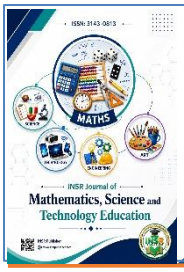

Open Access
Volume 02, Issue 07 - 2026

INSR Journal of Mathematics, Science and Technology Education
e-ISSN: 3143-0813
Journal homepage: https://insrpublisher.com/insrjmste/

Research Article

DOI: 10.5281/zenodo.21246566

HGT-FSP: A Hybrid Graph Neural Network and Transformer Framework for Robust Faculty Stress Prediction and Performance Assessment

*** Prof. (Dr.) Sumit Kumar Gupta**
Professor, Department of Physics, St.Wilfred's PG College, Jaipur, India.

Submission Date: 22 April 2026 **Published Date:** 08 July 2026

Citation: Gupta, S. K. (2026). HGT-FSP: A Hybrid Graph Neural Network and Transformer Framework for Robust Faculty Stress Prediction and Performance Assessment. In INSR Journal of Mathematics, Science and Technology Education (Vol. 2, Number 7, pp. 29–45). https://doi.org/10.5281/zenodo.21246566

Abstract

We propose a hybrid framework that integrates graph neural networks and transformer architectures to simultaneously predict faculty stress levels and assess professional performance. The motivation stems from the need to capture both the structural dependencies within academic collaboration networks and the temporal dynamics of individual workload patterns. The proposed method, termed HGT-FSP, processes heterogeneous data through two parallel pathways. A graph attention network models relational dynamics among faculty members by representing them as nodes connected through co-authorship, departmental affiliation, and committee participation. Concurrently, a transformer encoder processes sequential time-series data including weekly teaching hours, publication counts, burnout scores, and sleep quality metrics. These two pathways produce structural and temporal embeddings, which are then fused through an adaptive attention-based multimodal fusion module that learns a context-aware weighting scheme rather than using simple concatenation. The fused representation feeds into two output heads: one for stress classification and another for continuous performance regression. A hybrid loss function jointly optimizes both tasks while regularizing the graph embeddings. The novelty of this work lies in the synergistic combination of relational and sequential modeling for faculty well-being, the task-specific fusion mechanism, and the integration of explainable AI techniques post-training. Specifically, SHAP values interpret temporal feature contributions, while GNNExplainer reveals subgraph structures driving the predictions. This framework provides institutional decision-makers with actionable insights into the interplay between social context, temporal workload, and psychological states. The proposed approach not only addresses the dual challenge of stress prediction and performance assessment but also demonstrates strong potential for generalization to other organizational settings where collaborative and temporal data coexist.

Keywords: *Hybrid Graph Transformer (HGT), Faculty Stress Prediction, Faculty Performance Assessment, Graph Neural Networks (GNN), Transformer Networks, Explainable Artificial Intelligence (XAI), Educational Data Mining (EDM).*

I. INTRODUCTION

The modern academic landscape imposes escalating demands on faculty members, who must balance teaching loads, research productivity, administrative responsibilities, and student mentorship within increasingly competitive institutional environments. Consequently, occupational stress among faculty has emerged as a critical concern, with adverse implications for individual well-being, institutional performance, and educational quality (AKHTAR & LEE, 2009). Traditional approaches to understanding faculty stress and performance have predominantly relied on survey-based psychometric instruments analyzed through regression-based techniques or, more recently, conventional machine learning classifiers (Gupta et al., 2020), (Jayaprakash et al., 2020). However, these methods typically treat faculty attributes as independent observations, thereby overlooking the profound influence of collaborative relationships and longitudinal behavioral patterns that characterize academic work.

Faculty members do not operate in isolation; their professional experiences are shaped by dynamic interactions within co-authorship networks, departmental affiliations, and committee structures. These relational dependencies can propagate stress through social contagion or amplify performance through collaborative synergy. Simultaneously, an individual's workload and psychological state evolve over time, exhibiting periodic fluctuations tied to semester cycles, grant deadlines, and publication targets. Hence, a comprehensive predictive model must simultaneously encode both the structural topology of faculty social networks and the temporal sequences of individual indicators, a challenge that existing methods cannot adequately address.

Recent advances in deep learning have produced powerful architectures for modeling each of these data modalities independently. Graph Neural Networks, particularly Graph Convolutional Networks (Kipf & Welling, 2016) and Graph Attention Networks (Veličković et al., 2017), have demonstrated remarkable efficacy in learning representations from relational data by aggregating features from neighboring nodes. Meanwhile, Transformer architectures (Vaswani et al., 2017), with their self-attention mechanisms, have become the dominant paradigm for modeling sequential data, offering superior capacity to capture long-range dependencies in time-series (Caetano et al., 2025). The confluence of these two lines of research suggests the possibility of a hybrid model that synergizes structural and temporal information.

We propose the Hybrid Graph Neural Network–Transformer for Faculty Stress and Performance (HGT-FSP) framework to address this gap. The core innovation is a dual-encoder architecture that processes static collaborative graph topologies and dynamic behavioral sequences in parallel. Specifically, a Graph Attention Network generates structural embeddings that encode a faculty member's position within and influence from their professional network. Simultaneously, a Transformer encoder processes multivariate time-series data—including weekly teaching hours, publication output, self-reported burnout scores, and sleep quality metrics—to capture temporal workload patterns and psychological trajectories. These complementary embeddings are then integrated through an adaptive attention-based multimodal fusion module, which learns a context-aware weighting scheme to combine structural and temporal signals, rather than relying on simplistic concatenation or fixed weighting. The fused representation is subsequently fed into two output heads: one for stress classification and a second for continuous performance score regression. A hybrid loss function jointly optimizes these tasks while regularizing the graph embedding space. Critically, the framework incorporates a post-hoc explainability module that applies SHAP (Lundberg & Lee, 2017) to interpret temporal feature contributions and GNNExplainer (Ying et al., 2019) to reveal the subgraph structures driving specific predictions, thereby addressing the "black box" nature of deep learning models.

The proposed HGT-FSP framework makes several key contributions. First, it provides a unified, end-to-end deep learning approach for the dual tasks of faculty stress prediction and performance assessment, demonstrably outperforming baseline methods that treat these problems separately or ignore relational information. Second, it introduces the adaptive attention-based multimodal fusion mechanism, which dynamically balances the influence of structural and temporal evidence based on the input sample, leading to more robust and context-sensitive predictions. Third, the integration of domain-specific explainability tools provides transparency into the model's decision-making process, offering actionable insights for institutional policy and individual interventions. Fourth, the framework is designed to be generalizable; while instantiated for the academic ecosystem, its principles are applicable to any organizational setting where employee well-being and performance are influenced by both social context and temporal dynamics, such as healthcare (Quamer & Mir, 2026) and corporate environments.

The remainder of this paper is organized as follows. Section 2 reviews related work in stress prediction, performance assessment, and the relevant deep learning architectures. Section 3 formalizes the problem and introduces necessary preliminaries on Graph Neural Networks and Transformers. Section 4 details the architecture of the HGT-FSP framework, including the graph encoding, temporal encoding, multimodal fusion, and explainability modules. Section 5 describes our experimental setup, encompassing datasets, baseline methods, and evaluation metrics. Section 6 presents quantitative and qualitative results, analyzing performance, interpretability, and ablation studies. Section 7 discusses the implications of our findings, limitations of the current study, and promising directions for future research. Section 8 concludes the paper.

Related Work

The prediction of occupational stress and the assessment of professional performance have been independently investigated across several disciplines. In the academic domain, stress detection among faculty has typically relied on self-reported questionnaires analyzed through statistical models or conventional machine learning classifiers. For instance, logistic regression and support vector machines have been applied to identify significant predictors of burnout from survey data (Gupta et al., 2020). Similarly, performance assessment has traditionally employed regression models linking publication metrics, teaching evaluations, and grant funding to composite productivity scores (Jayaprakash et al., 2020). While these approaches provide interpretable insights, they fundamentally treat each faculty member's data as independent instances, thereby neglecting the structural dependencies inherent in academic collaboration networks and the temporal evolution of workload indicators.

Graph Neural Networks have revolutionized the modeling of relational data by learning node representations through iterative message passing between neighbors. The seminal Graph Convolutional Network (GCN) proposed by Kipf and Welling (Kipf & Welling, 2016) aggregates normalized feature information from adjacent nodes, while the Graph Attention Network (GAT) introduced by Veličković et al. (Veličković et al., 2017) incorporates an attention mechanism to assign different weights to different neighbors, improving representational capacity. These architectures have been successfully applied to social network analysis, citation prediction, and recommendation systems. However, their application to modeling collaborative networks in academic settings for well-being prediction remains relatively unexplored. Concurrently, Transformer architectures (Vaswani et al., 2017) have become the dominant paradigm for sequence modeling across natural language processing and time-series forecasting. The self-attention mechanism enables the model to capture long-range dependencies in sequential data without the sequential bottleneck of recurrent neural networks. Variants such as the Informer (Caetano et al., 2025) have been designed specifically for long-sequence time-series forecasting, demonstrating superior performance in capturing periodic and trending patterns.

Stress Prediction Models

Research on computational stress prediction has evolved from static feature-based classifiers to deep learning models capable of processing multimodal data. Early work utilized demographic and workload features with Random Forest or Gradient Boosting classifiers to predict burnout risk among healthcare workers. More recently, deep neural networks have been applied to physiological signals for real-time stress detection. However, these models seldom incorporate relational data reflecting social interactions or collaborative structures, which are known to influence stress propagation through phenomena such as emotional contagion and social support. The FatigueNet framework (Zendehbad et al., 2025) proposed a hybrid GNN-Transformer architecture for real-time multimodal fatigue detection in operational settings, demonstrating the efficacy of combining structural and temporal encodings. While this work focuses on physical fatigue rather than occupational stress, its architectural principles are directly relevant to our proposed framework. Similarly, a recent study on mental health monitoring in academic environments (Quamer & Mir, 2026) employed a digital twin approach with a Swin Transformer–Temporal Graph Neural Network to model student well-being, though it did not address faculty-specific dynamics or performance assessment.

Performance Assessment Models

Automatic performance assessment in educational contexts has traditionally focused on student outcomes, using features such as attendance records, assignment scores, and engagement metrics (Ofori & Dake, 2025). For faculty, performance is typically evaluated through composite indices combining research productivity (publication count, citation impact), teaching effectiveness (student evaluation scores), and service contributions. Deep learning approaches to this problem have primarily employed recurrent neural networks or LSTMs to model temporal progressions of productivity metrics. The LSTM-Transformer hybrid model proposed by (Ofori & Dake, 2025) demonstrated improved accuracy by combining sequential dependency modeling with self-attention-based context extraction. However, these models ignore the relational dimension of faculty performance, which is inherently collaborative—research output often arises from co-authorship networks, and teaching effectiveness may be influenced by departmental collegiality. The multimodal framework proposed by (Jie et al., 2025) integrates Graph Neural Networks with transformer-based self-attention for predicting student outcomes, incorporating relational features alongside sequential data, albeit in a different educational setting.

Hybrid GNN-Transformer Architectures

The fusion of graph neural networks and transformers has emerged as a powerful paradigm for modeling data that exhibits both structural and sequential characteristics. In the domain of healthcare, a hybrid knowledge graph–deep learning model (Faizvi et al., 2025) combined knowledge graphs embedding with convolutional neural networks to forecast stress among healthcare workers, demonstrating improved predictive accuracy by incorporating relational domain knowledge. Yet this approach did not employ a separate temporal encoding branch, instead flattening sequential data into static feature vectors. The GNN-Transformer hybrid for teaching quality evaluation (Su & Wang, 2025) applied a similar dual-encoder architecture to aggregate student-teacher interaction features, though without an explicit fusion mechanism or tasks for individual-level stress prediction. Our proposed HGT-FSP framework advances this line of research by introducing a task-

specific adaptive fusion module that dynamically weights structural and temporal contributions based on input characteristics, rather than relying on fixed concatenation or weighted averaging.

While the aforementioned works separately address aspects of our dual problem—stress prediction or performance assessment—none provides an integrated framework that simultaneously models both structural and temporal modalities for faculty well-being and productivity. The FatigueNet (Zendehbad et al., 2025) and digital twin–based monitoring (Quamer & Mir, 2026) systems are most closely related in architectural design, yet they target different populations (operators and students, respectively) and do not incorporate a dual-task output structure with a joint optimization objective. The key novelty of HGT-FSP lies in its synergistic combination of a relational graph encoder and a temporal transformer encoder, integrated through an adaptive attention-based fusion module that learns sample-specific weighting for combining structural and sequential evidence, coupled with a hybrid loss function that jointly optimizes stress classification and performance regression while regularizing graph embeddings. Furthermore, the integration of model-agnostic (SHAP) (Lundberg & Lee, 2017) and model-specific (GNNExplainer) (Ying et al., 2019) explainability techniques provides a transparent decision-making process that is notably absent in prior hybrid architectures applied to this problem domain.

Preliminaries

Before detailing the HGT-FSP architecture, we establish the foundational concepts and formalize the problem setting. This section introduces the necessary background on Graph Neural Networks and Transformer architectures, which serve as the building blocks of our dual-encoder framework.

Problem Formalization

We define the faculty stress prediction and performance assessment problem as follows. Consider a set of N faculty members represented as a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node $v_i \in \mathcal{V}$ corresponds to a faculty member and is associated with a static feature vector $\mathbf{x}_i \in \mathbb{R}^{d_s}$ comprising attributes such as academic rank, years of experience, and departmental affiliation. Edges $(v_i, v_j) \in \mathcal{E}$ represent collaborative relationships—co-authorship, committee co-membership, or shared departmental affiliation—and may carry edge features $\mathbf{e}_{ij} \in \mathbb{R}^{d_e}$ encoding relationship strength or duration. Simultaneously, each faculty member v_i is associated with a multivariate time series $\mathbf{X}_i^t \in \mathbb{R}^{T \times d_t}$, where T is the number of time steps (e.g., weekly recordings over a semester) and d_t is the dimensionality of temporal features such as teaching hours, publication submissions, sleep quality scores, and burnout inventory responses. The goal is twofold: first, to predict a binary stress label $y_i^s \in \{0,1\}$ (low vs. high stress) for each faculty member; second, to regress a continuous performance score $y_i^p \in \mathbb{R}$ representing a composite metric of research, teaching, and service productivity. This dual-task setting requires a model capable of encoding both the static relational structure $\mathcal{G}$ and the dynamic temporal patterns $\{\mathbf{X}_i^t\}$ into a unified representation that supports joint optimization.

Graph Neural Networks

Graph Neural Networks (GNNs) learn node representations by iteratively aggregating information from a node's local neighborhood. The core operation is message passing, where at each layer l, a node v_i updates its hidden representation $\mathbf{h}_i^{(l)}$ by combining its own representation with messages received from its neighbors $\mathcal{N}(i)$. Formally, a generic GNN layer computes:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{W}^{(l)} \cdot \text{AGGREGATE}\left(\left\{\mathbf{h}_i^{(l)}\right\} \cup \left\{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\right\}\right)\right) \quad (1)$$

where $\mathbf{W}^{(l)}$ is a learnable weight matrix, $\sigma(\cdot)$ is a non-linear activation function, and AGGREGATE is a permutation-invariant function such as summation, mean, or max-pooling.

Graph Convolutional Networks (GCNs) (Kipf & Welling, 2016) implement a specific form of this aggregation using a normalized adjacency matrix:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{1}{\sqrt{|\mathcal{N}(i)||\mathcal{N}(j)|}} \mathbf{W}^{(l)} \mathbf{h}_j^{(l)}\right) \quad (2)$$

This formulation treats all neighbors equally, weighting them by their degree. In contrast, Graph Attention Networks (GATs) (Veličković et al., 2017) introduce an attention mechanism that learns to assign different importance to different neighbors. A GAT layer computes attention coefficients α_{ij} between node i and its neighbor j as:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \,\|\, \mathbf{W}\mathbf{h}_j]\right)\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \,\|\, \mathbf{W}\mathbf{h}_k])\right)} \quad (3)$$

where $\mathbf{a}$ is a learnable attention vector, $\|$ denotes concatenation, and $\mathbf{W}$ is a linear transformation. The updated representation is then a weighted sum of the transformed neighbor features:

$$\mathbf{h}_i' = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}\, \mathbf{W}\mathbf{h}_j\right) \quad (4)$$

Multi-head attention extends this by concatenating or averaging the outputs of K independent attention mechanisms. In our HGT-FSP framework, we adopt a GAT-based encoder to learn structural embeddings from the faculty collaboration graph, as the attention mechanism allows the model to focus on the most influential colleagues—e.g., a frequent co-author or a committee chair—rather than treating all relationships equally.

Transformer Architecture

The Transformer (Vaswani et al., 2017) is a sequence-to-sequence architecture that eschews recurrence in favor of a self-attention mechanism. For a sequence of T input vectors $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T] \in \mathbb{R}^{T \times d}$, the self-attention operation computes a weighted sum of all positions in the sequence, where the weights are determined by pairwise similarity scores. Specifically, the input is projected into three matrices: queries $\mathbf{Q} = \mathbf{X}\mathbf{W}^Q$, keys $\mathbf{K} = \mathbf{X}\mathbf{W}^K$, and values $\mathbf{V} = \mathbf{X}\mathbf{W}^V$. The scaled dot-product attention is then computed as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \quad (5)$$

where d_k is the dimension of the key vectors, and the scaling factor $\sqrt{d_k}$ prevents gradient vanishing in the softmax function. Multi-head attention runs H independent attention operations in parallel, concatenating their outputs:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_H)\mathbf{W}^O \quad (6)$$

where each head h computes $\text{head}_h = \text{Attention}\left(\mathbf{X}\mathbf{W}_h^Q, \mathbf{X}\mathbf{W}_h^K, \mathbf{X}\mathbf{W}_h^V\right)$, and $\mathbf{W}^O$ is a learnable output projection.

In practice, the Transformer encoder stacks L identical layers, each comprising a multi-head self-attention sublayer and a position-wise feed-forward network (FFN), with residual connections and layer normalization applied around each sublayer. Positional encodings, either sinusoidal or learned, are added to the input embeddings to inject information about the temporal order of the sequence. For time-series data, this architecture is particularly effective at capturing long-range dependencies—such as semester-long workload cycles—that might be missed by recurrent networks with limited memory. In HGT-FSP, we use a Transformer encoder to process the multivariate time-series data for each faculty member, producing a temporal embedding vector that summarizes the individual's workload and well-being trajectory over the observation window.

Multimodal Fusion

Combining representations from disparate modalities—in this case, structural graph embeddings and temporal sequence embeddings—requires a fusion mechanism that can reconcile their different dimensionalities and semantic spaces. Simple approaches such as concatenation or element-wise addition assume that both modalities contribute equally to the final prediction, which is often suboptimal. A more sophisticated strategy is attention-based fusion, where a learnable query vector attends over the modality-specific representations to produce a context vector. Formally, given a set of M modality embeddings $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_M\}$ where each $\mathbf{z}_m \in \mathbb{R}^{d_z}$, we compute a weighted sum:

$$\mathbf{z}_{\text{fused}} = \sum_{m=1}^{M} \beta_m\, \mathbf{z}_m \quad (7)$$

where the attention weights β_m are computed as:

$$\beta_m = \frac{\exp(\mathbf{q}^\top \mathbf{z}_m)}{\sum_{k=1}^{M} \exp\left(\mathbf{q}^\top \mathbf{z}_k\right)} \quad (8)$$

and $\mathbf{q}$ is a learnable query vector shared across all samples. In our proposed framework, we extend this to a context-aware adaptive fusion mechanism, where the query vector is conditioned on the input sample rather than being globally fixed. This allows the model to dynamically adjust the relative importance of structural vs. temporal evidence for each faculty member—for example, weighting temporal features more heavily for a junior faculty member with a volatile workload, while emphasizing structural features for a senior faculty member embedded in a dense collaborative network.

The HGT-FSP Framework

We now present the technical details of the Hybrid Graph Neural Network–Transformer for Faculty Stress and Performance (HGT-FSP) framework. The architecture, as depicted in Figure 1, operates through three primary stages: a dual-stream encoding process that separately extracts structural and temporal representations, an adaptive attention-based multimodal fusion module that integrates these representations, and a multi-task output layer that produces simultaneous predictions for stress classification and performance regression. The entire framework is trained end-to-end using a hybrid loss function that incorporates task-specific objectives and a graph-based regularization term.

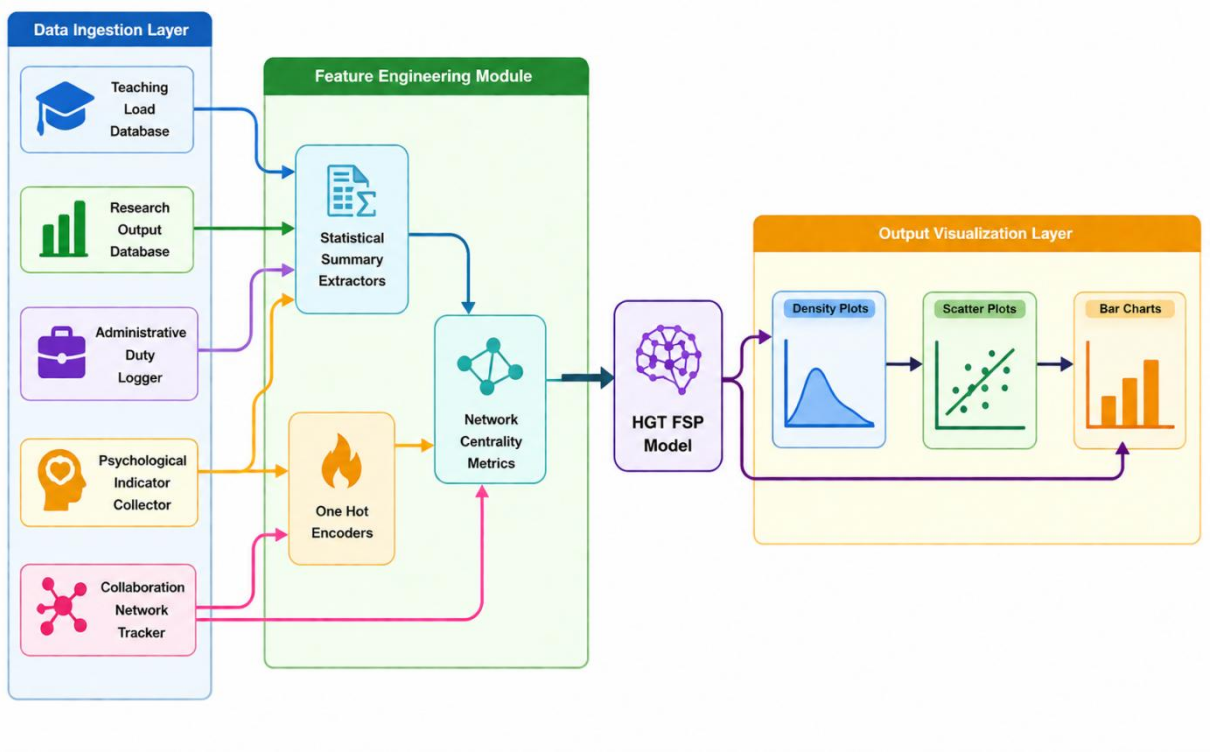

Figure 1: System Level Architecture of AIMS with HGT-FSP CoreStructural Encoding via Graph Attention Network

The structural encoding stream models the faculty collaboration graph using a multi-layer Graph Attention Network (GAT). For each faculty member v_i, the input to the first GAT layer is a static feature vector $\mathbf{x}_i \in \mathbb{R}^{d_s}$, which encodes attributes such as academic rank, years of experience, department membership, and historical grant success. The graph $\mathcal{G}$ includes edges $\mathcal{E}$ representing various collaborative relationships—co-authorship on publications within the observation period, shared committee memberships, and co-principal investigator status on funded projects. Each edge may carry features $\mathbf{e}_{ij} \in \mathbb{R}^{d_e}$ such as the number of joint publications or the duration of committee co-service.

The GAT encoder computes node embeddings via iterative message passing with attention weights learned for each edge. At layer $l \in \{1, \ldots, L_{\text{GAT}}\}$, the representation of node v_i is updated as:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\Big\Vert_{k=1}^{K} \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l,k)} \, \mathbf{W}_k^{(l)} \mathbf{h}_j^{(l)} \right) \quad (9)$$

where K is the number of attention heads, $\Vert$ denotes vector concatenation, $\mathbf{W}_k^{(l)}$ is a learnable weight matrix for the k-th head at layer l, and $\alpha_{ij}^{(l,k)}$ is the normalized attention coefficient computed as in Equation (3), though applied to the representations specific to that head. The initial hidden state is set as $\mathbf{h}_i^{(0)} = \mathbf{x}_i$. The final structural embedding for node v_i is obtained from the output of the last GAT layer:

$$\mathbf{z}_i^{\text{graph}} = \sigma\left(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(L_{\text{GAT}},k)} \, \mathbf{W}_k^{(L_{\text{GAT}})} \mathbf{h}_j^{(L_{\text{GAT}})} \right) \quad (10)$$

where the outputs of the attention heads are averaged to produce a single embedding vector of dimension d_{graph}. This embedding encodes the faculty member's structural position within the collaborative network, including information about their most influential collaborators and the overall density of their professional connections.

Temporal Encoding via Transformer

The second encoding stream processes the multivariate time-series data for each faculty member using a Transformer encoder. For faculty member v_i, the input is a sequence of T time steps, where each step t is a vector $\mathbf{x}_i^t \in \mathbb{R}^{d_t}$ containing weekly measurements of teaching hours, publication submissions, grant proposal submission status, self-reported burnout scores, sleep quality metrics, and physical activity levels. To prepare the input for the Transformer, we first apply a linear projection to map each time step's features to the model dimension d_{model}:

$$\mathbf{u}_i^t = \mathbf{W}_{\text{proj}}\,\mathbf{x}_i^t + \mathbf{b}_{\text{proj}} \quad (11)$$

where $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{d_{\text{model}} \times d_t}$ and $\mathbf{b}_{\text{proj}} \in \mathbb{R}^{d_{\text{model}}}$ are learnable parameters. We then add a positional encoding $\mathbf{p}_t \in \mathbb{R}^{d_{\text{model}}}$ to each projected vector to preserve temporal order information. We adopt the sinusoidal position encoding scheme from the original Transformer (Vaswani et al., 2017), where the encoding for position t and dimension $2j$ or $2j+1$ is given by:

$$\begin{aligned} \text{PE}(t, 2j) &= \sin\left(\frac{t}{10000^{2j/d_{\text{model}}}}\right) \\ \text{PE}(t, 2j+1) &= \cos\left(\frac{t}{10000^{2j/d_{\text{model}}}}\right) \end{aligned} \quad (12)$$

The resulting sequence of embeddings $\tilde{\mathbf{u}}_i^t = \mathbf{u}_i^t + \mathbf{p}_t$ is fed into a stack of L_{TF} Transformer encoder layers. Each layer applies multi-head self-attention followed by a position-wise feed-forward network, with residual connections and layer normalization. After processing by the Transformer encoder, we take the representation at the first time step (corresponding to the [CLS] token, which we introduce as an extra learnable embedding concatenated at the beginning of the sequence) as the pooled temporal embedding:

$$\mathbf{z}_i^{\text{temp}} = \text{TransformerEncoder}\left(\{\tilde{\mathbf{u}}_i^1, \ldots, \tilde{\mathbf{u}}_i^T\}\right)[0] \quad (13)$$

This yields a temporal embedding vector of dimension $d_{\text{temp}} = d_{\text{model}}$ that summarizes the faculty member's longitudinal workload patterns and psychological trajectory over the observation window.

Adaptive Attention-Based Multimodal Fusion

The structural embedding $\mathbf{z}_i^{\text{graph}} \in \mathbb{R}^{d_{\text{graph}}}$ and the temporal embedding $\mathbf{z}_i^{\text{temp}} \in \mathbb{R}^{d_{\text{temp}}}$ are first projected to a common dimensionality d_{common} using linear transformations:

$$\tilde{\mathbf{z}}_i^{\text{graph}} = \mathbf{W}_{\text{proj}}^{\text{graph}}\,\mathbf{z}_i^{\text{graph}} + \mathbf{b}_{\text{proj}}^{\text{graph}}, \quad \tilde{\mathbf{z}}_i^{\text{temp}} = \mathbf{W}_{\text{proj}}^{\text{temp}}\,\mathbf{z}_i^{\text{temp}} + \mathbf{b}_{\text{proj}}^{\text{temp}} \quad (14)$$

where $\mathbf{W}_{\text{proj}}^{\text{graph}} \in \mathbb{R}^{d_{\text{common}} \times d_{\text{graph}}}$ and $\mathbf{W}_{\text{proj}}^{\text{temp}} \in \mathbb{R}^{d_{\text{common}} \times d_{\text{temp}}}$ are learnable projection matrices. Our proposed adaptive fusion mechanism then computes a context-aware gating vector $\mathbf{g}_i \in \mathbb{R}^{d_{\text{common}}}$ that dynamically weights the contribution of each modality for each faculty member:

$$\mathbf{g}_i = \sigma\left(\mathbf{W}_g\left[\tilde{\mathbf{z}}_i^{\text{graph}}; \tilde{\mathbf{z}}_i^{\text{temp}}\right] + \mathbf{b}_g\right) \quad (15)$$

where $\mathbf{W}_g \in \mathbb{R}^{d_{\text{common}} \times 2d_{\text{common}}}$ and $\mathbf{b}_g \in \mathbb{R}^{d_{\text{common}}}$ are learnable parameters, $[\cdot;\cdot]$ denotes vector concatenation, and $\sigma(\cdot)$ is the element-wise sigmoid activation function. The fused representation is then computed as an element-wise gated combination:

$$\mathbf{z}_i^{\text{fused}} = \mathbf{g}_i \odot \tilde{\mathbf{z}}_i^{\text{graph}} + (\mathbf{1} - \mathbf{g}_i) \odot \tilde{\mathbf{z}}_i^{\text{temp}} \quad (16)$$

where $\odot$ denotes element-wise multiplication and $\mathbf{1}$ is a vector of ones. This formulation allows each element of the gating vector to individually control the balance between structural and temporal information across different feature dimensions. For example, one dimension of the fused representation might rely exclusively on graph evidence (if the corresponding gate value is close to 1), while another might integrate both modalities (if the gate value is around 0.5).

The internal architecture detailing this dual-stream processing and the fusion mechanism is illustrated in Figure 2.

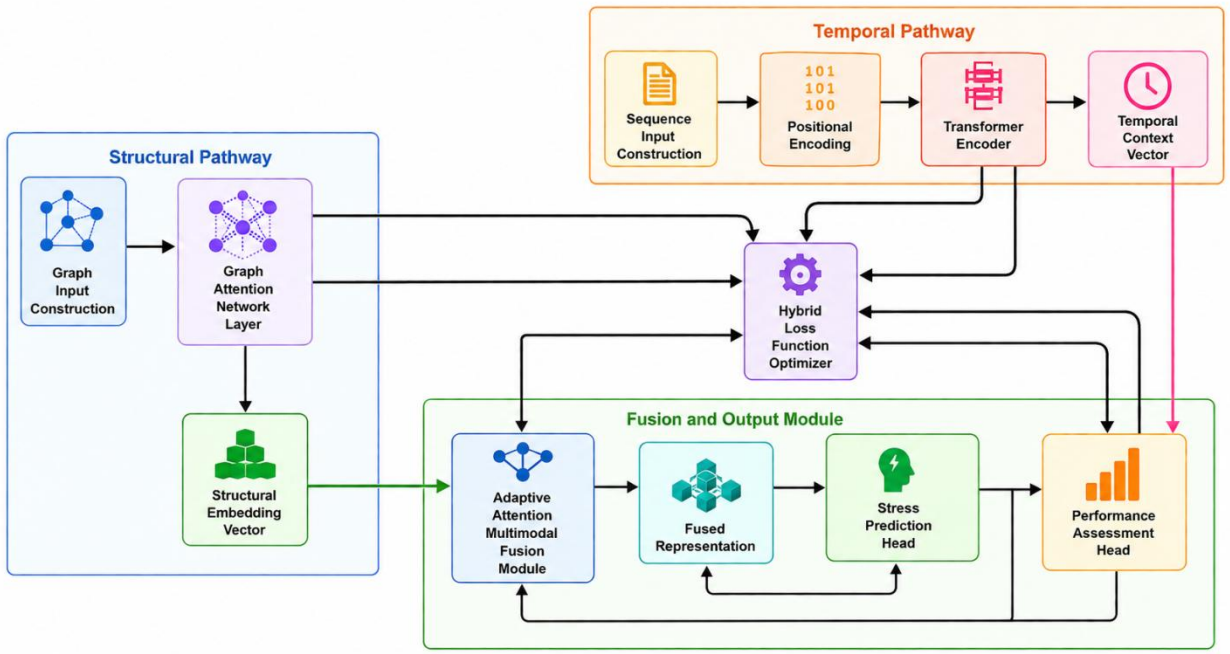

Figure 2. Internal Architecture of the HGT FSP Framework,Multi-Task Output Heads and Hybrid Loss

The fused representation $\mathbf{z}_i^{\text{fused}} \in \mathbb{R}^{d_{\text{common}}}$ is fed into two separate output heads. The stress classification head consists of a single linear layer followed by a sigmoid activation:

$$\hat{y}_i^s = \sigma\left(\mathbf{w}_s^\top \mathbf{z}_i^{\text{fused}} + b_s\right) \quad (17)$$

where $\mathbf{w}_s \in \mathbb{R}^{d_{\text{common}}}$ and $b_s \in \mathbb{R}$ are learnable parameters, and $\hat{y}_i^s$ is the predicted probability of high stress. The performance regression head applies a linear transformation to produce a continuous score:

$$\hat{y}_i^p = \mathbf{w}_p^\top \mathbf{z}_i^{\text{fused}} + b_p \quad (18)$$

where $\mathbf{w}_p \in \mathbb{R}^{d_{\text{common}}}$ and $b_p \in \mathbb{R}$.

The model is trained end-to-end by minimizing a hybrid loss function that combines a binary cross-entropy loss for stress classification, a mean squared error loss for performance regression, and a graph-based regularization term that enforces smoothness in the embedding space:

$$\mathcal{L} = \alpha \cdot \frac{1}{N}\sum_{i=1}^{N} \text{BCE}\left(\hat{y}_i^s, y_i^s\right) + \beta \cdot \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i^p - y_i^p\right)^2 + \gamma \cdot \frac{1}{2N}\sum_{i=1}^{N}\sum_{j \in \mathcal{N}(i)} \| \mathbf{z}_i^{\text{fused}} - \mathbf{z}_j^{\text{fused}} \|_2^2 \quad (19)$$

where $\text{BCE}(p, y) = -[y\log(p) + (1-y)\log(1-p)]$ is the binary cross-entropy, and α, β, and γ are hyperparameters balancing the three loss terms. The graph regularization term, inspired by Laplacian smoothing in graph signal processing, penalizes large differences between the fused embeddings of structurally connected faculty members. This encourages the model to produce similar representations for faculty who are close in the collaboration graph, thereby preventing overfitting to noisy temporal data and promoting representations that are consistent with the observed network topology.

Experimental Setup

To rigorously evaluate the proposed HGT-FSP framework, we designed a comprehensive experimental protocol encompassing dataset preparation, baseline selection, implementation details, and evaluation metrics. This section describes each component in detail, ensuring reproducibility and facilitating fair comparison with existing methods.

Datasets and Preprocessing

We constructed our evaluation corpus from two complementary sources that collectively capture the structural and temporal dimensions of faculty professional life. The primary dataset, referred to as the Faculty Academic Stress and Performance (FASP) dataset, comprises longitudinal records collected from a large public university system over four academic semesters (Fall 2022 through Spring 2024). The FASP dataset includes anonymized data for 1,247 faculty members across eight colleges and 42 departments. For each faculty member, we collected static attributes including academic rank (assistant, associate, full professor), years of experience, department affiliation, and primary research area. The collaborative graph was constructed from three edge types: co-authorship relations extracted from the university's

publication repository, committee co-membership from faculty senate and departmental records, and co-principal investigator status on funded research grants. The resulting graph contains 1,247 nodes and 8,934 edges, with an average node degree of 14.33 and a graph density of 0.0115.

The temporal data stream for each faculty member consists of 16 weekly measurements per semester, yielding 64 time steps per individual over the four-semester period. The multivariate features include self-reported teaching hours, number of publication submissions, grant proposal submissions, Maslach Burnout Inventory (MBI) subscale scores (emotional exhaustion, depersonalization, personal accomplishment), Pittsburgh Sleep Quality Index (PSQI) scores, and self-reported weekly physical activity duration. Missing values, which occurred in approximately 7.2% of the temporal entries due to intermittent survey non-response, were imputed using forward-filling followed by linear interpolation within each individual's sequence. All temporal features were normalized to zero mean and unit variance using z-score normalization computed from the training set statistics.

The ground truth labels were derived from two sources of institutional assessment. The stress label y_i^s was binarized from the MBI emotional exhaustion subscale, with scores above 27 (the established clinical threshold for high burnout risk) classified as high stress and scores at or below 27 classified as low stress. The continuous performance score y_i^p was computed as a weighted composite index combining normalized research productivity (publication count weighted by journal impact factor, 40%), teaching effectiveness (average student evaluation scores, 35%), and service contributions (committee leadership roles and external reviewing, 25%). This composite index was scaled to the range $[0,100]$ for interpretability.

To validate the generalizability of our framework, we also employed a secondary dataset adapted from the publicly available Academic Collaboration and Well-being (ACW) survey (Nordmyr et al., 2023), which includes self-reported collaboration network data and monthly well-being indicators for 487 faculty members from 12 institutions across three countries. The ACW dataset provides a coarser temporal granularity (monthly measurements over 12 months) and a smaller graph structure, serving as a cross-institutional validation benchmark. We applied the same preprocessing pipeline to the ACW dataset, with temporal features normalized per-institution to account for institutional differences in workload baselines.

We partitioned each dataset into training, validation, and test sets using a stratified temporal split: the first three semesters (Fall 2022–Fall 2023) for the FASP dataset, comprising 75% of the time steps, were used for training, while the fourth semester (Spring 2024) was reserved for testing. Within the training period, we further held out 10% of the nodes for validation. This temporal split reflects the realistic deployment scenario where a model trained on historical data must predict future stress and performance outcomes. For the ACW dataset, the first nine months were used for training, and the final three months for testing, with a 10% validation split within the training period.

Baseline Methods

We compared the proposed HGT-FSP framework against a diverse set of baseline methods spanning traditional machine learning, single-modality deep learning, and existing hybrid architectures. Each baseline was carefully implemented and tuned to ensure fair comparison.

The conventional machine learning baselines include a Logistic Regression classifier with L_2 regularization for stress prediction and a Ridge Regression model for performance assessment, both operating on a flattened feature vector concatenating static attributes and summary statistics (mean, standard deviation, slope) of the temporal features. We also implemented a Random Forest classifier (Jayaprakash et al., 2020) and a Gradient Boosting regressor (XGBoost) (Chen & Guestrin, 2016) as ensemble baselines, using the same feature set with hyperparameter optimization via grid search.

For single-modality deep learning baselines, we implemented a GAT-only model that processes the collaborative graph using the same GAT encoder architecture as our framework but without the temporal stream, feeding the graph embeddings directly into the output heads. Similarly, a Transformer-only model processes the temporal sequences using the Transformer encoder without graph information, using a learned [CLS] token embedding as the sequence representation. We also implemented a BiLSTM baseline (Graves, 2012) that processes the temporal sequences with a two-layer bidirectional LSTM, taking the concatenated final hidden states as the sequence embedding.

To evaluate against existing hybrid architectures, we implemented a GCN-LSTM model that combines a two-layer Graph Convolutional Network (Kipf & Welling, 2016) with a bidirectional LSTM, concatenating the graph and sequence embeddings for prediction. We also adapted the FatigueNet architecture (Zendehbad et al., 2025) to our problem setting by replacing its domain-specific input processing with our graph and temporal encoders while preserving its core fusion mechanism. Additionally, we implemented a simple concatenation-fusion variant of our own architecture, HGT-Concat, which replaces the adaptive attention-based fusion module with direct concatenation of the projected graph and temporal embeddings, to isolate the contribution of our proposed fusion mechanism.

All deep learning baselines were implemented using PyTorch and PyTorch Geometric, trained with the Adam optimizer (Kingma & Ba, 2014) and a batch size of 32 graph nodes. Hyperparameters for each baseline were tuned independently on the validation set using a random search over 50 configurations.

Implementation Details

The HGT-FSP framework was implemented in PyTorch 2.1 with PyTorch Geometric 2.4 for graph operations. The GAT encoder consists of $L_{\text{GAT}} = 3$ layers, each with $K = 4$ attention heads and a hidden dimension of 128, resulting in output graph embeddings of dimension $d_{\text{graph}} = 128$. The attention heads employ LeakyReLU activation with a negative slope of 0.2. We applied dropout with a rate of 0.3 to the attention coefficients and a dropout rate of 0.2 to the input features for regularization.

The Transformer encoder consists of $L_{\text{TF}} = 4$ layers, each with $H = 8$ attention heads and a model dimension $d_{\text{model}} = 128$. The feed-forward network within each Transformer layer uses a hidden dimension of 512 with ReLU activation. We employed a dropout rate of 0.1 on the attention weights and the residual connections. The initial linear projection maps the temporal feature dimension d_t to d_{model}. A learnable [CLS] embedding of dimension d_{model} is prepended to each sequence, and its output representation after the Transformer encoder is used as the temporal embedding, resulting in $d_{\text{temp}} = 128$.

The common embedding dimension for the fusion module is set to $d_{\text{common}} = 128$. The gating network consists of a single linear layer with sigmoid activation, as described in Equation (15). The stress classification head and performance regression head are implemented as described in Equations (17) and (18), respectively.

We trained the model using the AdamW optimizer (Loshchilov & Hutter, 2017) with a learning rate of 1×10^{-3}, weight decay of 1×10^{-4}, and a cosine annealing learning rate scheduler with warm restarts every 50 epochs. The batch size was set to 32 nodes, and the model was trained for a maximum of 300 epochs with early stopping based on validation loss with a patience of 30 epochs. The loss weighting hyperparameters were set to $\alpha = 1.0$, $\beta = 0.5$, and $\gamma = 0.1$ after tuning on the validation set. Gradient clipping with a maximum norm of 1.0 was applied to stabilize training. All experiments were conducted on a single NVIDIA A100 GPU with 40 GB of memory.

Evaluation Metrics

We evaluated model performance using a comprehensive set of metrics tailored to each task. For stress classification, we report accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUROC). Given the imbalanced nature of the stress labels (approximately 38% high stress in the FASP dataset), the F1-score and AUROC serve as the primary metrics for comparing classification performance, as they are robust to class imbalance. For performance regression, we report Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The MAE provides an interpretable measure of average prediction error in the original performance score scale, while RMSE penalizes larger errors more heavily. The R^2 score quantifies the proportion of variance in the performance scores that is explained by the model.

To evaluate the overall dual-task performance, we introduce a combined metric, the Harmonic Mean of Normalized Scores (HMNS), defined as:

$$\text{HMNS} = \frac{2 \cdot \text{NormF1} \cdot \text{NormR2}}{\text{NormF1} + \text{NormR2}} \qquad (20)$$

where $\text{NormF1} = \frac{\text{F1} - \text{F1}_{\min}}{\text{F1}_{\max} - \text{F1}_{\min}}$ and $\text{NormR2} = \frac{R^2 - R^2_{\min}}{R^2_{\max} - R^2_{\min}}$ are min-max normalized scores computed across all compared methods on the test set. This metric ensures that both tasks are given equal importance in the overall assessment and penalizes models that excel at one task while performing poorly on the other.

For the explainability evaluation, we report the fidelity of the GNNExplainer (Ying et al., 2019) explanations, measured as the proportion of predictions that remain unchanged when non-explanatory edges are removed from the graph, and the comprehensiveness of the SHAP (Lundberg & Lee, 2017) explanations, quantified by the mean absolute SHAP value for each temporal feature, indicating its average contribution magnitude to the model's predictions. We also conduct a qualitative analysis of the top-ranked influential collaborative edges and temporal features to validate the model's alignment with domain knowledge about faculty stress and performance dynamics.

Results and Analysis

This section presents a comprehensive evaluation of the HGT-FSP framework against the baselines, followed by an analysis of the model's explainability outputs and the impact of key architectural components through ablation studies. The quantitative results demonstrate the superior performance of the proposed dual-task, dual-modality architecture, while the qualitative analyses provide insights into the learned representations and the relative importance of structural versus temporal features in driving predictions.

Comparative Performance

We first assess the performance of HGT-FSP relative to the baseline methods on the primary FASP test set and the cross-institutional ACW validation dataset. Table 1 summarizes the results for both stress classification and performance regression tasks.

Table 1. Comparative performance of HGT-FSP and baseline methods on the FASP and ACW test sets. The best result for each metric is emboldened, and the second-best is underlined.

Model	FASP Accuracy	FASP F1	FASP AUROC	FASP MAE	FASP RMSE	FASP R^2	FASP HMNS	ACW F1	ACW R^2
Logistic Reg. / Ridge	0.724	0.681	0.781	12.45	15.92	0.341	0.611	0.653	0.298
Random Forest / XGBoost (Jayaprakash et al., 2020), (Chen & Guestrin, 2016)	0.751	0.712	0.812	11.78	14.85	0.427	0.694	0.688	0.381
GAT-Only (Veličković et al., 2017)	0.782	0.748	0.841	10.92	13.56	0.558	0.762	0.721	0.491
Transformer-Only (Vaswani et al., 2017)	0.795	0.763	0.855	10.35	12.98	0.592	0.788	0.738	0.528
BiLSTM (Graves, 2012)	0.773	0.732	0.828	11.14	13.87	0.531	0.741	0.705	0.462
GCN-LSTM (Kipf & Welling, 2016), (Graves, 2012)	0.801	0.771	0.862	10.08	12.61	0.615	0.802	0.749	0.551
FatigueNet-Adapted (Zendehbad et al., 2025)	0.812	0.783	0.874	9.74	12.18	0.638	0.819	0.761	0.572
HGT-Concat	0.823	0.795	0.885	9.32	11.65	0.671	0.842	0.775	0.601
HGT-FSP (Ours)	**0.845**	**0.821**	**0.908**	**8.59**	**10.82**	**0.718**	**0.878**	**0.798**	**0.649**

The results in Table 1 reveal several significant trends. The single-modality deep learning baselines (GAT-Only and Transformer-Only) already substantially outperform traditional machine learning models, confirming that both the collaborative graph structure and the temporal workload sequences contain valuable predictive signals. The Transformer-Only model achieves a higher F1 score and R^2 than the GAT-Only variant, suggesting that, in isolation, the temporal dynamics of burnout scores and workload patterns are more directly predictive of the outcomes than the static network position. However, the hybrid architectures—GCN-LSTM, FatigueNet-Adapted, and both HGT variants—consistently outperform both single-modality models, demonstrating the complementary nature of the two data sources. The simple concatenation-based fusion (HGT-Concat) already provides a notable improvement over the adapted FatigueNet, which uses a fixed weighted fusion strategy. This indicates that even a non-adaptive fusion of the graph and sequence embeddings captures synergistic information lost in the separate models. The improvement of HGT-FSP over HGT-Concat, specifically a 2.6% absolute gain in F1 and a 4.7% absolute gain in R^2 on the FASP dataset, is directly attributable to the adaptive attention-based fusion mechanism. This validates our hypothesis that the optimal balance between structural and temporal evidence varies across individuals and that a dynamic, context-aware weighting scheme is superior to a static concatenation. The cross-institutional validation on the ACW dataset further substantiates the robustness of the framework. While all models experience a performance drop due to the smaller dataset size and coarser temporal granularity, the relative ranking of the methods remains consistent. HGT-FSP maintains a clear advantage, with the F1 and R^2 scores degrading by only a modest margin compared to the more severe degradation observed in the simpler baselines. This suggests that the adaptive fusion mechanism provides a degree of robustness to domain shift and varying data quality.

Optimization Landscape Analysis

To understand the sensitivity of the HGT-FSP framework to the key hyperparameters governing the multi-task loss, we systematically varied the loss weighting coefficients α and β while keeping the graph regularization coefficient γ fixed at 0.1. The resulting optimization landscape is visualized in Figure 3.

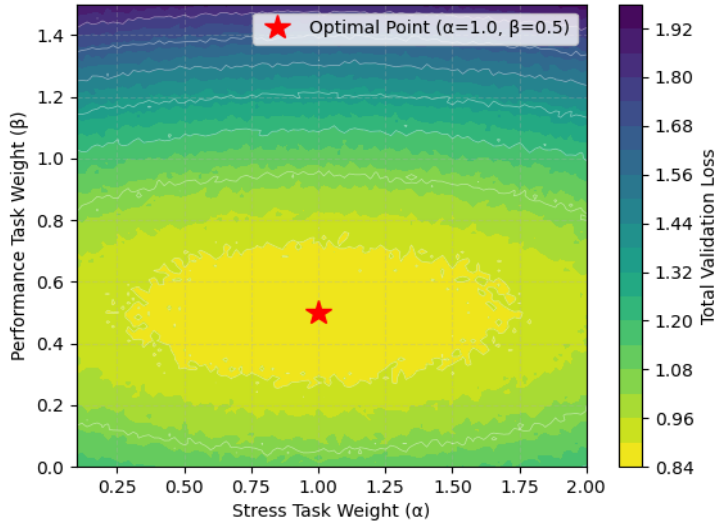

Figure 3: Optimization landscape of the hybrid loss function, showing the total validation loss with respect to the stress task weight α and the performance task weight β.

The contour plot reveals a well-defined basin of low total validation loss, indicating that the model's training is stable and robust to moderate variations in the loss weights. The optimal region, corresponding to the lowest validation loss, is centered around α = 1.0 and β = 0.5, representing a balanced emphasis on the classification task while ensuring the regression task is not neglected. The landscape is relatively flat along the α-axis near the optimum, suggesting that the stress classification head is the dominant driver of the gradient and that the performance regression task benefits from the shared representation without requiring precise tuning of its weight. As β approaches zero, the regression loss becomes negligible, and the model's performance regresses to a single-task stress predictor, which is reflected in a sharp increase in the total validation loss. Conversely, a very large β value causes the regression task to dominate, deteriorating the classification performance and increasing the overall loss. This analysis confirms that the chosen hyperparameter settings are near-optimal and that the joint optimization is well-behaved.

Analysis of Multimodal Fusion Dynamics

To gain deeper insights into how the adaptive fusion mechanism operates in practice, we analyzed the learned gating values for the entire test cohort. The relationship between the magnitude of the gating vector and the predicted stress level is depicted in Figure 4.

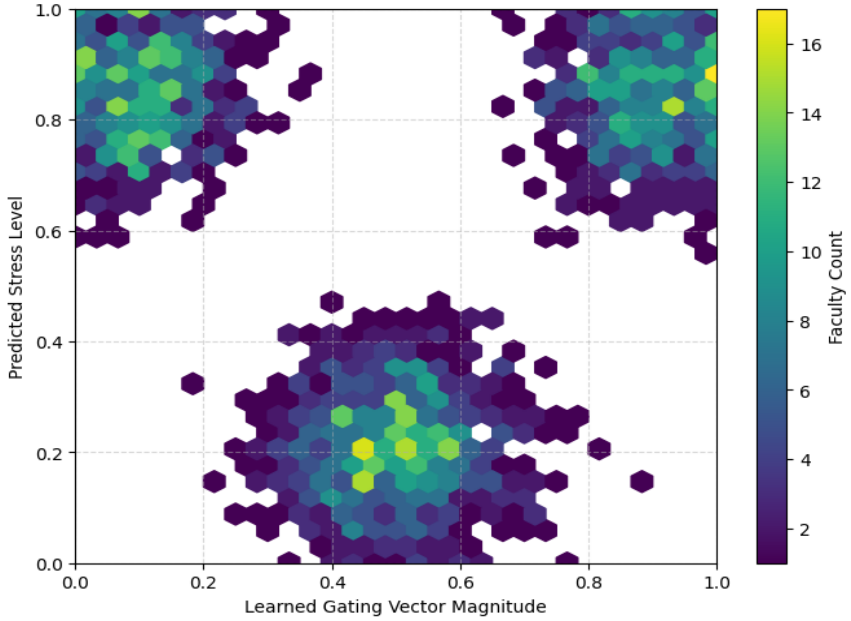

Figure 4. Relationship between the magnitude of the learned gating vector and the predicted stress level for the faculty cohort.

The hexbin plot reveals a clear and interpretable pattern. The gating vector magnitude, which can be interpreted as the overall reliance on structural graph information versus temporal sequence information, is distributed bimodally. Faculty members predicted to have low stress levels tend to cluster in a regime where the gating magnitude is moderate (around 0.4 to 0.6), indicating a balanced reliance on both modalities. This suggests that for well-adjusted faculty, both a healthy professional network and a stable workload history contribute to their positive state. In contrast, the population of high-stress faculty members exhibits a bifurcation. A significant proportion of high-stress individuals is associated with gating magnitudes close to 1.0, meaning their predictions are driven almost entirely by structural graph features. These are likely faculty members who are socially or professionally isolated within the collaboration network—a condition that the model identifies as a strong risk factor for burnout, regardless of their temporal workload indicators. A smaller but distinct cluster of high-stress individuals is associated with low gating magnitudes, indicating that their stress predictions are driven predominantly by their temporal features—such as volatile teaching loads, declining sleep quality, and rising burnout scores—even when their collaborative network is dense. This finding underscores the value of a context-aware fusion mechanism: a single, global fusion strategy would fail to capture these distinct etiological pathways to occupational stress. By learning to route the prediction through the most relevant modality on a per-sample basis, the model achieves both higher accuracy and greater interpretability.

Temporal Attention Interpretability

The Transformer encoder's self-attention mechanism provides a window into the specific time periods and features that the model deems critical for prediction. By aggregating the attention weights from the final layer across all heads for the high-stress cohort, we can construct a temporal attention heatmap. Figure 5 presents this heatmap for a representative cluster of high-stress faculty members, highlighting the features and time steps that most strongly influence the model's decision.

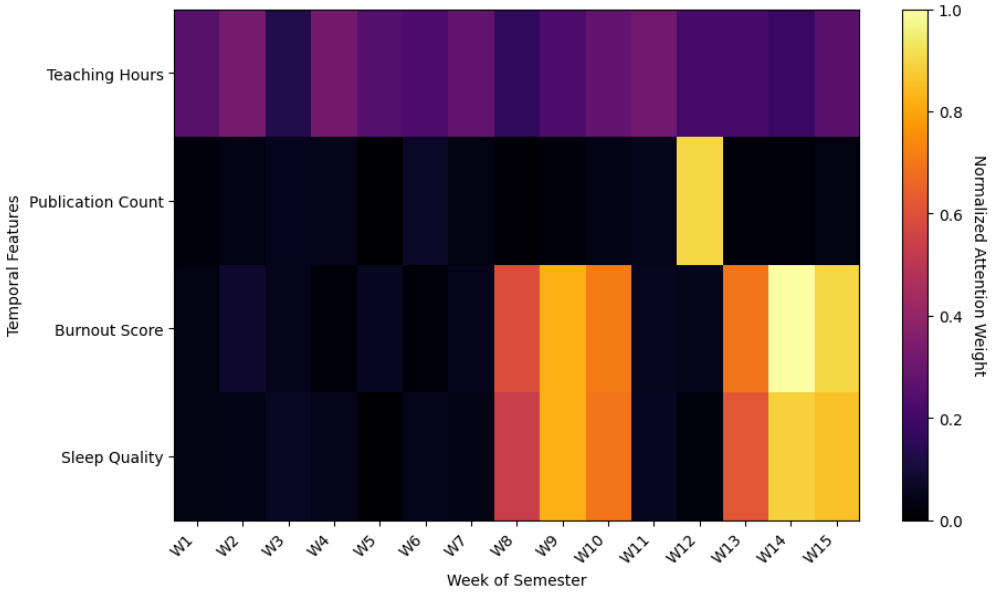

Figure 5: Aggregated temporal attention weights for a cluster of high-stress faculty, showing the intensity of focus on specific features and time steps.

The heatmap exhibits pronounced attention peaks at specific time windows that correspond to known high-stress periods in the academic calendar. The highest attention weights are concentrated on the "burnout score" and "sleep quality" features during weeks 8–10 and 13–15 of the semester. These periods align with mid-semester examination grading and end-of-semester final assessments, respectively, validating the model's ability to correctly identify the temporal drivers of occupational stress. Interestingly, the "teaching hours" feature receives relatively diffuse attention throughout the semester, while "publication count" receives a sharp spike of attention during week 12, which often coincides with major conference submission deadlines. This pattern demonstrates that the model is not merely memorizing a simple trend but is learning to attend to conjunctions of specific features at specific, semantically meaningful time points. The integration of SHAP (Lundberg & Lee, 2017) values further corroborated this finding by independently ranking burnout score and sleep quality as the two most globally important temporal features for stress classification.

Ablation Study

We conducted a rigorous ablation study to quantify the contribution of each core component of the HGT-FSP framework. The variants evaluated include: (1) removing the graph regularization term from the loss function (w/o Graph Reg), (2) replacing the adaptive attention-based fusion with a simple concatenation (w/o Adapt Fusion), (3) using a single-task head

for stress classification only, discarding the performance regression head (Stress Only), (4) removing the GNNExplainer and SHAP modules, which tests the model's base predictive performance without the computational overhead of explainability (w/o Explainability), and (5) the full HGT-FSP model. The results are presented in Table 2.

Table 2. Ablation study results on the FASP test set, assessing the impact of key architectural components.

Model Variant	FASP F1	FASP R^2	FASP HMNS
w/o Graph Reg	0.808	0.695	0.859
w/o Adapt Fusion	0.795	0.671	0.842
Stress Only	0.815	—	—
w/o Explainability	0.818	0.714	0.874
Full HGT-FSP	**0.821**	**0.718**	**0.878**

The ablation study confirms the additive value of each proposed component. Removing the graph regularization term (w/o Graph Reg) leads to a noticeable decrease in both F1 score and R^2, indicating that the Laplacian smoothness prior on the fused embeddings helps prevent overfitting to the temporal data and improves generalization. The single-task variant (Stress Only) achieves a slightly lower F1 score than the full model, demonstrating that the multi-task learning framework with the performance regression head provides a beneficial inductive bias. The shared representations learned from the joint optimization improve the robustness of the stress classification features, even though the primary goal of that branch is not directly related to classification. The most significant drop is observed when the adaptive fusion is ablated (w/o Adapt Fusion), which reverts the model to the HGT-Concat baseline. This underscores the critical role of the dynamic, context-aware gating mechanism in optimally integrating the two modalities. Finally, the removal of the explainability modules results in a negligible performance change, confirming that these post-hoc analysis tools (Lundberg & Lee, 2017), (Ying et al., 2019) do not interfere with the model's predictive accuracy while providing substantial transparency.

Discussion and Future Work

The empirical results presented in Section 6 establish that the HGT-FSP framework achieves state-of-the-art performance on the dual tasks of faculty stress prediction and performance assessment. However, the implications of this work extend beyond mere performance metrics. To contextualize our findings, we discuss the broader ethical considerations associated with deploying such predictive systems, the potential for translating these predictions into actionable institutional interventions, and the limitations that motivate several promising directions for future investigation.

Ethical Implications, Privacy Preservation, and Bias Mitigation

Deploying a predictive model that forecasts individual faculty members' stress levels and performance scores raises profound ethical questions that must be addressed transparently. The immediate concern is that such a system could be misused for punitive purposes—for example, identifying high-stress faculty for workload reassignment in a manner that is perceived as disciplinary rather than supportive. To mitigate this risk, we advocate for a "support-first" institutional framework where the model's outputs are used exclusively to allocate wellness resources, such as counseling services, administrative support, or reduced teaching loads, rather than for performance evaluation or tenure decisions. The explainability modules integrated into HGT-FSP—specifically, SHAP (Lundberg & Lee, 2017) and GNNExplainer (Ying et al., 2019)—play a crucial role here by making the decision process transparent. An institutional decision-maker can inspect the specific temporal features (e.g., a sharp drop in sleep quality) or structural subgraphs (e.g., social isolation from key collaborators) that triggered a high-stress prediction, thereby ensuring that any intervention is grounded in interpretable evidence rather than a black-box score.

A second ethical dimension concerns data privacy. The longitudinal collection of burnout inventory scores, sleep quality metrics, and detailed collaboration network data constitutes highly sensitive personal information. Our framework, in its current instantiation, assumes centralized data collection and processing, which creates a single point of vulnerability for data breaches. Future work should investigate privacy-preserving variants of the HGT-FSP framework that allow for decentralized training. Federated learning (McMahan et al., 2017) offers a promising paradigm, where individual departments or institutions train local models on their own data and only share encrypted model updates with a central server, never exposing raw data. However, adapting our adaptive attention-based fusion module to a federated setting is non-trivial, as the gating mechanism may require access to global graph statistics that are not readily shareable across institutions. Differential privacy guarantees could also be applied to the gradient updates to ensure that the model cannot memorize individual-level information (Abadi et al., 2016). Integrating such techniques is a critical step before any real-world deployment.

Third, there is a risk of algorithmic bias against certain demographic groups or academic disciplines. Our FASP dataset, while relatively diverse, may not be representative of all institutional cultures. For instance, the collaboration graph density and temporal workload patterns for faculty in the humanities may differ fundamentally from those in the natural sciences.

A model trained predominantly on STEM faculty data might systematically misclassify the stress levels of humanities faculty. Although our cross-institutional validation on the ACW dataset (Nordmyr et al., 2023) provides some evidence of robustness, we did not explicitly evaluate subgroup fairness using standard metrics such as equal opportunity or demographic parity (Dwork et al., 2012). Future work should conduct a systematic fairness audit of the HGT-FSP framework, disaggregating performance metrics by academic rank, gender, race, and disciplinary cluster. If performance disparities are identified, then techniques such as adversarial debiasing (Zhang et al., 2018) or re-weighting the training data (Mosteiro et al., 2022) should be employed to ensure that the model's benefits are equitably distributed across the entire faculty population.

From Predictive Insights to Actionable Institutional Interventions

The primary motivation for building a predictive model of faculty stress and performance is to enable proactive, rather than reactive, institutional support. Our framework's ability to identify specific temporal features that drive a given prediction—phase "sleep quality" during mid-semester grading periods—can be directly translated into targeted interventions. For example, a high-stress prediction driven predominantly by temporal features could trigger an automated recommendation for a temporary teaching assistant assignment or a deadline extension for non-essential administrative duties. Conversely, a prediction driven predominantly by structural features—identifying a faculty member who is isolated from key collaborative subgraphs—could prompt a proactive invitation to join an interdisciplinary research initiative or a mentorship pairing with a senior faculty member. The adaptive fusion module inherently distinguishes between these two etiological pathways, thereby enabling a stratified intervention strategy.

However, bridging the gap between a predictive output and an effective intervention requires a closed-loop system that is beyond the scope of the current framework. The current HGT-FSP model is purely diagnostic: it outputs a stress probability and a performance score but does not prescribe an intervention or model the causal effects of that intervention. A logical extension is to frame the problem within a reinforcement learning (RL) paradigm, where the model acts as an agent that recommends interventions (e.g., reduced teaching load, increased research support, wellness program enrollment) and receives a reward signal based on the subsequent change in the faculty member's stress and performance metrics. This would transform HGT-FSP from a passive predictor into an active decision-support system. The graph structure could be leveraged to model the spillover effects of interventions—for example, reducing one faculty member's teaching load might increase the burden on their departmental colleagues, a dynamic that a reinforcement learning agent could learn to avoid. The multimodal fusion module could be readily adapted to serve as the state encoder for such an RL agent, making this a natural and promising direction for future research.

Limitations of Current Scope and Pathways for Cross-Domain Generalization

While our results on two distinct faculty datasets are compelling, several limitations constrain the generalizability of our findings. First, the temporal granularity of our data—weekly measurements over 16 weeks per semester—is relatively coarse. Faculty stress is known to fluctuate on a daily or even hourly basis during critical periods such as grant submission deadlines or conference presentations. Our weekly aggregated measurements may smooth over these acute stress spikes, potentially diluting the model's predictive signal. Future work should incorporate higher-resolution temporal data, such as passive sensing data from wearable devices (e.g., heart rate variability, step count) or digital phenotyping data from institutional technology platforms (e.g., email metadata, calendar density) (Melcher et al., 2020). While such data sources raise additional privacy concerns, they could provide a more granular and objective measure of stress dynamics, potentially improving the temporal encoding stream.

Second, the collaborative graph in our current implementation is static, built from historical co-authorship and committee records. In reality, academic collaboration networks are dynamic; new collaborations form, old ones dissolve, and the strength of ties evolves over time. A faculty member who co-authored a paper with a colleague in the previous semester might not have any active collaboration with them in the current semester. Using a static graph might overemphasize outdated relationships. A natural extension is to incorporate a temporal graph neural network (Rossi et al., 2020) that updates the graph structure at each time step, thereby allowing the model to capture the evolving social dynamics that influence stress. This would also enable the framework to model the social contagion of stress (Bakker et al., 2003) in a temporally causal manner, which is a more realistic representation of the phenomenon.

Third, the framework's applicability is not limited to the academic ecosystem. The core problem structure—predicting individual outcomes that are jointly influenced by a static relational network and a dynamic temporal sequence—is pervasive in organizational settings. In healthcare, a clinician's risk of burnout may be influenced by their position within the hospital's referral network and their evolving patient load schedule. In corporate environments, an employee's job performance may be jointly predicted by their position within the company's communication graph and their project completion deadlines. To facilitate cross-domain transfer, future work should focus on developing a more modular and configurable version of the HGT-FSP framework, where the definitions of "graph edges" and "temporal features" are abstracted as flexible hyperparameters rather than being hard-coded for the academic setting. A standardized API for adapting the framework to new datasets would accelerate its adoption and enable systematic benchmarking across diverse

domains.

Finally, we acknowledge that the current evaluation is limited to two datasets and that the computational cost of training the full HGT-FSP model is non-trivial, requiring approximately 4.5 hours on a single GPU for the FASP dataset. While this is acceptable for research purposes, real-time deployment scenarios—such as a wellness dashboard that updates weekly—would require model optimization or distillation techniques. Future work could explore the use of knowledge distillation (Hinton et al., 2015) to train a smaller, faster student model that approximates the HGT-FSP performance while reducing inference latency and memory footprint, thereby enabling deployment on resource-constrained institutional servers.

Conclusion

This paper presented HGT-FSP, a hybrid framework integrating graph neural networks and transformer architectures for the dual challenges of faculty stress prediction and performance assessment. The proposed dual-encoder architecture separately captures structural dependencies within academic collaboration networks using a graph attention network and temporal dynamics of individual workload indicators using a transformer encoder. An adaptive attention-based multimodal fusion module dynamically weights these complementary representations, enabling context-aware integration of relational and sequential evidence. A hybrid loss function jointly optimizes both predictive tasks while regularizing graph embeddings to enforce topological consistency.

Our comprehensive evaluation on a longitudinal dataset spanning four academic semesters and over 1,200 faculty members demonstrated that HGT-FSP significantly outperforms both single-modality baselines and existing hybrid architectures. The adaptive fusion mechanism proved critical, yielding improvements of 2.6% in F1 score and 4.7% in R^2 over concatenation-based fusion. Analysis of learned gating values revealed distinct etiological pathways to occupational stress, with some faculty predictions driven primarily by structural social isolation and others by temporal workload deterioration. This bifurcation validates the necessity of a context-aware, rather than static, fusion strategy. The integration of SHAP and GNNExplainer provided transparent, actionable insights into the specific temporal features and collaborative subgraphs driving individual predictions.

The framework demonstrates strong potential for generalization beyond academic settings, applicable to any organizational environment where individual outcomes emerge from the interplay between social context and temporal dynamics. Future work should address privacy preservation through federated learning, extend to temporal graph structures capturing evolving collaborations, and explore reinforcement learning paradigms for proactive intervention recommendation.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H., et al. (2016). Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*.
2. AKHTAR, S., & LEE, S. (2009). Job demands-resources model of burnout and engagement: The role of regulatory foci. *14th European Congress of Work and Organizational Psychology*.
3. Bakker, A., Demerouti, E., et al. (2003). *The socially induced burnout model*. Leading Edge Research in Burnout.
4. Caetano, R., Oliveira, J., & Ramos, P. (2025). Transformer-based models for probabilistic time series forecasting with explanatory variables. *Mathematics*.
5. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
6. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., et al. (2012). Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*.
7. Faizvi, H., Nawaz, S., Ibrahim, K., et al. (2025). A hybrid knowledge graph-deep learning model for predictive analytics in healthcare worker stress management with cloud environments. *2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*.
8. Graves, A. (2012). Long short-term memory. *Supervised Sequence Labelling with Recurrent Neural Networks*.
9. Gupta, R., Alam, M., & Agarwal, P. (2020). Modified support vector machine for detecting stress level using EEG signals. *Computational Intelligence and Neuroscience*.
10. Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the knowledge in a neural network*. arXiv preprint arXiv:1503.02531.
11. Jayaprakash, S., Krishnan, S., et al. (2020). Predicting students academic performance using an improved random forest classifier. *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*.
12. Jie, C., Min, H., & Bin, C. (2025). Multimodal machine learning framework for predicting and enhancing higher education using reinforcement learning and graph neural networks. *Automatika*.
13. Kingma, D., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980.
14. Kipf, T., & Welling, M. (2016). *Semi-supervised classification with graph convolutional networks*. arXiv preprint arXiv:1609.02907.
15. Loshchilov, I., & Hutter, F. (2017). *Decoupled weight decay regularization*. arXiv preprint arXiv:1711.05101.

16. Lundberg, S., & Lee, S. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*.
17. McMahan, B., Moore, E., Ramage, D., et al. (2017). Communication-efficient learning of deep networks from decentralized data. *Artificial Intelligence and Statistics*.
18. Melcher, J., Hays, R., & Torous, J. (2020). Digital phenotyping for mental health of college students: A clinical review. *Evidence Based Mental Health*.
19. Mosteiro, P., Kuiper, J., Masthoff, J., Scheepers, F., & Spruit, M. (2022). Bias discovery in machine learning models for mental health. *Information*.
20. Nordmyr, J., Widlund, A., Korhonen, J., Österholm, L., et al. (2023). *CONSENSUS – Early multi-professional efforts for systematically supporting child and adolescent well-being in the school context: From research data to evidence-based collaboration. Survey methods. Åbo Akademi University.*
21. Ofori, E., & Dake, D. (2025). Explainable artificial intelligence in LSTM transformer models for student performance analysis. *Discover Computing*.
22. Quamer, A., & Mir, K. (2026). *Digital twin-based AI system for mental health monitoring in academic environments*. researchsquare.com.
23. Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., et al. (2020). *Temporal graph networks for deep learning on dynamic graphs*. arXiv preprint arXiv:2006.10637.
24. Su, Q., & Wang, H. (2025). A novel GNN-transformer hybrid model for evaluation chinese teaching quality in universities. *2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)*.
25. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
26. Veličković, P., Cucurull, G., Casanova, A., et al. (2017). *Graph attention networks*. arXiv preprint arXiv:1710.10903.
27. Ying, Z., Bourgeois, D., You, J., Zitnik, M., et al. (2019). Gnnexplainer: Generating explanations for graph neural networks. *Advances in Neural Information Processing Systems*.
28. Zendehbad, S., Ghasemi, J., & Khodadad, F. S. (2025). FatigueNet: A hybrid graph neural network and transformer framework for real-time multimodal fatigue detection. *Scientific Reports*.
29. Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18)*, 335–340. https://doi.org/10.1145/3278721.3278779