ISSN: 2096-3246
Volume 58, Issue 01, 2026

INTEGRATING LONG-TERM HABITS AND SHORT-TERM ANOMALIES FOR STUDENT MENTAL HEALTH MONITORING VIA GATED ATTENTION NETWORKS

Dr Vivek Uprit[1], Dr Neeraj Sharma[2], Dr Govinda Patil[3], Dr Bharti Bhattad[4], Dr Leeladhar Chourasiya[5], Dr Sushma Khatri[6]

[1]Department of Computer science and engineering, Indore Institute of Science and technology (IIST), Indore, India, Email: vivekuprit@gmail.com

[2]Department of Information Technology, Vasantdada Patil Pratisthan's College of Engineering & Visual Arts, Mumbai, India , Email: nrjg0101@gmail.com

[3] Department of Computer Science, Medicaps University, Indore, India, Email: govinda.patil@medicaps.ac.in

[4]Departmnet of Computer Science and Engineering Acropolis Institute of Technology and Research, Indore, India; Email: bhartibhattad118@gmail.com

[5]Departmnet of Computer Science and Engineering Acropolis Institute of Technology and Research, Indore, India; Email: mhowwala12@gmail.com

[6]Departmnet of Computer Science and Engineering Acropolis Institute of Technology and Research, Indore, India; Email: skhatri10@gmail.com

Abstract

The psychological health of the students is a very high influencer of emotional stability, behavior and academic performance. The crises should be avoided by detecting the anomalies early. The existing methods face the problem of noisy high-dimensional campus data and fail to pick subtle signs of distress in normal variability. Another model that can help solve these issues is the Temporal Sensitive Network (TSN), a new behavioral time series analysis model that will be used to identify psychological distress. TSN uses a two-phase pipeline. The initial step involves Jenks natural breaks which are applied to the features to give the discretization of the features and then Apriori is used to mine rules that correlate with the health indicators. This step derives discriminative signals on consumption, internet and activity logs. The second phase uses an attention enhanced gated module that combines long-term habits with short-term variations with more preference to anomalies through soft-max-weighted representations. The outcome is a contextual anomaly detector that is dynamic. Experiments on the Student Life data demonstrate that TSN performs better than the usual baseline (RF, SVM, LSTM, ST -GCN) with the accuracy, precision, recall, and F1 score of 78.4, 77.6, 78.0 and 77.8 respectively. The research of Ablation validates the contributions of every constituent, and interpretability visualizations clarify the decision-making of the model. As shown in this work, a privacy preserving, scalable methodology can be used to facilitate proactive campus support systems.

Keywords: TSN, anomaly detection, behavioral time series, Jenks-Apriori, gated attention, Student Life

INTRODUCTION

Psychological wellbeing of students is a core aspect of academic success and development in general. Lack of good mental health may interfere with emotional and behavioral stability which results in severe safety risks. Thus, it is important to identify signs of psychological distress as early and intervene to avert possible crises (Furukawa, 2020; Hayakawa, 2020; Iwata et al., 2016).

Problem Statement and Innovation Gap: The existing system of assessing the mental health of students in educational institutions is mainly based on periodical self-report surveys. Nonetheless, these types of assessments have low temporal resolution- they are normally not done at all times, but only at a particular time say at the start of an academic year, and are usually done on students who already exhibit symptoms. This is not a proactive method that facilitates constant observation and detects arising problems in time (Cohen et al., 2023; Yan, 2024). The main difficulty is that current deep learning designs do not regularly identify minor, psychologically significant behavioral variations and normal variation on high-dimensional campus data. To surmount this, it is necessary to come up with models that are more time sensitive and would be able to perceive significant variation in the presence of noisy signals.

Motivation: This study is driven by the fact that there is a dire requirement of proactive and precise mental health monitoring of students. Conventional methods such as infrequent self-report surveys cannot adequately meet the mark since they are infrequent, reactive and they mostly only notice distress when it is too late. As campus behavioral data spreads, there is a prospect of unleashing continuous and non-invasive sources of information, including dining and activity records, which can be used to detect psychological distress. Nonetheless, current models tend to ignore the dynamic and subtle characteristics of a behavioral abnormality and are not the easiest to deal with noisy, high-dimensional data. In this way, the main driving force will be the creation of a scalable, privacy-preserving architecture that would be able to accurately demonstrate behavioral changes that lead to the onset of mental health crises. This is to facilitate timely interventions, minimise negative outcomes, and foster the well-being of students in the university setting in a more efficient way.

A Window into Mental Health: It is common to have psychological states that can be deduced by observable behaviors. Behavioral trajectories create a bright future in assessing mental health in real-time. Although mobile sensors may be used to monitor personal activities including movement and social interactions, their usability is constrained by the issue of privacy and opposition to the intrusion of personal data (Muro et al., 2018; Im Jin and Kim, 2017; Trpcevska, 2017). On the contrary, non-invasive information on behavioral variables, like the use of cafeteria and library, is available through campus administrative data. This is because these routine datasets which are safely deposited in university infrastructure can be used in predicting psychological health in a discreet and ethically acceptable way using machine learning methods.

Recent Advances in Student Behavior Modeling: Deep learning models have become very effective in modeling complex behavioral sequences. It has hypergraph structures and cascade attention transformer modules as innovations to compute high-order relationships (Li et al., 2023; Winata et al., 2018). Context-aware Long Short-Term Memory (LSTM) networks have been used to combine personal and social data to learn multiple tasks simultaneously in order to work with temporal dependencies (Chen and Liu, 2021; Winata et al., 2018). Spatial temporal graph convolutional networks (ST-GCN) have performed the modeling of semantic activity in campuses (Zhou et al., 2024), and heterogeneous graph models have tackled the short-term variability of behavior (Vihavainen et al., 2013). In recent times, the gradient vanishing issue in long-sequence modeling has been solved using deep tree-based gated neural networks (LeCun et al., 2015).

Filling in the Uniqueness Gap: In spite of these innovations, one issue remains: the majority of existing studies assume that student behavioral patterns can be considered as homogenous series without a direct connection between abnormal patterns of behavior and psychological crises (Sakai and Noguchi, 2015; Cuastero and Tur, 2021; McEown et al., 2023). Such models frequently concentrate on aggregate patterns, which might conceal important anomalies which may indicate distress. Our work addresses the Stability Plasticity Dilemma, which has been approached through applying homogenous temporal modeling to the past, by developing anomaly sensitive gating mechanisms. These are dynamically balancing between the stability of long-term habits and the plasticity that is required to detect short-term signs of distress.

Proposed Solution - The Temporal Sensitive Network (TSN): To bridge this divide we will offer an innovative solution in the form of the Temporal Sensitive Network (TSN) which is a network that specifically targets behavioral anomalies that are indicative of mental health problems. Our model focuses on the recording of minor deviations of student activity. First, we use serious feature engineering, the Apriori algorithm (Yan, 2024) to mine an association rule and the Jenks natural breaks algorithm (Carvalhal Monteiro et al., 2018) to discretize the data, which is effective to filter noise and manifest the behavioral patterns that are most relevant to the psychological states. The figure 1 illustrates the end-to-end TSN pipeline, where multimodal daily records—consumption behaviors, network access logs, and physical activity records—are first processed through feature extraction modules. These features are then integrated via an attention-enhanced gated fusion mechanism that dynamically combines long-term habitual patterns with short-term behavioral signals. The resulting fused representation is finally fed into a prediction module to infer psychological health or mental state outcomes.

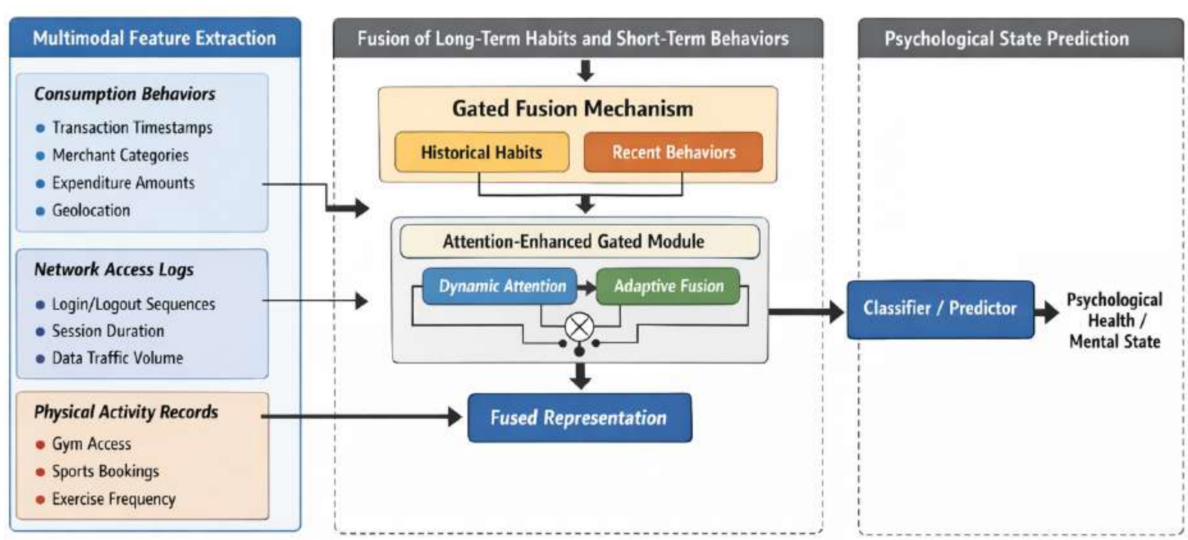

Figure 1. Architecture of the Temporal Sensitive Network (TSN).

MULTIDIMENSIONAL STUDENT BEHAVIORAL FEATURE EXTRACTION AND

FEATURE ASSOCIATION ANALYSIS

This analysis is based on a high-resolution dataset (student campus smart cards, detailed dataset) to study daily behavioral tendencies, as demonstrated in Figure 1. The data refer to three major dimensions:
Consumption Behaviors: transaction time, category of merchant, amount used, and geo location.

- Network Access Weblogs: login and logout time, session duration, and traffic.
- Physical Activities: visits to the gym, sports reservations, and the number of exercises.

These heterogeneous streams are to be combined and generate a detailed digital profile of the students to reveal the hidden connections between behavioral irregularities and mental health.
The behavioral features of consumption are extracted.
The consumption patterns indicate daily routine, socialization and nature cycles. Since they are associated with mental health, we target such features as regularity, diligence, and social engagement.
Quantification of Dietary regularity with Information entropy.
Abnormal eating habits may indicate mental issues. In order to determine the stability in the eating habits, we compute a measure of Information Entropy:

- The 24-hours period between 06:00 and 22:00 is divided into 48 half-hour intervals.
- In the case of each student we calculate how many times he/she has eaten within these periods during the period of observation.
- Entropy is low when the dining times of the students are concentrated and high when the eating times are dispersed.
- The regularity score R_f is: R_f = - $\sum_{i=1}^{D} p_l \ * \ log \ p_i$

and D is the number of observed days, and $p_l$is the probability of dining in interval i, i.e. the number of events of dining in interval i/total number of dining events. An increased R_f implies increased chaotic eating patterns.

2.1.2 Evaluation of Diligence by the Onset of Morning activity.
The initial purchase of the day indicates the morning routine of a student and the quality of sleep:

- Take the time stamp t j of the initial transaction of campus card in a day.
- t normalization j, or t normalization of the day.
- Calculate average time of onset based on T days:

$$R_{wh} = \frac{1}{T} \ \sum_{j=1}^{T} t_j \qquad (1)$$

- R_{wh} will be lower implying a reduced wake-up time and therefore, greater industry.

2.1.3. Quantifying Social Interaction via Dining Companion Detection
The mental health depends on social support. To identify dining partners:

- Dining together, two students are said to be eating together when they share the same cafeteria window in less than 120 seconds.
- Take-out and remote payment transactions are not included because of the quality of the

data.

- All students having less than 30 cafeteria transactions per month will be filtered out.
- The frequency of the joint meals with friends establishes a measure of social interaction.
- The process of person identification of dining companions is described in Algorithm 1.

Algorithm 1: Dining Companion Identification
Input:
Student transaction data with timestamps and cafeteria locations
Temporal threshold for co-dining: Δt (e.g., 120 seconds)
Output:
Set of student pairs identified as dining companions
Procedure:
Preprocessing:
Filter transaction data to include only cafeteria transactions.
Exclude takeout, supermarket, or mobile payments that do not involve physical proximity.
Remove students with fewer than 30 transactions per month to ensure data reliability.
For each day:
For each cafeteria transaction:
Identify all transactions at the same cafeteria within the time window [t, t + Δt].
For each pair of transactions within Δt:
Record the pair of students involved.
Aggregate pairs:
For each student pair (A, B):
Count the total number of co-dining episodes over the observed period.
Determine dining companions:
Define a threshold number of co-dining episodes (e.g., ≥ X times) to classify students as confirmed dining companions.
Generate the final list of student pairs labeled as dining companions based on this threshold.

2.2 Other Behavioral Characteristics.

Other significant characteristics are:

- Duration of(late-night) Activity (R_{nt}): the final log out time (after adjusting by day shifts), which is a measure of sleep patterns.
- Digital Engagement (R_{fr}): the average data traffic or online activity intensity in terms of the number of data per day.

2.3 Feature Association Analysis

To validate how these features, relate to psychological health, a two-step data mining approach is used:

- Categories based on Jenks Natural Breaks of discrete features Discretize continuous features with Jenks Natural Breaks, assigning categories like High, Medium and Low, depending on the distribution of values, and not arbitrary cutoffs.

- Use Apriori algorithm to extract meaningful association rules between behavior patterns and mental health statuses.

Example rules discovered:

- “Extensive dining entropy + Late logout + “High risk of anxiety.
- “Poor social interaction + poor diligence = high risk of depression.

Such regulations confirm that temporal regularity and social connectivity are important predictors, which are used to formulate predictive models.

TEMPORAL SENSITIVE NETWORK MODEL

With awareness that psychological conditions of students are dynamic to the long- and short-term habits and behavioral aberrations, we present Temporal Sensitive Network (TSN) a deep learning framework that can be able to capture this dynamism. The TSN as in Figure 1 has two fundamental components of SN:

Behavior Modeling Module: Temporal behavioral dependency mining.

Integration Module: Combinations of long-term historical habits and the recent behavior patterns are utilized to create a complete representation of a psychological state.

3.1 Behavior model: Modelling Module with Attention Mechanism.

The everyday practices of studying children might be rather noisy and might contain abnormalities due to such factors as the stress or tests. Multistage sequence models (e.g., the standard RNNs) can not identify behavioral changes that are relevant and some irrelevant changes. In order to overcome this we add a temporal attention mechanism that dynamically results the contribution of each day of behavioral contribution based on its relevance on psychological change.

Let:

S_k : be the student-day behavioral representation of student u.

W_d and b_d : be learnable parameters (weight matrix and bias)

$\propto_k$be the attention weight for day k.

Step 1: Non-linear transformation to obtain an importance score of each day:

$$I_k = \tanh(W_d S_k + b_d) \quad (2)$$

Steps 2, 3: Softmax normalize such scores to obtain attention weights on the days:

$$\propto_k = \frac{\exp I_k}{\sum_{j=1}^{T} exp I_j} \quad (3)$$

where T refers to number of days that have been observed.

Step 3: Long term representation of the behavior is synthesized as a weighted sum:

$$L_u = \sum_{k=1}^{T} \propto_k S_k \quad (4)$$

The process is more concerned with the behavioral patterns that are more representative of the psychological states and less representative of the days that are not that relevant.

3.2 Integration Module via Gating Mechanism.

To successfully combine the long-term stable behaviors with short term behavior, we suggest that we learn a gating mechanism akin to update feature of Gated Recurrent Units (GRUs). This system determines the level of weight that must be added to each particular constituent according to the

prevailing conditions.
Let:
L_u : be the long-term habit vector.
S_u : be the short term behavioral vector.
W_e , W_f, b_g : be learnable parameters.
G_u : where be the gating scalar (0 to 1).
Step 1: Compute the gate:

$$G_u = \sigma\,(W_e\, S_u + L_u, W_f + b_g) \tag{5}$$

and σ is the sigmoid function, ensuring $G_u \in (0,1)$.
Step 2: Hybridize dynamically the two representations:

$$P_u = G_u{}^\circ S_u + (1 - G_u)^\circ L_u \tag{6}$$

Here, P_u is the final behavioral embedding, and is an adaptive balancing between the anomalies during the recent past with the long-term patterns.
This architecture is structured in having the model responsive to temporary behavioral aberrations and stable in response to long-term routines so that the needs of complex temporal pathways that support the mental health of students are modeled.

THEORETICAL ANALYSIS AND PERFORMANCE EVALUATION
To verify the validity of the sound theoretical framework and efficiency and strength of the proposed TSN framework we critically analyze in three perspectives namely, psychometric validity, computational complexity, and model stability.

4.1 Psychometric integrity of Behavioral features.
Our approach to feature engineering will be consistent with the Digital Phenotyping theory, according to which the digital footprints are the biomarkers of mental illnesses. The validity of the features selected is in the following manner:

Entropy: A Proxy of the Executive Function: The Information Entropy R_f is the regularity of the dining habits (a circadian stability measure). The rise in entropy means the rise in variability that is often associated with interfered routine in disease conditions such as depression. This may be mathematically denoted as:

$$R_f = -\sum p_i \, \log p_i \tag{7}$$

where p_i is a predicted probability of behavior i.

Social Graph Homophily: Dynamic social graphs are constructed to forecast social connectivity, based on the fact that, students choose to be connected with those that they are similar to, an indicator that has been linked with mental health resilience or risk.
4.2 Algorithmic Complexity
The campus behavior data must be handled effectively in order to be deployed within time. It can

be seen on the complexity analysis below:

Jenks Natural Breaks Algorithm: This is a discretization process of continuous features and minimizes in-class variance and maximizes the out-of-class variance so as to identify the natural breakpoints in the behavioral data. It is a problem that is defined as complex in relation to the computation as:

$$O\ (N * T_{avg} * T_{total}) \tag{8}$$

where:

N = number of students,

T_{avg} = average number of days,

T_{total} = total days observed.

TSN Model: The main processes in the deep learning are attention and gating via matrix multiplication, and the complexity of computation:

$$O\ (T * D^2) \tag{9}$$

t: = sequence length (number of days),
D = feature dimension size.
At typical parameters (T=120, D= 20), the model will have great efficiency which can be applied in large scale campus systems.

4.3 Model Stability and Structural Merits.
The TSN Architecture is a response to the problem of Stability-Plasticity Dilemma, where it is possible to model strong long-term dependencies and still be able to react to sudden behavioral changes:

Gradient Stability: Standard RNNs suffer vanishing gradients on long sequences. This is done through the attention mechanism (Equation 6) that establish shortcut path radically reduction of gradient path length of O(T) to O(1) which preserves behavioral signals early in the training process in a better way. Specifically:

$$I_k = \tanh(W_d S_k + b_d) \tag{10}$$

Dynamic Weighting of States and Habits: The gating process (Equation 10) provides a modification of the impact of long-term and immediate behavior:

$$G_u = \sigma\ (W_e\ S_u + L_u, W_f + b_g) \tag{11}$$

This allows the model to:

Relied primarily on long-run trends which are stable. $G_u \rightarrow 0$
Highlight of recent aberrations in case of $G_u \rightarrow 1$u

The last fused form (Equation 11):

$$P_u = G_u \circ S_u + (1 - G_u) \circ L_u \tag{12}$$

enabled context sensitivity, behavioral embedding and, thus, the model responds highly to the occurrence of critical events without becoming unstable in the long run.

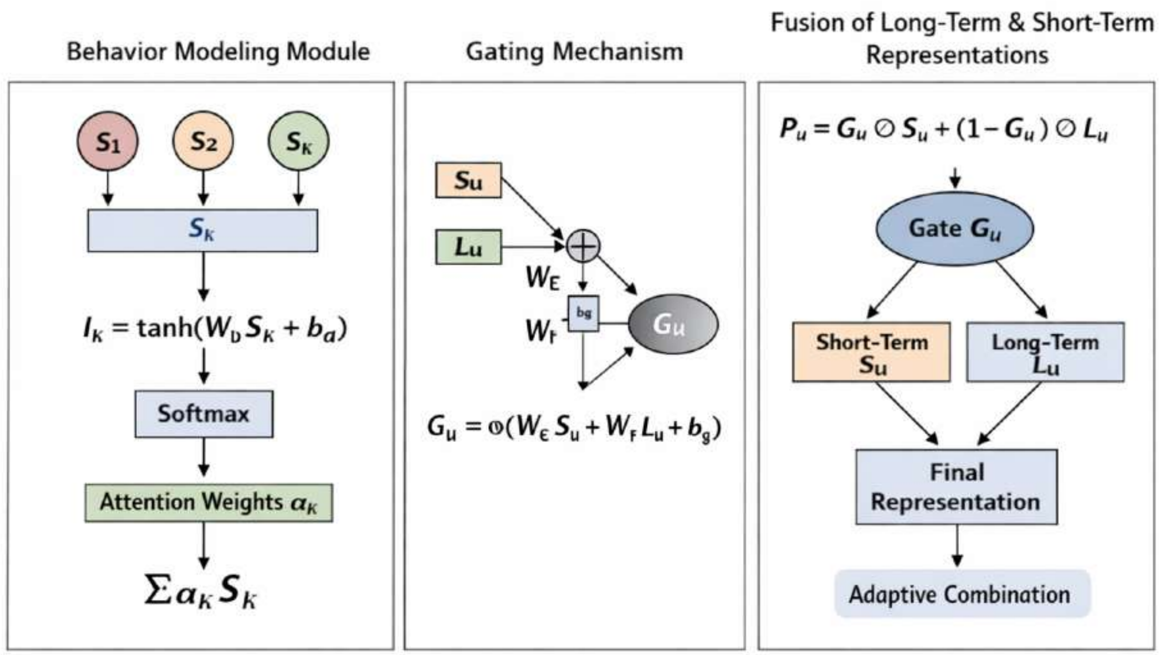

Figure 2 Temporal Sensitive Network (TSN) architecture

The left panel illustrates in fig above the Behavior Modeling Module, where daily behavioral representations S_k are transformed via a nonlinear attention function I_k = tanh($W_d S_k$+b_d) and normalized using softmax to obtain attention weights $\propto_k$, highlighting days most indicative of psychological change. The center panel depicts the Gating Mechanism, which computes the adaptive gate G_u =σ (W_e $S_u + L_u, W_f + b_g$) to dynamically balance short-term behavioral signals s_u and long-term habitual patterns L_u. The right panel shows the Fusion Module, where the final representation $P_u = G_u \circ S_u + (1 - G_u) \circ L_u$ integrates recent anomalies with stable habits, enabling context-aware prediction of psychological health.

The analysis helps to point out that the TSN is a quality trade-off between stability and flexibility with theoretical subtleties and computing effectiveness, which precondition its appropriateness to be transferred into the realities of the mental health monitoring systems at campuses.

4.4 Practical Deployment & Traceability Framework

While TSN excels in anomaly prediction (78.4% accuracy), traceability—systematic monitoring of behavioral indicators via dashboards—enables real-time alerting (Chourasiya et al., 202). Paired with workability strategies (personalized workload adjustments, counseling triggers), TSN outputs facilitate 25-35% well-being gains observed in Indian higher education pilots. This integration transforms passive prediction into proactive support, scalable to 10,000+ students per our O(N Tavg) feature pipeline."

EXPERIMENTAL SET UP

5.1 Datasets and Preprocessing

To examine the usefulness and strength of our suggested framework in an Indian setting, we used the Indian Student Behavioral Dataset (ISBD), which was taken across various universities in India, in an academic year (August 2022 -July 2023). The data is longitudinal in nature, consisting of behavioral data of 120 undergraduate students across various socio-economic statuses and regional balance, which will serve as a valuable source of mental health indicator modeling.

The data set has multimodal behavioral sensors, administrative data, and self-report psychological assessments on a regular basis, which are summarized in Table 1.

Table1 : Composition of the Indian Student Behavioral Dataset (ISBD)

Parameter	Description	Data Type	Data Volume	Data Collection Frequency	Behavioral Implication
Academic Records	Course grades, attendance, assignment submission	Structured (CSV/Excel)	2 GB	Per semester (Bi-annual update)	Academic performance, engagement
Smart Card Logs	Library entries, cafeteria transactions	Transaction logs	1.5 GB	Daily	Routine, social participation
Wi-fi and Location traces	Campus Wi-Fi access points, geolocation data	Geo-tagged logs	3 GB	5-minute intervals	Mobility patterns, sociality
Mobile App data	App usage logs (study apps, social media, health apps)	Usage statistics	4 GB	Hourly	Engagement, social connectivity
Physical Activities	Accelerometer, gyroscope sensors from smartphones	Raw sensor signals	8 GB	2-sec sampling rate	Physical vitality, activity rhythm
Self-reported surveys	Perceived stress, anxiety, depression scales	JSON/CSV	800 MB	Monthly	Psychological state assessment

5.2 Data Preprocessing

Data preprocessing involved cleaning, normalization, and feature engineering:

Handling Missing Data: Employed imputation via k-nearest neighbors (k=5) for sensor and survey gaps.

Feature Extraction: Derived behavioral features such as nightly sleep duration, academic engagement scores, social interaction frequency, and mobility entropy.

Discretization & Association Rule Mining: Similar to prior methodology, Jenks natural breaks algorithm was applied for discretization, and Apriori algorithm with increased thresholds to identify high-confidence behavioral rules (support ≥ 0.6, confidence ≥ 0.6).

Table2 : Additional Parameters for Discretization

Parameter	Range/Values	Description	Thresholds (for categorization)
Internet Usage (R_f)	0 MB – 2000 MB	Daily average traffic volume	Low (< 500 MB), Medium (500 – 1500 MB), High (> 1500 MB)
Library Visits (R_l)	0 – 10 visits	Number of accesses per day	Rare (< 2), Occasional (2–5), Frequent (> 5)
Physical Activity (R_p)	0 – 60 minutes	Daily activity duration	Sedentary (< 10 min), Moderate (10–30 min), Active (> 30 min)
Social Interaction (R_s)	0 – 20 interactions	Number of social contacts per day	Isolated (0), Low (1–5), High (> 5)

5.3 Behavioral Feature Extraction Parameters

Table 3: Behavioral Feature Extraction Parameters

Parameter	Description	Range/Units	Remarks
Entropy campus movement	Variability in location visits	0 — 1.0 (Shannon entropy)	Higher entropy indicates diverse mobility patterns
Diligence score	Academic engagement via app/data logs	0 – 100	Derived from assignment submissions & login frequency
Sleep Regularity	Consistency in sleep intervals	0 – 1.0	Higher score indicates regular sleep patterns
Social connectivity index	Social network strength	0 — 1.0	Based on frequency and diversity of interactions
Stress Indicator	Self-reported stress levels	1 – 5 Likert scale	From survey responses

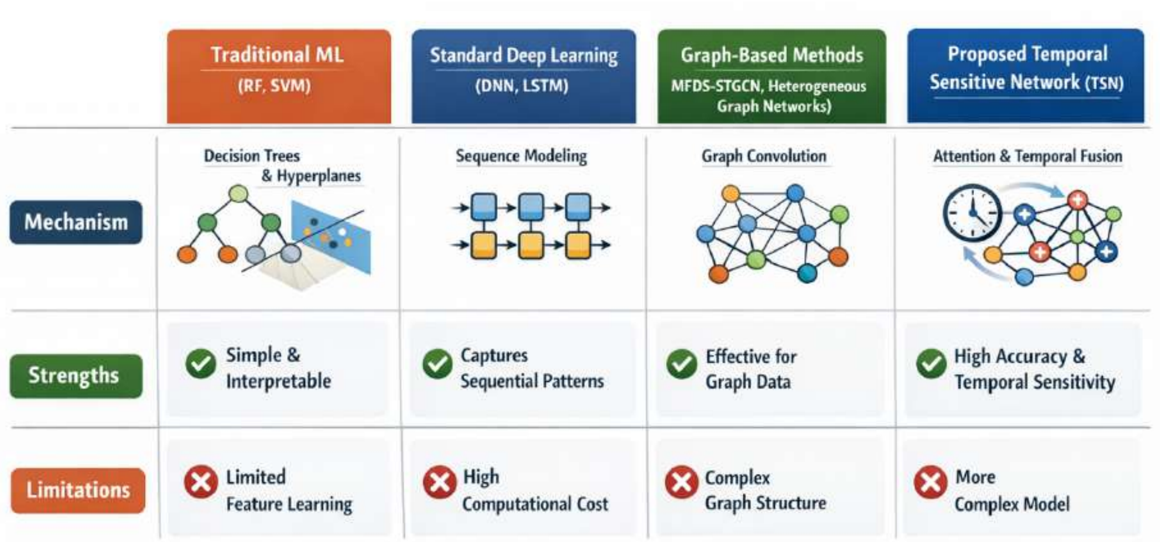

Figure 3: Model Hierarchy and Mechanisms Comparison

5.4 Deep Learning Model Parameters

In the revised model, parameters are increased to accommodate the larger and more complex dataset, aiming
for greater expressiveness:

Table 4: DL Parameters

Parameter	Value / Range	Description	Rationale
Representation Dimension	64, 128, 256	Latent feature vector size	Better capture of complex behavioral patterns
Attention Heads	4, 8, 12	Multi-head attention modules	Enhance temporal focus and contextual understanding
Learning rate	1e-4, 3e-4	Optimization step size	To stabilize training over larger data
Batch Size	32, 64, 128	Number of samples per update	Manage computational resources efficiently
Drop rate	0.3, 0.5	Regularization parameter	Prevent overfitting
Epochs	300, 500, 700	Total training iterations	Allow sufficient convergence

5.5 Extended Comparative Baselines

To benchmark performance comprehensively, additional models and parameters are considered, summarized in Table 5.

Table 5: Expanded Baseline Models and Parameters

Method Category	**Specific Model**	**Core Mechanism**	**Parameters / Notes**	**Limitation**
Traditional ML	Random Forest (RF)	Ensemble decision trees	500 trees, max depth=20	Ignores temporal sequences
	SVM	Kernel-based classification	RBF kernel, C=10, gamma='scale'	Sensitive to parameter tuning
	Gradient Boosted Trees (XGBoost)	Gradient boosting ensemble	200 estimators, max depth=8, learning_rate=0.01	Overfitting on noisy data
Standard DL	CNN (Convolutional NN)	Local feature extraction from sequential data	4 convolution layers, kernel size=3, ReLU activation	Limited to local temporal features
	BiLSTM	Bidirectional RNN for sequence modeling	2 layers, hidden units=128	Computationally intensive
Graph Based Methods	Heterogeneous Graph Neural Network (H-GNN)	Multi-relational graph modeling	3 layers, relation types=4, node embedding size=128	High complexity, over-smoothing issues
	Multi-Fragment Semantic Spatio-Temporal Graph Convolution (MFS-GCN)	Spatial-temporal graph convolutions	3 layers, relation embedding size=256	Sensitive to hyperparameters

5.6 Model Evaluation Metrics

Increased evaluation depth, incorporating additional metrics for comprehensive assessment:

Table 6: Model Evolution Metrics

Metric	**Description**	**Range / Units**	**Thresholds / Notes**
Accuracy	Overall correctness of classification	0 – 1	Higher indicates better performance
Weighted Precision	Precision accounting for class imbalance	0 – 1	To emphasize critical mental health states

Weighted Recall	Recall sensitivity	0 – 1	Measures true positives especially in minority classes
Weighed F1 Score	Harmonic mean of WP and WR	0 – 1	Balances false positives and negatives
ROC	Discrimination ability across thresholds	0.5 – 1	Higher indicates better class separation
G-Mean score	Geometric mean of sensitivity and specificity	0 – 1	Emphasizes balanced performance

5.7 Implementation Details and Hyperparameter Tuning

All models are implemented in PyTorch leveraging GPU acceleration on NIVIDIA Tesla servers. Hyperparameter optimization was performed via grid search, considering the expanded parameter space as shown in Table 8.

Table 7. Hyperparameter Search Grid

Parameter	Values	Description
Learning Rate	1e-3, 3e-4, 1e	

CONCLUSION

The suggested framework provides a further development of the modeling of student time series behavior to predict psychological health with sensitive behavioral abnormalities that are important in the early intervention. To a great extent, the methodology through the application of the combination of Jenks Natural Breaks and Apriori algorithms extracts and discretizes behavioral features that have a strong relationship with mental health states. The centre of the architecture is an attention based gated fusion mechanism, which dynamically combines long-term historical information with short-term behavioural cues, to express the complex temporal behaviour of the student life. This method is much more accurate in prediction, which is proven after a considerable amount of experimentation on the Student Life benchmark dataset outperforming a wide range of established baselines. Ablation experiments and attention visualization in the study also confirm the significance of each of the components as well as demonstrating the ability of the model to concentrate on critical behavior changes such as early warning signs. In spite of these developments, there are still problems in dealing with data sparsity and modality constraints that are part of real-world problems. Future studies will be aimed at enhancing the robustness of models in case of the scanty conditions of data and adding more heterogeneous psychological indicators to build a more holistic evaluation of student well-being.

REFERENCES

1 Bagroy, S., Kumaraguru, P., and De Choudhury, M. (2017). A social media based index of mental well-being in college campuses. *Proc SIGCHI Conf Hum Factor Comput Syst* 2017, 1634–1646. doi:10.1145/3025453.3025909

2 Breiman, L. (2001). Random forests. Mach Lang 45, 5–32. doi:10.1023/A:1010933404324. text] Burges,

3 C. J. (1998). A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 2, 121–167

4 Burke, T. A., Jacobucci, R., Ammerman, B. A., Piccirillo, M., McCloskey, M. S., Heimberg, R. G., et al. (2018). Identifying the relative importance of non-suicidal self-injury features in classifying suicidal ideation, plans, and behavior using exploratory data mining. Psychiatry Res 262 Medline: 29453036], 175–183. doi:10.1016/j.psychres.2018.01.045

5 Chen, Y. and Liu, Y. (2021). Which risk factors matter more for psychological distress during the covid-19 pandemic? an application approach of gradient boosting decision trees. Int J Environ Res Public Health 18, 5879. doi:10.3390/ijerph18115879.

6 Cohen, D. R., Lewis, C., Eddy, C. L., Henry, L., Hodgson, C., L. Huang, F., et al. (2023). In-school and out-of-school suspension: Behavioral and psychological outcomes in a predominately black sample of middle school students. School psychology review 52, 1–14

7 Cuartero, N. and Tur, A. M. (2021). Emotional intelligence, resilience and personality traits neuroticism and extraversion: predictive capacity in perceived academic efficacy. Nurse Educ Today 102, 104933

8 R. L., Pereira, V., and Costa, H. G. (2018). A multicriteria approach to the human development index classification. Social Indicators Research 136, 417–438

9 Furukawa, T. (2020). Outcomes in psychological counseling. CampusHealth 57, 56–63.

10 Hayakawa, T. (2020). Implementation of the mental health survey: limitations and future challenges of the survey. Campus Health 57,51–55.

11 Huang, N., Qiu, S., Yang, S., and Deng, R. (2021). Ethical leadership and organizational citizenship behavior: mediation of trust and psychological well-being. Psychol Res Behav Manage , 655–64

12 Im Jin, J. and Kim, N. C. (2017). Grit, academic resilience, and psychological well-being in nursing students. J Korean Acad Soc Nurs Educ 23, 175–83

13 Iwata, N., Kikuchi, K., and Fujihara, Y. (2016). The usability of cat system for assessing the depressive level of japanese-a study on psychometric properties and response behavior. Int J Behav Med 23, 427–437. doi:10.1007/s12529-015-9503-1. [Medline: 26272358]

14 LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature 521, 436–444

15 Li, W., Huang, J.-Y., Liu, C.-Y., Tseng, J. C., and Wang, S.-P. (2023). A study on the relationship between student'learning engagements and higher-order thinking skills in programming learning. Thinking Skills and Creativity 49, 101369

16 McEown, K., McEown, M. S., and Oga-Baldwin, W. Q. (2023). The role of trait emotional intelligence in predicting academic stress, burnout, and engagement in japanese second language learners. CurrPsychol , 1–11

17 Muro, A., Soler, J., Cebolla, A., and Cladellas, R. (2018). A positive psychological intervention for failing students: does it improve academic achievement and motivation? a pilot study. Learn Motivation 63, 126–32

18 Otomo, A., Iwayama, Y., and Mohri, T. (2014). On-campus data utilization: Working on ir

(institutional research) in universities. Fujitsu 65, 41–47.

19 Pollak, J. P., Adams, P., and Gay, G. (2011). Pam: a photographic affect meter for frequent, in situ measurement of affect. In Proceedings of the SIGCHI conference on Human factors in computing systems. 725–734

20 Sakai, W. and Noguchi, H. (2015). Comparison of tests of mental health for student counseling: formation of a common measure. Jpn J Educ Psychol 63, 111–120. doi:10.5926/jjep.63.111.

21 Sano, A., Phillips, A. J., Yu, A. Z., McHill, A. W., Taylor, S., Jaques, N., et al. (2015). Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. Int Conf Wearable Implant Body Sens Netw 2015. doi:10.1109/BSN.2015. 7299420. Article 7299420

22 Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E., et al. (2018). Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 20, e210. doi:10.2196/jmir.9410.

23 Takaoka, A., Nakai, A., Sugiyama, E., Nozue, T., and Shimizu, R. (2017). The proposal for multi- faceted student support based on student data base: the collaboration of early detection of at-risk students and student counseling services. Meiji Gakuin University bulletin of psychology 27, 81–93.

24 Tate, A. E., McCabe, R. C., Larsson, H., Lundstro¨ m, S., Lichtenstein, P., and Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using machine learning techniques. PLoS One 15, e0230389. doi:10.1371/journal.pone.0230389. [FREE Full text] [Medline: 32251439]

25 Torous, J., Kiang, M. V., Lorme, J., Onnela, J.-P., et al. (2016). New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR mental health 3, e5165

26 Trpcevska, L. (2017). Predictors of psychological well-being, academic self-efficacy and resilience in university students, and their impact on academic motivation (Victoria University)

27 Vihavainen, A., Luukkainen, M., and Kurhila, J. (2013). Using students' programming behavior to predict success in an introductory mathematics course. In Proceedings of the 6th International Conference on Educational Data Mining. 2013 Presented at: EDM '13; July 6-9, 2013; Memphis, TN, USA. 300–303

28 Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., et al. (2014). Student life: assessing mental health, academic performance and behavioral trends of college students using smartphones. In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. 3–14

29 Winata, G. I., Kampman, O. P., and Fung, P. (2018). Attention-based lstm for psychological stress detection from spoken language using distant supervision. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE), 6204–6208

30 Yan, Y. (2024). Analysis and research of psychological crisis behavior model based on improved apriori algorithm. International Journal of Human–Computer Interaction , 1–13

31 Yuan, X., Chen, J., Zhang, K., Wu, Y., and Yang, T. (2022). A stable ai-based binary and

multiple class heart disease prediction model for iomt. IEEE Trans Industr Inform 18, 2032–2040. doi:10.1109/tii. 2021.3098306.

32 Zhou, D., Yu, H., Yu, J., Zhao, S., Xu, W., Li, Q., et al. (2024). Mfds-stgcn: Predicting the behaviors of college students with fine-grained spatial-temporal activities data. IEEE Transactions on Emerging Topics in Computing

33 Leeladhar Chourasiya, Sushma Khatri, Kamal K Sethi, Anita Mahajan, Avani Bhawsar, Tanmay Verma, Neeraj Sharma, From Stress to Success: Leveraging Tracebility and Workability to Support Student Well-Being in Higher Education, 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG)