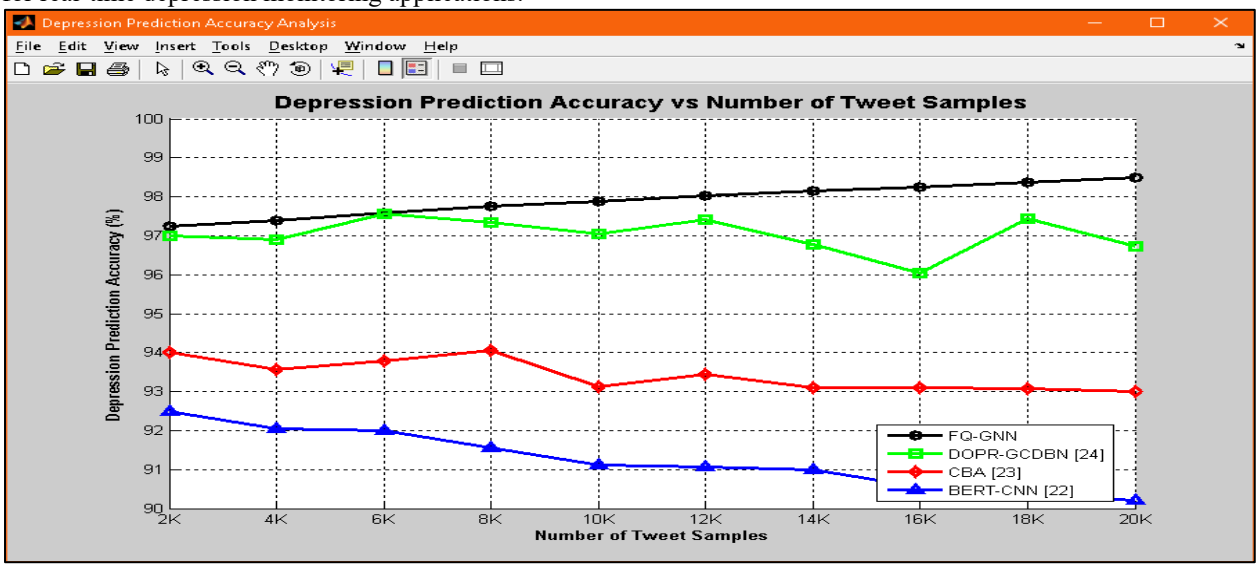

Figure – 1. Depression Prediction Accuracy Analysis
Table – 2. Depression Prediction Accuracy

TPM Vol. 32, No. S9, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Number of Tweet Samples	FQ-GNN	DOPR-GCDBN [24]	CBA [23]	BERT-CNN [22]
2000	97.25	97.00	94.00	92.50
4000	97.40	96.89	93.56	92.05
6000	97.58	97.56	93.78	92.00
8000	97.76	97.33	94.05	91.56
10000	97.88	97.05	93.12	91.11
12000	98.02	97.41	93.45	91.06
14000	98.14	96.78	93.11	91.00
16000	98.25	96.05	93.10	90.56
18000	98.38	97.45	93.08	90.33
20000	98.50	96.74	93.00	90.22

Table – 3. Precision, Recall and Specificity

Number of Tweet Samples	Precision				Recall				Specificity			
	FQ-GNN	DOPR-GCDBN [24]	CBA [23]	BERT-CNN [22]	**FQ-GNN**	DOPR-GCDBN [24]	CBA [23]	BERT-CNN [22]	**FQ-GNN**	DOPR-GCDBN [24]	CBA [23]	BERT-CNN [22]
2000	**0.973**	0.962	0.933	0.923	**0.986**	0.980	0.951	0.932	**0.975**	0.970	0.941	0.927
4000	**0.975**	0.965	0.931	0.915	**0.987**	0.975	0.948	0.931	**0.977**	0.969	0.939	0.922
6000	**0.976**	0.966	0.928	0.905	**0.988**	0.972	0.944	0.928	**0.979**	0.968	0.935	0.916
8000	**0.978**	0.958	0.932	0.912	**0.989**	0.978	0.943	0.926	**0.981**	0.967	0.937	0.918
10000	**0.980**	0.963	0.937	0.917	**0.990**	0.977	0.940	0.929	**0.982**	0.969	0.938	0.922
12000	**0.982**	0.967	0.939	0.916	**0.991**	0.981	0.947	0.927	**0.984**	0.973	0.942	0.921
14000	**0.983**	0.959	0.937	0.911	**0.992**	0.975	0.943	0.925	**0.985**	0.966	0.939	0.917
16000	**0.984**	0.963	0.927	0.907	**0.993**	0.982	0.942	0.920	**0.986**	0.972	0.934	0.913
18000	**0.985**	0.958	0.930	0.917	**0.994**	0.976	0.945	0.923	**0.987**	0.966	0.937	0.919
20000	**0.986**	0.964	0.928	0.913	**0.995**	0.979	0.938	0.920	**0.988**	0.971	0.932	0.916

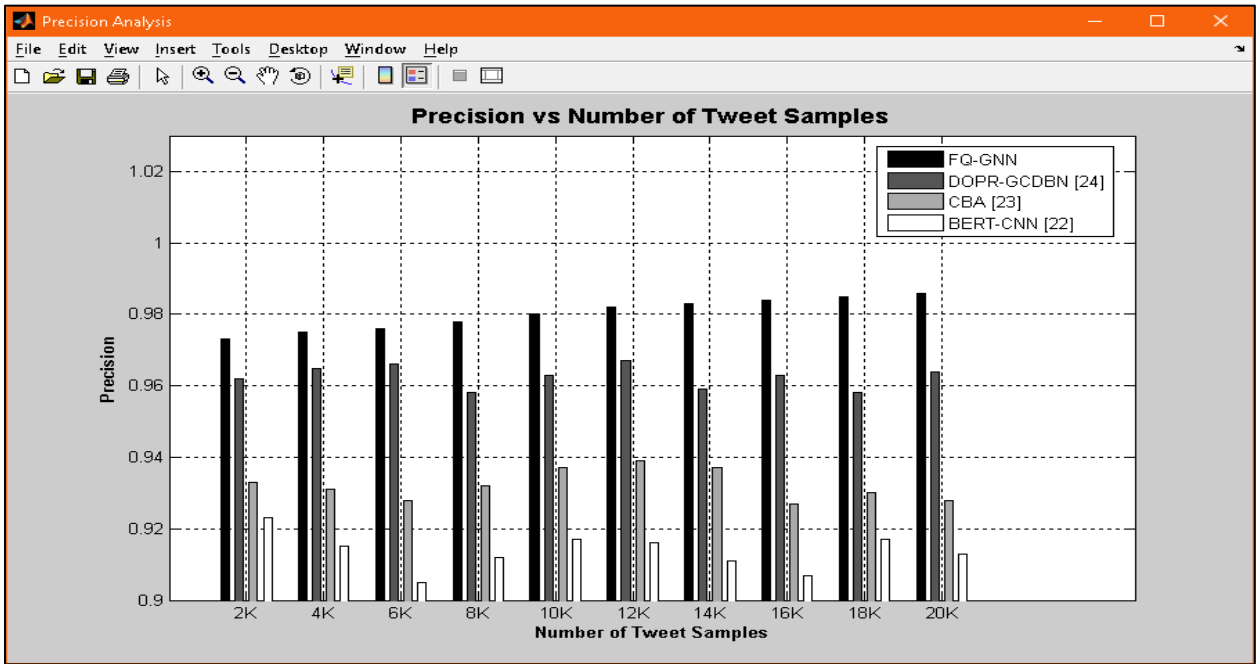

Figure – 2. Precision Analysis

TPM Vol. 32, No. S9, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

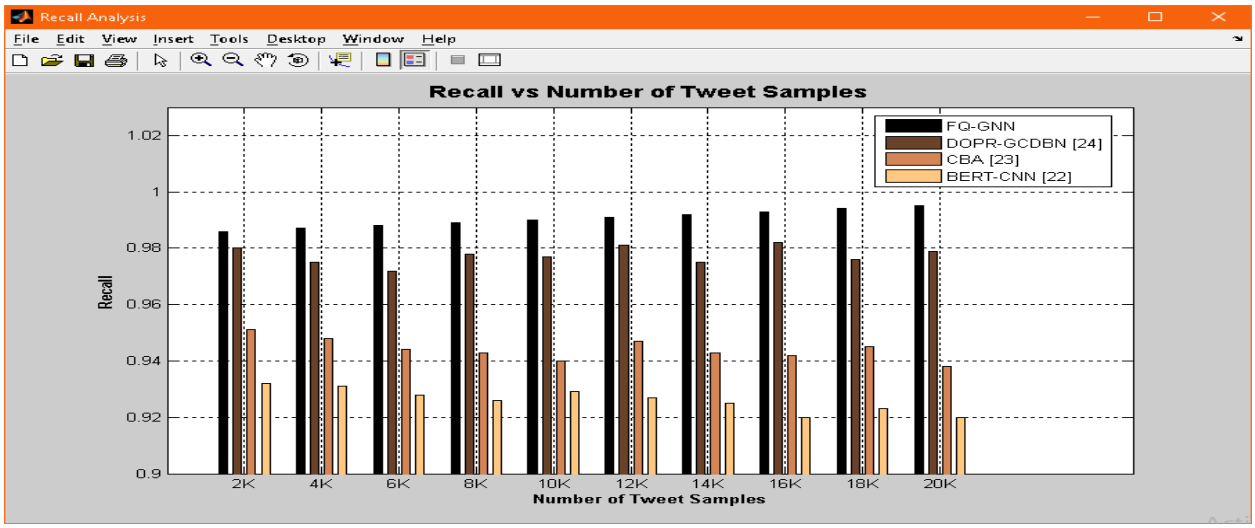

Figure – 3. Recall Analysis

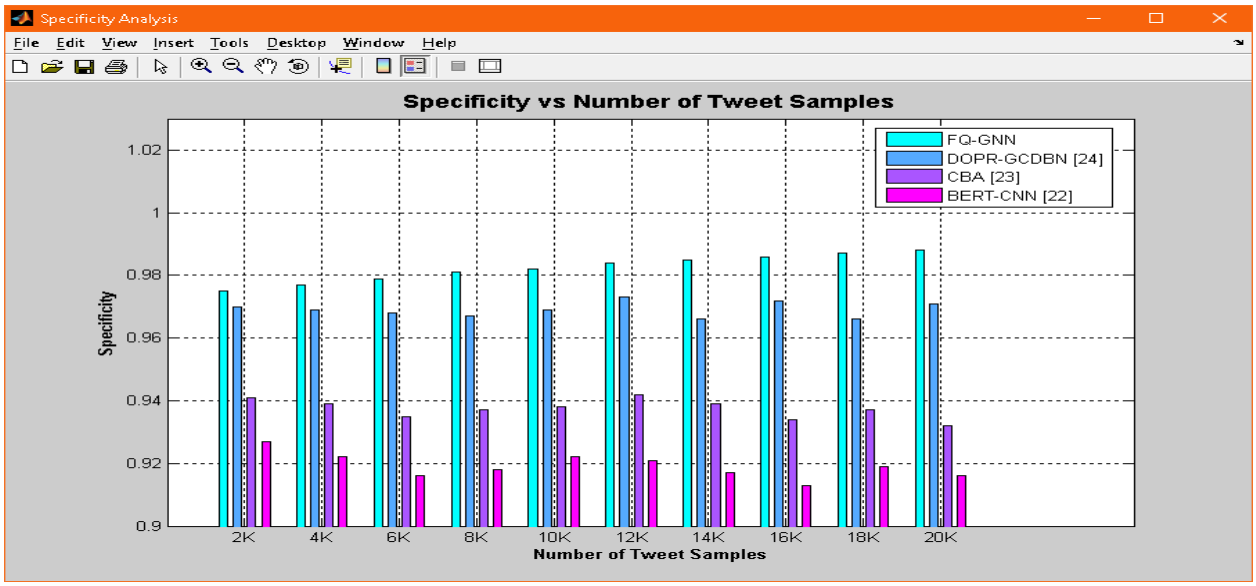

Figure – 4. Sensitivity Analysis

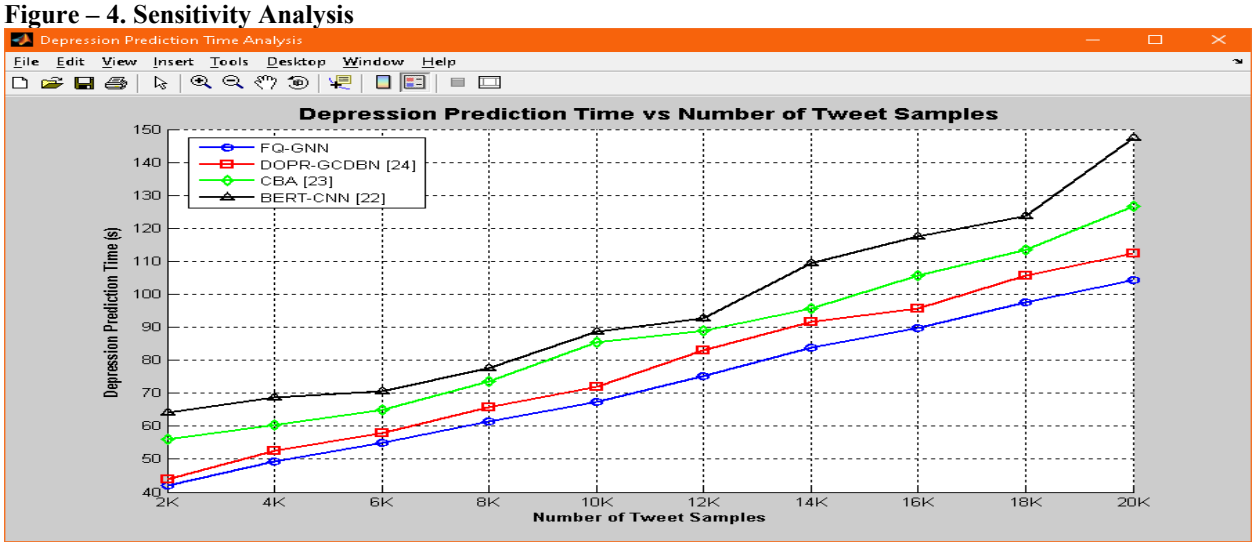

Figure – 5. Depression Prediction Time Analysis

Table – 4. Depression Prediction Time

Number of Tweet Samples	FQ-GNN	DOPR-GCDBN [24]	CBA [23]	BERT-CNN [22]
2000	**42.1**	44.0	56.0	64.0
4000	**49.2**	52.6	60.3	68.6

6000	**55.0**	58.0	65.0	70.5
8000	**61.5**	65.7	73.6	77.6
10000	**67.3**	72.0	85.5	88.7
12000	**75.1**	83.0	89.0	92.6
14000	**83.9**	91.6	95.6	109.4
16000	**89.8**	95.7	105.7	117.6
18000	**97.5**	105.6	113.5	123.8
20000	**104.2**	112.5	126.6	147.5

VI. CONCLUSION

The proposed FQ-GNN framework effectively mixestemporal, semantic, and organizationaldata from social media data to improve contextual depression detection. By firefly-based hyper-parameter optimization and leveraging quantum-aware graph convolutions, the model attainsclassification performance and superior feature representation, compared to traditional methods. Further, the adaptive fusion and neighbor-aware refinement stages improve interpretability and robustness, representing the potential of combining quantum computing principles with bio-inspired optimization in mental health analytics. The upcoming work will explore about the real-time deployment and extension to multi-modal social media signals for wider mental health monitoring.

REFERENCES

[1] D.J. Conti and W.N. Burton, "The economic impact of depression in a workplace," J. Occup. Med., vol. 36, no. 9, pp. 983–988, Sep. 1994.
[2] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," Proc. Int. Conf. Weblogs Soc. Media (ICWSM), 2013.
[3] H. Jalonen, "Social media – an arena for venting negative emotions," Online J. Commun. Media Technol., vol. 4, no. 3, pp. 1–10, Jul. 2014.
[4] S.C. Guntuku, D.B. Yaden, M.L. Kern, and L.H. Ungar, "Detecting depression and mental illness on social media: an integrative review," Curr. Opin. Behav. Sci., vol. 18, pp. 43–49, Dec. 2017.
[5] S. Evans-Lacko, D. London, L. Osborn, S. Hall, and D. Rüsch, "Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys," Psychol. Med., vol. 48, no. 9, pp. 1560–1571, Jul. 2018.
[6] F. Cacheda, D. Fernandez, F.J. Novoa, and V. Carneiro, "Early detection of depression: social network analysis and random forest techniques," J. Med. Internet Res., vol. 21, no. 6, p. e12554, Jun. 2019.
[7] R.T. Adek, R.K. Dinata, and A. Ditha, "Online newspaper clustering in Aceh using the agglomerative hierarchical clustering method," Int. J. Eng. Sci. Inf. Technol., vol. 12, no. 1, pp. 1–8, Jan. 2022.
[8] M.H. Ahmed, M.A. Hossain, and M. Rahman, "Short text clustering algorithms, application and challenges: a survey," Appl. Sci., vol. 12, no. 11, p. 3578, Jun. 2022.
[9] H. Kour and M.K. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," Multimedia Tools and Applications, vol. 81, no. 17, pp. 23649–23685, 2022.
[10] J. Singh, M. Wazid, D.P. Singh, and S. Pundir, "An embedded LSTM based scheme for depression detection and analysis," Procedia Comput. Sci., vol. 215, pp. 166–175, Dec. 2022.
[11] M. Kabir, T. Ahmed, M.B. Hasan, M.T.R. Laskar, T.K. Joarder, H. Mahmud, and K. Hasan, "DEPTWEET: A typology for social media texts to detect depression severities," Comput. Hum. Behav., vol. 107, p. 107503, Feb. 2023.
[12] D. Ren, X. Ma, H. Qin, and S. Lei, "A multi-criteria decision-making method based on discrete Z-numbers and Aczel-Alsina aggregation operators and its application on early diagnosis of depression," Eng. Appl. Artif. Intell., vol. 115, p. 105484, Jan. 2025.
[13] F.I. Kurniadi and N.L.P.S. Paramita, "BERT and RoBERTa models for enhanced detection of depression in social media text," Procedia Comput. Sci., vol. 245, pp. 202–209, Jan. 2024.
[14] X. Wei, Z. Zhang, H. Huang, and Y. Zhou, "An overview on deep clustering," Neurocomputing, vol. 487, pp. 127–144, Jan. 2024.
[15] M. Kanahuati-Ceballos, A. Alhindi, and M. Alhindi, "Detection of depressive comments on social media using RNN, LSTM, and random forest: Comparison and optimization," Soc. Netw. Anal. Mining, vol. 14, no. 1, p. 1, Dec. 2024.
[16] J. Cha, S. Kim, D. Kim, and E. Park, "MOGAM: A multimodal object-oriented graph attention model for depression detection," arXiv preprint arXiv:2403.15485, Mar. 2024.
[17] L. Bendebane, Z. Laboudi, A. Saighi, and S.E. Bouziane, "Fine-tuning the BERT model to predict depression and anxiety using multi-labeled Twitter data," Proc. 4th Int. Conf. Machine Learning Data Eng., 2025.

[18] M.M. Aldarwish, M.A. Hossain, and M. Rahman, "Predicting depression levels using social media posts," Appl. Sci., vol. 12, no. 11, p. 3578, Jun. 2022.
[19] Y. Li, R. Mihalcea, and S.R. Wilson, "Text-based detection and understanding of changes in mental health," Proc. 2018 Int. Conf. Soc. Inform., 2018.
[20] S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: a critical review," npj Digit. Med., vol. 3, p. 43, Mar. 2020.
[21] S. Mirjalili, "Firefly algorithm: Recent advances and applications," Swarm Evol. Comput., vol. 40, pp. 1–12, 2018.
[22] Joel Philip Thekkekara, Sira Yongchareon, and Veronica Liesaputra, "An Attention-Based CNN-BiLSTM Model for Depression Prediction on Social Media Text," Expert Systems with Applications, vol. 249, part c, 2024.
[23] Cao Xin, and LailatulQadriZakaria, "Integrating Bert with CNN and BiLSTM for Explainable Detection of Depression in Social Media Contents," IEEE Access, vol. 12, pp. 161203-161212, 2024.
[24] R. Geetha, D. Vimal Kumar , "Dragonfly Optimised Regressive Gradient Convolutional Deep Belief Network for Depression Prediction using Social Media Texts ", Int. J. of Engg Trends and Tech, vol. 73, no. 6, pp. 76-90, 2025.