Computer Science and Information Systems 00(0):0000–0000 https://doi.org/10.2298/CSIS260725053M

Paper Type: Research Paper

Cross-Sectional Classification of Mental Health Risk Indicators in University Students Using Graph Convolutional Networks: The AtRisk-GCN Framework

Seifeddine Mechti[1], Anis Saddoud[2], Azza Allouch[3], Salma Sassi[4], Richard Chbeir[5], Hajer Brahmi[6], Boudour Ammar[6], and Mehdi Chlif[3,7]

[1] Faculty of Economic Science and managemement of Sfax, Larodec Laboratory, Tunis le Bardo, Tunis 3000, Tunisia
seif.mechti@isims.usf.tn
[2] High Institute of Sport and Physical Education of Sfax, University of Sfax, Sfax 3000, Tunisia
[3] Faculty of Mathematical, Physical, and Natural Sciences of Tunis, University of El Manar, Tunisia / Robotics and Internet of Things Laboratory, Prince Sultan University, Riyadh, Saudi Arabia
[4] Department of Computer Engineering, Ecole Centrale d'Electronique, Lyon, 69007, France
[5] University Pau & Pays de l'Adour, E2S UPPA, LIUPPA, Anglet, 64600, France
richard.chbeir@univ-pau.fr (corresponding author)
[6] REGIM-Lab: REsearch Groups in Intelligent Machines, National Engineering School of Sfax (ENIS), BP 1173, Sfax 3038, Tunisia Engineering School of Sfax (ENIS), Department of Management IT, Higher Business School of Sfax (ESCS), University of Sfax, Sfax 3038, Tunisia
[7] Research Laboratory: Sports Performance Optimisation, National Center of Medicine and Science in Sports (NCMSS), Tunis 263, Tunisia

Abstract. Identifying students at psychological risk is a growing public-health concern in higher education. Conventional self-report questionnaires are subject to response bias and provide only static snapshots of mental health status. Graph Convolutional Networks (GCNs) offer a principled way to model feature-based relational similarities between individuals, going beyond independent-sample assumptions made by classical classifiers. In this paper we present AtRisk-GCN, a GCN-based framework for cross-sectional classification of questionnaire-derived and text-derived mental-health risk indicators in student populations. Two datasets are studied: (*i*) a survey-derived dataset of 1 000 records, comprising real medical students from a single institution (Faculty of Biology and Medicine, University of Lausanne, and Lausanne University Hospital, CHUV, Switzerland), drawn from the publicly available data release accompanying Carrard et al. [5]) (Data 1, continuous psychological scale scores as node features) and (*ii*) a publicly available mental-health text corpus of 27 977 samples (Data 2, TF-IDF representations as node features, no synthetic augmentation). In both cases, edges are constructed from cosine-similarity between node feature vectors using a k-nearest-neighbour (k-NN) scheme; no observed social-network data are available or assumed. On Data 1, AtRisk-GCN achieves an accuracy of 0.861 and an F1-score of 0.857, outperforming LSTM, CNN, and Transformer baselines. On Data 2 the model attains 0.84 accuracy and an F1-score of 0.84, surpassing, under matched input representation, TF-IDF-based Logistic Regression and Linear SVM baselines, and clearly outperforming the graph-free MLP trained on identical features.

Keywords: Graph Convolutional Networks, Mental Health Risk Classification, Student Well-being, Feature-similarity Graph, Node Classification, Cross-sectional Study, Data Augmentation Leakage.

1. Introduction

In recent years, mental health disorders among students have emerged as a global public-health concern. Studies indicate that approximately one-third of university students experience mental health problems, with depression and anxiety being the most frequently reported [32,4]. These conditions affect student well-being and have profound implications for academic performance, leading to decreased productivity and increased dropout rates. Timely identification of elevated mental-health risk is therefore important for guiding support and intervention strategies; however, this remains challenging because of the subtle and heterogeneous nature of these conditions.

Traditional mental-health assessment within educational institutions relies primarily on self-report questionnaires and clinical interviews. Although these methods have an established track record, they are subject to response bias, subjectivity, and resource intensiveness. In addition, they may not capture the complex interplay of factors that affect mental health, such as academic stress, social interaction, and individual behaviour [32,4].

These limitations have motivated the use of machine learning (ML) techniques, which can analyse multi-dimensional data in an objective and scalable manner. Classical ML models—Random Forests, Support Vector Machines (SVMs), and logistic regression—have been applied to mental-health screening with varying degrees of success [30]. However, these models often require extensive feature engineering and may not fully capture nonlinear and high-dimensional relationships inherent in mental-health data.

Deep learning architectures (CNNs, RNNs, Transformers) have since been applied to mental-health-related classification tasks, demonstrating improved pattern recognition from text, speech, physiological, and neuroimaging signals [38,39,37]. Nevertheless, these models treat each sample independently and therefore do not exploit potential relational structure between individuals.

Graph Convolutional Networks (GCNs) offer a complementary perspective: they propagate information along edges of a graph, allowing the model to integrate both individual node features and pairwise relational information [16,8]. In settings where genuine social-network data are unavailable—as is the case for most student survey datasets—edges can be derived from feature-based similarity between individuals (e.g. cosine similarity of questionnaire scores or text representations), thereby constructing a *feature-similarity graph*. This is a supervised graph learning approach rather than a social-network analysis. Despite the growing interest in GCNs, their application to student mental-health *classification* remains underexplored.

We note explicitly that neither dataset used in this study was collected or designed by the authors. Both are pre-existing, publicly accessible secondary resources. The compounding leakage risks it introduces when combined with the pre-existing label-construction design of Data 1, is disclosed and discussed in full in Section 2.3 and Section 5.

1.1. Research Objectives

1. Develop a GCN-based model (AtRisk-GCN) that constructs feature-similarity graphs from student survey and text data and performs node-level risk classification.
2. Evaluate AtRisk-GCN against traditional ML (Logistic Regression, Random Forest) and deep learning (CNN, LSTM, Transformer) baselines under a consistent evaluation protocol, distinguishing fair, matched-representation comparisons from secondary ones.
3. Fully and transparently disclose the methodological limitations affecting each dataset—in particular, for Data 1.

1.2. Contributions

1. We propose AtRisk-GCN, a GCN framework that constructs k-nearest-neighbour feature-similarity graphs from tabular and text-based student data and performs node-level classification of questionnaire-derived risk labels.
2. We conduct controlled experiments on two publicly accessible datasets using a unified evaluation protocol, reporting AtRisk-GCN's performance relative to classical ML and deep learning baselines under matched-input-representation comparisons where applicable.

2. Literature Review of Mental Health Detection

The increasing prevalence of mental-health concerns among students has stimulated extensive inquiry into the complex interplay between psychological, social, and academic determinants. Various ML and deep learning techniques have been developed to detect mental-health conditions; however, validation remains limited due to scarce data and inconsistent evaluation methods.

2.1. Student Mental Illness

Prevalent Disorders and Contributing Factors. University students often experience a variety of mental disorders, including schizophrenia, depression, bipolar disorder, and anxiety disorders. These conditions can substantially affect both academic performance and social functioning; they are frequently linked to academic stress, social isolation, and transitional challenges during higher education [20]. Table 1 summarises the most common mental-health disorders observed among students, their primary characteristics, and key contributing factors.

Institutional Responsibility and Intervention. Despite increasing awareness, many universities continue to provide insufficient mental-health support, with limited resources and preventive programmes [9]. The transition to higher education heightens vulnerability to psychological distress, while stigma remains a major barrier to help-seeking. The literature emphasises the need for comprehensive institutional strategies, including expanding accessible mental-health services, conducting anti-stigma campaigns, and implementing evidence-based interventions [9,20].

Table 1. Summary of Common Mental Disorders Among Students

Disorder	Main Characteristics	Contributing Factors	Ref.
Schizophrenia	Hallucinations, delusions, social withdrawal; anxiety in up to 65% of cases	Stress, isolation, academic pressure	[3]
Depression	Persistent sadness, loss of interest, irritability, suicidal ideation	Academic overload, social isolation	[20]
Bipolar Disorder	Alternating manic and depressive episodes	Irregular lifestyle, stress, lack of support	[20]
Anxiety Disorders	Excessive worry, fear, panic attacks	Academic stress, uncertainty about future	[22]

2.2. Mental Health Detection Techniques

Traditional ML models (Random Forest, SVM, CATBoost, and hybrids) have been applied to depression, anxiety, schizophrenia, ADHD, and stress using questionnaires, social-media content, biosignals (EEG, fMRI), and academic records [25,7,21,30,11,6,4,33,10]. Deep learning models (CNNs, RNNs, autoencoders) outperform classical ML on complex, high-dimensional data [17,12]. Graph Neural Networks (GNNs) and GCNs represent a further step: GCNs achieved 85.8% accuracy for schizophrenia detection compared to 80.9% with SVM, while GNNs achieved 77.78% for major depressive disorder prediction [19,16,35,8,1]. An overview of these approaches is provided in Table 2.

Table 2. Classification of Studies on ML and Deep Learning Techniques in Mental Health Detection.

Category	Target	Techniques	Data	Findings	Ref.
Traditional	General	Interviews, questionnaires	Surveys	Subjective, limited	—
ML	General	RF, AdaBoost, LR, KNN, DT	Questionnaires, OSN	High accuracy	[26,18,29]
ML	Depression	RF, NB, SVM, NN	Public datasets	RF: 86.8%	[25,7,21]
ML	Stress	SVM, wearables	EEG, physiological	99.27–99.98%	[30]
ML	Anxiety	CATBoost, RF, LR	Clinical	CATBoost: 89.3%	[11]
ML	ADHD	SVM, Hybrid, CNN	fMRI, EEG	CNN: 97.78%	[6,4]
ML	Schizophrenia	XGBoost, DNN	EEG, imaging	AUC: 0.94	[33,10]
DL	Multiple	CNN, RNN, AE	Social media, imaging	High accuracy	[17,12]
GNN	Multiple	GCN, EV-GCN, GCRN	Neuroimaging, fMRI	GCN: 85.8%	[19,16,35,8,1]
Challenges	General	Interpretability	Multimodal	Data dependency	[15,1]

2.3. Mental Health Datasets

Several datasets have been developed to study mental health using surveys, behavioural data, physiological signals, and text, including DASS-21 [24], DAIC-WOZ [13], eRisk [23],

and StudentLife [36]. For this study, we use two datasets: (*i*) a medical student mental-health survey dataset and (*ii*) the publicly available Mental Health Corpus [27]. Neither dataset was collected by the authors; Data 1 additionally includes a synthetic augmentation component produced by the authors, described in full below.

Dataset 1: Medical Student Mental Health Survey. Dataset 1 originates from the publicly available data release accompanying Carrard et al. [5], a cross-sectional study of medical students' empathy, mental health, and burnout conducted at the Faculty of Biology and Medicine, University of Lausanne, and Lausanne University Hospital (CHUV), Switzerland, supported by the Swiss National Science Foundation (grant 10001C_197442).

The data were obtained via an anonymous online survey and cover demographic and academic variables (gender identity, curriculum year, mother tongue, partnership status, paid employment status, self-rated health satisfaction, weekly study hours) and validated psychological scales: the Center for Epidemiologic Studies Depression Scale (CES-D), the Spielberger Trait Anxiety Inventory (STAI-T), and the Maslach Burnout Inventory (MBI; Emotional Exhaustion, Cynicism, and Academic Efficacy subscales), as well as a binary indicator of psychiatric/psychotherapy consultation in the previous 12 months. The original release additionally includes empathy-related measures (JSPE, QCAE, AMSP, GERT), which are **not used** in the present study.

Table 3. Summary of Dataset 1 Characteristics.

Category	Details
Real records	886 (Carrard et al. [5])
Synthetic records (SMOTE)	114 (11.4%)
Total sample size	1 000
Institution (real records)	Faculty of Biology and Medicine, University of Lausanne & Lausanne University Hospital (CHUV), Switzerland (single site)
Collection method	Anonymous online survey (real records)
Demographic features	Gender identity, mother tongue, partnership status, employment status, self-rated health satisfaction
Psychological scales	CES-D, STAI-T, MBI (Exhaustion, Cynicism, Academic Efficacy), psychiatric consultation (past 12 months)
Academic features	Weekly study hours, curriculum year
Class distribution	At Risk: 612 (61.2%); Not at Risk: 388 (38.8%)

The real-record component is self-reported, cross-sectional, single-institution (Switzerland), and thus subject to reporting bias and limited generalisability. The synthetic component is discussed below as a distinct, additional limitation.

Feature preprocessing. Categorical variables (gender identity, curriculum year, employment status, mother tongue) were encoded with a label encoder. Continuous features (age, study hours, CES-D total score, STAI-T total score, MBI subscale scores) were standardised using z-score normalisation.

Label construction and compounding leakage risk. A binary "At Risk" label was constructed from established clinical thresholds: CES-D ≥ 16, or STAI-T ≥ 40, or MBI-Exhaustion ≥ 27, or psychiatric consultation $= 1$.

We identify and disclose three distinct, compounding leakage channels affecting Data 1.

Channel 1 – Feature-level label leakage. The total CES-D, STAI-T, and MBI-Exhaustion scores used to construct the label are also included as node features in the GCN.

Channel 2 – Graph-level (structural) label leakage. These same label-defining scores also contribute to the cosine-similarity computation used to build the k-NN graph (Section 3.1), so that nodes with similar label-defining scores are more likely to be connected *and* to share the same label.

Channel 3 – SMOTE augmentation leakage (pre-split, pre-graph). The 114 synthetic records were generated by interpolating between real, same-class nearest neighbours *before* the train/test split and *before* k-NN graph construction, and using *all* variables (including the Channel-1/2 label-defining scores). As a direct consequence: (i) a synthetic record assigned to the test set is, by construction, a near-interpolation of one or more real records that may be assigned to the training set (or vice versa), creating a direct and severe train/test leakage pathway independent of Channels 1–2; (ii) the k-NN graph may connect synthetic nodes to their real "parent" nodes with near-maximal cosine similarity, regardless of which split each node is later assigned to, further reinforcing label propagation specifically for these nodes; and (iii) standard (continuous-variable) SMOTE interpolation is not designed for the categorical/ordinal variables present in Data 1 (e.g. gender identity, curriculum year), so synthetic values for these variables may not correspond to valid categories, raising an additional data-quality concern distinct from leakage per se.

These three channels compound one another: Channel 3 means that even if Channels 1–2 were resolved (e.g. by removing CES-D/STAI-T/MBI from the features and the graph-construction inputs), residual train/test leakage via synthetic near-duplicates would likely remain unless SMOTE is also moved to after the split and after graph construction, and restricted to the training partition only.

Important caveat: the total CES-D, STAI-T, and MBI-Exhaustion scores are also included as node features in the GCN. Consequently, the model has access to the primary variables from which the label is derived. This creates a target-leakage risk: a sufficiently flexible classifier could approximate the label-construction rule without learning clinically meaningful patterns. We address this concern in two ways. First, we report individual feature-to-label correlations (ranging from -0.21 to 0.32), which show that no single feature deterministically encodes the label, consistent with the OR-based composite rule used; however, given Channel 3, these correlations do not rule out near-duplicate-level leakage between specific train/test record pairs, which a global correlation statistic cannot detect.

Dataset 2: Mental Health Corpus (Namdari 2023) – A Technical Benchmark Only. We stress upfront that Dataset 2 serves exclusively as a **secondary, methodological**

benchmark used to assess whether the AtRisk-GCN mechanism transfers to a different data modality (text) and larger sample size. It does **not** serve as validation of AtRisk-GCN for the university-student mental-health classification task that motivates this study, and it involves no synthetic augmentation.

Dataset 2 is the publicly available Mental Health Corpus [27] (Kaggle). It comprises 27 977 short text samples, each associated with a binary label that indicates whether the text contains language related to mental distress (originally annotated as "toxic" vs. "non-toxic" in the context of online mental-health communities). For consistency with the classification objective we denote these labels as "At Risk" and "Not at Risk"; however, we stress that **this relabelling is a modelling convenience and does not constitute a clinically validated mental-health diagnosis**. The dataset was not collected exclusively from university students and does not provide verified demographic information; the source population is an unspecified online community, which does not correspond to the university-student population targeted by this study's research questions. Results on Dataset 2 therefore reflect the model's ability to classify distress-related language rather than to screen individuals for clinical mental-health disorders.

Highly correlated terms with the positive label include words such as "suicidal", "kill", "die", and "depression". This indicates that part of the classification signal may originate from explicit distress keywords (shortcut learning); this concern is directly compounded by the graph-construction procedure itself: because k-NN edges are derived from TF-IDF cosine similarity, texts sharing these same distress keywords are also more likely to become directly connected nodes, so the graph topology may largely encode lexical co-occurrence rather than latent psychological-distress structure. We discuss this as a limitation and recommend future work that investigates the model's behaviour on texts without overt distress lexicon.

Each text sample was preprocessed as follows: lowercasing, removal of non-alphabetical characters, and whitespace normalisation. TF-IDF vectorisation was then applied with a vocabulary restricted to the 300 most informative terms, yielding a feature matrix of size $(27{,}977 \times 300)$. TF-IDF was fitted exclusively on the training split to prevent information leakage into validation and test sets. The class distribution is approximately balanced: 13 836 "At Risk" samples (49.5%) and 14 141 "Not at Risk" samples (50.5%).

2.4. Limitations and Research Gaps

Although ML holds promise for mental-health screening, several limitations must be acknowledged [28]:

- **Label validity:** questionnaire-derived and proxy labels are not equivalent to independent clinical diagnoses.
- **Compounding** target leakage (Data 1): using the same psychometric scores as both features and graph-construction inputs, as well as label components, and as inputs to a pre-split, pre-graph synthetic augmentation procedure, can inflate performance estimates.
- **Generalisation:** the real-record component of Data 1 originates from a single Swiss institution; models trained on a single text corpus may not transfer to other contexts.
- **Interpretability:** GCN predictions are difficult to explain to clinicians and students.

– **Cross-sectional design:** snapshot data cannot establish causal or temporal relationships.

Addressing these gaps requires prospective cohort studies, independent outcome measurements, and rigorous graph-construction protocols .

3. Methods

3.1. Graph Construction

Neither dataset contains observed social-network data. Graphs are therefore constructed from pairwise feature similarity between samples, following the standard k-NN graph approach used in semi-supervised GCN literature [16].

Similarity metric.

$$s(i,j) = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \, \|\mathbf{x}_j\|}. \tag{1}$$

Edge construction. An undirected edge (i, j) is added if node j belongs to the k nearest neighbours of node i in cosine-similarity space (or vice versa), using $k = 10$ for Data 1 and $k = 15$ for Data 2. These values were selected by grid search on the validation set over the range $k \in \{5, 10, 15, 20, 25\}$. Edges are unweighted.

Transductive split. The graph is constructed on all nodes before splitting, which gives a transductive setting: the GCN can propagate information through edges connecting training and test nodes during inference. Importantly, because the k-NN edges are computed from pairwise feature similarities on the full node set, the test-node feature vectors also influence the neighborhood structure seen during training. While test labels remain unseen, this can yield performance estimates that are optimistic relative to an inductive deployment scenario in which new students/texts arrive only after training. We adopt this setting because it follows the standard evaluation protocol of the original GCN formulation [16].

Compounding with SMOTE augmentation (Data 1 only). For Data 1, the 114 SMOTE-generated synthetic nodes are also present in the graph at construction time, prior to the train/test split (Section 2.3, Channel 3). This means the transductive leakage described above is further compounded: synthetic nodes assigned to the test split may be directly connected, with very high cosine similarity, to their real "parent" nodes in the training split. This is a distinct and more severe leakage pathway than the standard transductive setting alone, and is not present in Data 2, which involves no synthetic augmentation.

To prevent feature-level leakage, TF-IDF (Data 2) and z-score normalisers (Data 1) are fitted solely on training-node features. As future work, an inductive variant should be evaluated (e.g., build the graph using training nodes only, then connect validation/test nodes to the training graph using k-NN computed against training features only, or use an inductive GNN such as GraphSAGE).

Graph statistics. Table 4 reports key structural properties of the two resulting graphs. For Data 1, these statistics are computed on the full $N = 1{,}000$.

Table 4. Graph Statistics for Data 1 and Data 2.

Property	Data 1	Data 2
Nodes	1 000	27 977
Edges	4 872	198 430
Edge density	0.0098	0.0005
Average degree	9.74	14.19
Connected components	1	1

3.2. Model Architecture: AtRisk-GCN

AtRisk-GCN performs **node-level binary classification**: each node (student record or text sample) receives an individual prediction. This is consistent with the task definition and the reported per-sample confusion matrices.

GCN layer. Node representations are updated according to the spectral graph convolution rule of Kipf & Welling [16]:

$$H^{(l+1)} = \sigma\Big(\tilde{D}^{-\frac{1}{2}}\tilde{A}\,\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\Big), \tag{2}$$

where $\tilde{A} = A + I$ is the adjacency matrix with added self-loops, $\tilde{D}$ is its degree matrix, $W^{(l)}$ is a trainable weight matrix, and σ is ReLU.

Output layer. After L graph-convolutional layers, each node embedding $h_i^{(L)}$ is passed directly to a linear output layer followed by a sigmoid activation for binary classification:

$$\hat{y}_i = \sigma_{\text{sig}}\Big(W_o\, h_i^{(L)} + b_o\Big). \tag{3}$$

No graph-level readout is applied; the task requires one prediction per node.

Architecture summary.

- **Data 1:** 2 GCN layers (hidden dim. 64), ReLU, dropout $p = 0.5$; Adam optimiser, $\text{lr} = 0.005$, weight decay 5×10^{-4}; early stopping with patience 15; 200 maximum epochs.
- **Data 2:** 2 GCN layers (hidden dim. 64), ReLU, dropout $p = 0.5$; Adam optimiser, $\text{lr} = 0.01$, weight decay 5×10^{-4}; 50 epochs.

Training uses binary cross-entropy loss.

3.3. Evaluation Protocol

To ensure a fair and reproducible comparison, **all models** are evaluated under the same protocol:

- **Data 1:** stratified 80/20 train–test split (random seed 42) on the $N = 1{,}000$ sample ; 10% of the training set is held out as a validation set (i.e. 72/8/20 overall). Five independent runs with different random seeds are performed; mean and standard deviation are reported.
- **Data 2:** stratified 72/8/20 train–validation–test split (random seed 42). Five independent runs are performed.

Performance metrics: accuracy, macro-precision, macro-recall, and macro-F1. For the GCN, the best checkpoint (highest validation accuracy) is selected, and final metrics are reported on the held-out test set.

3.4. Baseline Models

Data 1 baselines. Logistic Regression and Random Forest (Scikit-learn, default hyperparameters with class-weight balancing); DNN (3 fully connected layers, hidden dims 128-64-32, ReLU, dropout 0.3, Adam, lr 0.001); CNN (1D convolution, 3 layers, 64 filters, kernel size 3, global average pooling); LSTM (2 layers, hidden dim 64, dropout 0.3); Transformer encoder (2 heads, 2 layers, hidden dim 64). All deep learning models use the same train/validation/test split and are trained with early stopping (patience 10).

Role of baselines for Data 1. Logistic Regression and Random Forest constitute the principal tabular-data baselines for Data 1, and the DNN constitutes the principal *graph-free* comparison, since it is trained on exactly the same node features as AtRisk-GCN. CNN, LSTM, and Transformer architectures, while included for completeness and for comparability with the Data 2 experimental design, are not natively suited to low-dimensional tabular input and should be interpreted as secondary points of comparison. **All Data 1 baselines, including AtRisk-GCN, are trained and evaluated under the same compounding leakage conditions described in Section 2.3**; the relative ranking among them may therefore be more informative than their absolute performance values. Gradient-boosted tree models (e.g. XGBoost, LightGBM), which are standard strong baselines for clinical tabular data, were not included in this study and are recommended as a priority addition in future work.

Data 2 baselines. MLP (3 layers, hidden dims 256-128-64, ReLU, dropout 0.3, Adam); CNN and LSTM as above. Preliminary experiments showed that standard sequence-based CNN and LSTM architectures are ill-suited to TF-IDF bag-of-words features for discussion). Additional text-classification baselines (TF-IDF + Logistic Regression, TF-IDF + Linear SVM) are reported for reference.

4. Results

4.1. Results on Data 1: Medical Student Mental Health Survey

All models are trained and evaluated on the same 80/20 stratified split of the, augmented before splitting; see Section 2.3), yielding a test set of **200 observations** (123 “At Risk”; 77 “Not at Risk”). Results are averaged over five independent runs; standard deviations are reported in parentheses.

We reiterate that these results are reported under the compounding feature-level, graph-level, and augmentation-level leakage conditions disclosed in Section 2.3, and should be interpreted as an upper-bound estimate, not as unbiased generalisation performance.

Baseline and AtRisk-GCN Results. Table 5 presents the performance of all baseline models and AtRisk-GCN on Data 1.

Table 5. Results on Data 1 (Medical Student Mental Health Survey,), averaged over five runs (std in parentheses). Task: binary classification (At Risk / Not at Risk). Test set: $n = 200$. Results are affected by compounding leakage (Section 2.3) and should be read as upper-bound estimates.

Model	Acc.	Prec.	Rec.	F1
Logistic Regression	0.750 (0.012)	0.760 (0.014)	0.740 (0.013)	0.750 (0.013)
Random Forest	0.780 (0.011)	0.790 (0.012)	0.770 (0.013)	0.780 (0.012)
CNN	0.760 (0.014)	0.758 (0.015)	0.755 (0.016)	0.756 (0.015)
LSTM	0.780 (0.013)	0.782 (0.014)	0.778 (0.014)	0.780 (0.014)
DNN	0.810 (0.010)	0.820 (0.011)	0.800 (0.012)	0.810 (0.011)
Transformer	0.830 (0.009)	0.832 (0.010)	0.828 (0.011)	0.830 (0.010)
AtRisk-GCN	**0.861 (0.008)**	**0.854 (0.009)**	**0.860 (0.009)**	**0.857 (0.009)**

AtRisk-GCN achieves an accuracy of 0.861 ($\sigma = 0.008$), an improvement of 3.1 percentage points over the best baseline (Transformer, 0.830). The relative gain is:

$$\text{Relative Gain} = \frac{0.861 - 0.830}{0.830} \times 100 \approx 3.7\%. \tag{4}$$

As noted above, Transformer is a secondary comparison for this tabular dataset; the principal matched-feature comparison is against the graph-free DNN (0.810). Given the compounding leakage affecting all models in this table equally (Section 2.3), this relative ranking is more interpretable than the absolute accuracy values, which are very likely inflated relative to what would be observed on the 886 real records alone, evaluated without synthetic augmentation.

Training dynamics are shown in Figure 1: the training loss decreases steadily from 1.30 to 0.41, and the validation accuracy reaches 0.871 before early stopping is triggered.

Figure 2 shows the validation accuracy across epochs (left) and the confusion matrix on the 200-sample test set (right).

4.2. Results on Data 2: Mental Health Corpus

The test set contains **5 596 observations** (2 828 “Not at Risk”; 2 768 “At Risk”), corresponding to the 20% stratified split of the 27 977-sample corpus (no synthetic augmentation used). Table 6 reports the performance of all models.

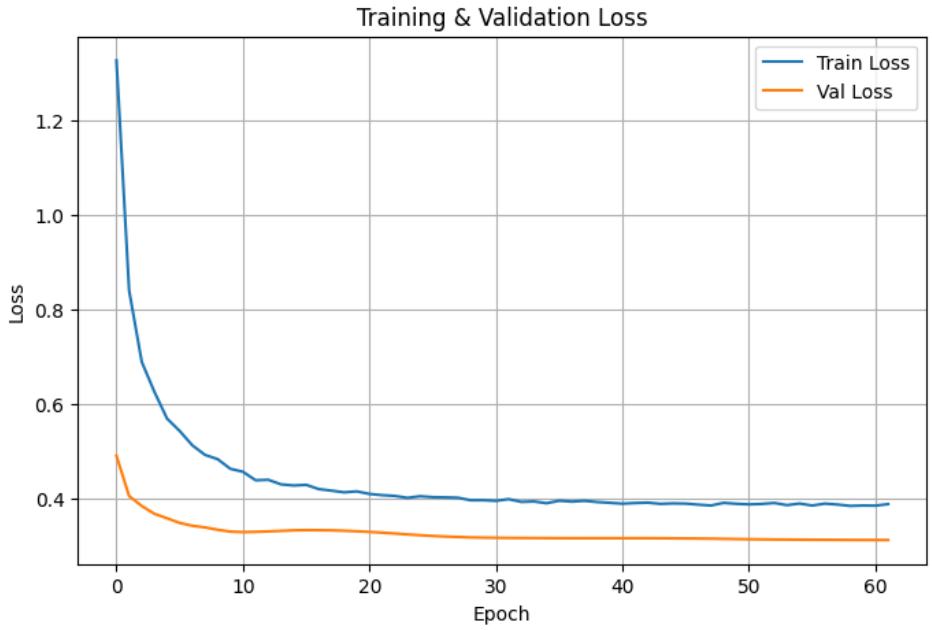


Fig. 1. Data 1 : AtRisk-GCN training and validation binary cross-entropy loss (y-axis) across epochs (x-axis). Although the maximum was 200 epochs, training stops earlier due to early stopping (curves shown up to the final executed epoch).

Primary vs. secondary baselines. We distinguish two roles among the Data 2 baselines. The *primary, fair comparison* for AtRisk-GCN is against TF-IDF + Logistic Regression and TF-IDF + Linear SVM, since these models operate on *exactly the same input representation* (300-dimensional TF-IDF vectors) as AtRisk-GCN's node features. Under this fair comparison, AtRisk-GCN (0.840 F1) outperforms TF-IDF + Logistic Regression (0.82 F1) and TF-IDF + Linear SVM (0.81 F1) by a modest but consistent margin. The MLP constitutes the principal *graph-free* comparison, since it is trained on exactly the same TF-IDF features as AtRisk-GCN. CNN and LSTM are reported for architectural completeness and consistency with the Data 1 experimental design, but are not well-suited to bag-of-words input and should **not** be treated as the primary basis for assessing the added value of AtRisk-GCN. A stronger secondary comparison – fine-tuned Transformer language models operating on raw text – is specified in Section 5.3 (Step 6) as necessary future work.

Remarks on CNN and LSTM performance. The near-chance accuracy of CNN and LSTM ($\approx$0.51–0.52) is explained by the mismatch between bag-of-words TF-IDF input and the sequential inductive bias of these architectures. When a 300-dimensional TF-IDF vector is treated as a sequence of 300 scalar timesteps, no meaningful temporal or local structure is present and the models effectively fail to learn. This result is consistent with the text-classification literature; TF-IDF features are better paired with linear classifiers or feedforward networks, as the MLP results confirm.

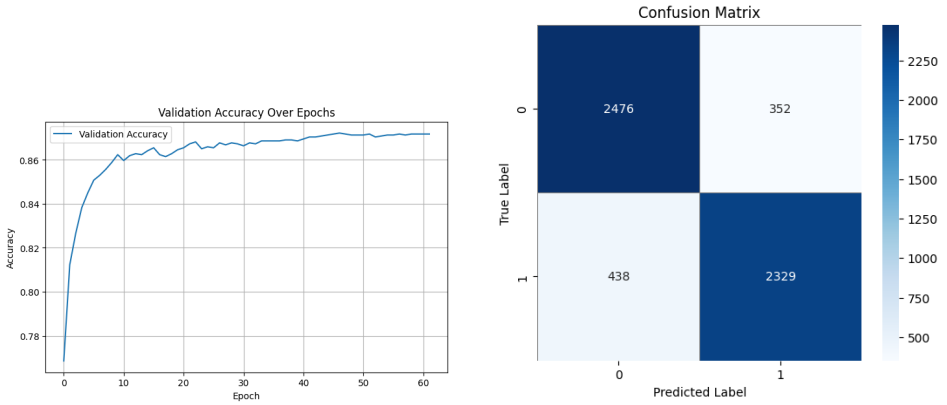


Fig. 2. Data 1: Validation accuracy (y-axis, 0–1) across epochs (x-axis) (left) and confusion matrix with absolute counts on the held-out test set ($n = 200$, right). The accuracy curve is shown up to the early-stopping epoch. Results affected by compounding leakage; see Section 2.3.

Training dynamics for AtRisk-GCN are shown in Figure 3, and Figure 4 shows the accuracy evolution and the confusion matrix on the test set.

A detailed classification report is provided in Table 7.

Comparison with prior studies on the Mental Health Corpus. Table 8 compares AtRisk-GCN with recent studies using the same corpus. AtRisk-GCN achieves an F1-score of 0.84; however, the comparison is approximate because evaluation protocols may differ across studies.

Table 6. Results on Data 2 (Mental Health Corpus), averaged over five runs (std in parentheses). Task: binary distress-language classification. Test set: $n = 5\,596$.

Model	**Acc.**	**Prec.**	**Rec.**	**F1**
TF-IDF + Logistic Regression	0.82 (0.004)	0.82 (0.004)	0.82 (0.005)	0.82 (0.004)
TF-IDF + Linear SVM	0.81 (0.003)	0.81 (0.004)	0.81 (0.004)	0.81 (0.004)
MLP	0.70 (0.008)	0.67 (0.009)	0.63 (0.010)	0.65 (0.009)
CNN†	0.52 (0.021)	0.51 (0.022)	0.50 (0.023)	0.50 (0.022)
LSTM†	0.51 (0.019)	0.50 (0.020)	0.50 (0.021)	0.50 (0.020)
AtRisk-GCN	**0.840 (0.006)**	**0.840 (0.007)**	**0.840 (0.007)**	**0.840 (0.006)**

CNN and LSTM are applied to TF-IDF bag-of-words vectors treated as fixed-length sequences. These architectures are designed for ordered sequential data and cannot exploit the permutation-invariant structure of TF-IDF vectors, which explains their near-chance performance. Results are included for completeness only; future comparisons should include fine-tuned transformer language models (e.g. BERT, RoBERTa).



Fig. 3. Data 2: AtRisk-GCN training and validation binary cross-entropy loss (y-axis) across epochs (x-axis).

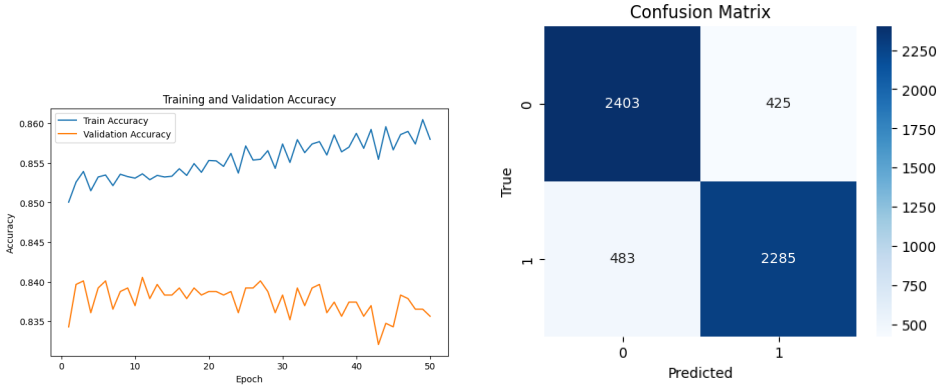


Fig. 4. Data 2: Accuracy (y-axis, 0–1) across epochs (x-axis) (left) and confusion matrix with absolute counts on the held-out test set ($n = 5\,596$, right).

Table 7. Classification report for AtRisk-GCN on the test set of Data 2 ($n = 5\,596$).

Class	Prec.	Rec.	F1	Support
Not at risk	0.83	0.85	0.84	2 828
At risk	0.85	0.83	0.84	2 768
Accuracy			0.84	5 596

Table 8. Macro F1-scores of different studies on the Mental Health Corpus.

Authors	Model	F1
Baydili et al. (2025) [2]	Feature selection	0.56
Tank et al. (2024) [31]	Semi-supervised Su-RoBERTa	0.70
Saeed et al. (2025) [34]	Behavioural + linguistic	0.74
AtRisk-GCN (2025)	GCN (this work)	**0.84**

4.3. Graph Contribution: Evidence from Existing Comparisons

While a complete, controlled ablation study (identical architecture with and without graph convolution, randomised graph connectivity, and a k-NN classifier baseline) was not performed in this work and is identified as a priority for future work (Section 5.3, Step 5), two comparisons already present in our experimental design provide partial, indirect evidence regarding the contribution of the feature-similarity graph.

For Data 1, the DNN baseline (Table 5) is trained on the *same standardised tabular node features* used by AtRisk-GCN, without any graph structure. AtRisk-GCN outperforms this graph-free model by 5.1 percentage points in accuracy (0.861 vs. 0.810). For Data 2, the MLP baseline (Table 6) is trained on the *same TF-IDF feature vectors* used by AtRisk-GCN, again without graph structure, and is outperformed by 14.0 percentage points (0.840 vs. 0.700). The Data 2 comparison is not affected by the Data 1 augmentation issue and is the more reliable of the two indirect indicators; for Data 1, the 5.1-point gap cannot be attributed to graph convolution alone, as both models are affected by the same compounding leakage described in Section 2.3.

These gaps are *consistent with* a positive contribution of the feature-similarity graph, but we emphasise that DNN/MLP and AtRisk-GCN also differ in other architectural respects; consequently, these comparisons cannot *isolate* the graph's contribution with the same rigour as a controlled ablation study.

5. Discussion

5.1. Interpretation of Results

AtRisk-GCN outperforms classical ML and deep learning baselines across two heterogeneous datasets. On Data 2 – this advantage (84.0% vs. 70.0% for the graph-free MLP, and vs. 82.0% for TF-IDF + Logistic Regression) is the more trustworthy evidence for a genuine contribution of the feature-similarity graph. On Data 1 it achieves 86.1% accuracy (F1 $= 0.857$) compared to 81.0% for the matched-feature graph-free DNN baseline;

however, this result must be interpreted in light of the compounding, severe limitations described below, and should not be read as evidence of the graph's standalone contribution or of genuine mental-health risk detection.

The performance advantage of AtRisk-GCN over standard feedforward models is consistent with the hypothesis that feature-similarity graphs provide complementary structural information that aids classification (most clearly supported by the Data 2 evidence). By aggregating representations of similar individuals, the GCN effectively performs a form of graph-regularised learning, which may reduce overfitting and improve generalisation; however, a controlled ablation is still required to confirm this mechanism rather than a difference in model capacity or training procedure.

5.2. Methodological Limitations

Compounding leakage in Data 1: label, graph, and synthetic augmentation. We summarise the three compounding leakage channels disclosed in Section 2.3: (1) CES-D/STAI-T/ MBI-Exhaustion are used simultaneously as label components and as node features; (2) these same variables are used to construct the k-NN graph. Channel 3 is, in our assessment, the most severe: it creates a direct mechanism for near-duplicate information to cross the train/test boundary, independent of Channels 1–2. We did not resolve these issues in the present study, and we do not consider the reported Data 1 performance to be a reliable estimate of AtRisk-GCN's ability to detect at-risk students from independent data. We regard this disclosure, together with the protocol in Section 2.3, as more scientifically important than the numerical performance values themselves, and we leave to the Editor and reviewers the determination of whether this level of disclosed limitation is acceptable for publication in its current form or requires the leakage-free re-evaluation specified below as a condition of acceptance.

Label validity in Data 2. The Mental Health Corpus was originally annotated for toxic language in online communities. Relabelling these annotations as "At Risk" is a modelling convenience with no clinical basis. The model learns to classify distress-related language, not to identify individuals with mental-health disorders. Combined with the lexical shortcut-learning risk discussed in Section 2.3 (keyword–graph coupling via TF-IDF cosine similarity), results on Data 2 should not be extrapolated to clinical screening scenarios and are treated in this paper strictly as a technical benchmark (Section 4.2), unaffected by Data 1's augmentation issues.

Cross-sectional design and generalisability. Both datasets are static snapshots. AtRisk-GCN assigns labels to current observations and cannot predict future mental-health deterioration. In addition, the real-record component of Data 1 originates from a single Swiss institution, and may not generalise to other student groups, institutions, countries, or languages; the online-community population underlying Dataset 2 provides a further, independent generalisability concern. No subgroup analysis is reported.

5.3. A Concrete Protocol for Leakage-Free and Inductive Re-Evaluation

To address the methodological concerns above in a verifiable manner, we specify here the exact protocol required for a rigorous re-evaluation of AtRisk-GCN, which we identify as

the immediate priority for future work. We have not conducted this re-evaluation in the present study and therefore do not claim that AtRisk-GCN's advantage on Data 1 would persist under this protocol.

Step 1 – Leakage-free feature set (Data 1). Remove CES-D, STAI-T, and MBI-Exhaustion from the node feature matrix.

Step 2 – Leakage-free graph construction (Data 1). Recompute the k-NN graph using only the reduced feature set from Step 1.

Step 3 – Independent outcome (preferred, Data 1). Where available, replace the composite threshold-based label with an independently measured outcome.

Step 4 – Inductive evaluation protocol (both datasets). Construct the k-NN graph on training nodes only; attach validation/test nodes via k-NN against training features only, or adopt an inductive architecture (e.g. GraphSAGE [14]).

Step 5 – Controlled ablation of the graph component. Compare AtRisk-GCN against a matched-capacity graph-free MLP, a k-NN classifier, and a randomised-graph AtRisk-GCN variant, all on the Step 0–2 leakage-free data.

Step 6 – Stronger, task-appropriate baselines. Gradient-boosted trees (Data 1); fine-tuned Transformer language models (Data 2).

5.4. Absence of Clinical Deployment Recommendation

Given the limitations above, this work does not recommend the deployment of AtRisk-GCN in real student support systems. Before any deployment, the following would be required. (b) prospective cohort studies with independently measured clinical outcomes; (c) subgroup fairness analysis; (d) calibration and uncertainty quantification; (e) clinical validation in partnership with mental-health professionals; and (f) ethical review including privacy, consent, and risk-of-harm analysis.

6. Conclusion

This study presented AtRisk-GCN, a Graph Convolutional Network framework designed for the cross-sectional classification of mental-health risk indicators using feature-similarity graphs. By constructing k-nearest-neighbour graphs from questionnaire-derived and text-based representations, the proposed framework incorporates relational information between observations that is not explicitly exploited by conventional graph-free classifiers.

The experimental results provide evidence that AtRisk-GCN can achieve competitive classification performance across two heterogeneous datasets. On the medical-student dataset, the model achieved an accuracy of 86.1% and an F1-score of 85.7%, while on the larger Mental Health Corpus it achieved 84.0% accuracy and 84.0% F1-score. The results obtained on the second dataset, which is not affected by synthetic augmentation, provide the more reliable evidence that feature-similarity graphs may offer complementary information

beyond the original feature representations. In particular, AtRisk-GCN outperformed both graph-free MLP and conventional TF-IDF-based classifiers under matched input representations.

Nevertheless, the findings should be interpreted in light of important methodological limitations. The medical-student dataset is affected by three interacting sources of information leakage: label-defining psychometric variables are included in the node features, the same variables contribute to graph construction, and SMOTE augmentation was performed before the train–test split and graph construction. Consequently, the reported performance on this dataset should be regarded as an upper-bound estimate rather than an unbiased measure of generalisation. Similarly, the second dataset is used strictly as a technical benchmark because its labels represent distress-related language rather than clinically validated mental-health outcomes. Therefore, the present results do not support direct clinical deployment or individual-level mental-health screening.

Future work should conduct a rigorous leakage-free and inductive re-evaluation of AtRisk-GCN by removing label-defining variables from the feature and graph-construction processes, applying data augmentation only within the training partition when appropriate, and using independently measured outcomes whenever available. Further experiments should include controlled graph ablations, stronger tabular and text baselines, prospective multi-institutional datasets, subgroup fairness analysis, calibration and uncertainty assessment, and independent clinical validation. Such evaluations will be necessary to determine whether the observed benefits of graph-based learning persist under realistic deployment conditions.

Overall, the study provides a methodological investigation of feature-similarity GCNs for cross-sectional mental-health indicator classification and highlights both their potential value and the importance of rigorous leakage control when applying graph learning to sensitive psychological data.

Ethics Statement. This study is a secondary analysis of pre-existing, publicly available data. The authors did not recruit participants, did not collect new data from human subjects, and had no access to directly identifying information.

The real-record component of Dataset 1 (886 records) derives from the publicly available data release accompanying Carrard et al. [5], deposited on Zenodo (`https://doi.org/10.5281/zenodo.5702895`). The original study was conducted at the Faculty of Biology and Medicine, University of Lausanne, and Lausanne University Hospital (CHUV), Switzerland, and supported by the Swiss National Science Foundation (grant 10001C_197442). No ethics approval/protocol number is stated in the cited data release or its associated publication; readers are referred to Carrard et al. [5] for the original study's institutional context. The 114 synthetic records were generated by the authors using SMOTE and do not correspond to real individuals; their generation and associated leakage risks are disclosed in full in Section 2.3.

Dataset 2 is the publicly available Mental Health Corpus [27], distributed under its original licence on Kaggle.

As no new human-subjects data collection was performed, no new institutional ethics board approval was sought for this secondary analysis.

Data Availability. The real-record component of Dataset 1 is available via Zenodo at `https://doi.org/10.5281/zenodo.5702895` [5]. Dataset 2 is publicly available on Kaggle [27]. Code for graph construction, model training, and evaluation is available upon reasonable request from the corresponding author.

References

1. Ahmedt-Aristizabal, D., Armin, M.A., Denman, S., Fookes, C., Petersson, L.: Graph-based deep learning for medical diagnosis and analysis: Past, present and future. Sensors 21 (2021)
2. Baydili, I., Tasci, B., Tasci, G.: Deep learning-based detection of depression and suicidal tendencies in social media data with feature selection. Behavioral Sciences 15(3), 352 (2025)
3. Braga, R.J., Mendlowicz, M.V., Marrocos, R.P., Figueira, I.L.: Anxiety disorders in outpatients with schizophrenia: prevalence and impact on the subjective quality of life. Journal of Psychiatric Research 39(4), 409–414 (2005)
4. Cao, M., Martin, E., Li, X.: Machine learning in attention-deficit/hyperactivity disorder: new approaches toward understanding the neural mechanisms. Translational Psychiatry 13(1), 236 (2023)
5. Carrard, V., Bourquin, C., Berney, S., Schlegel, K., Gaume, J., Bart, P.A., Preisig, M., Schmid Mast, M., Berney, A.: Dataset for the paper "the relationship between medical students' empathy, mental health, and burnout: A cross-sectional study" published in medical teacher (2022). Zenodo (2022)
6. Chen, T., Tachmazidis, I., Batsakis, S., Adamou, M., Papadakis, E., Antoniou, G.: Diagnosing attention-deficit hyperactivity disorder (adhd) using artificial intelligence: a clinical study in the uk. Front Psychiatry 14, 1164433 (2023)
7. Chen, W., Fan, Y., Zhou, J.: Depression prediction using machine learning algorithms. Applied and Computational Engineering 5(1), 379–385 (2023)
8. Choi, K.S., Kim, S., Kim, B.H., Jeon, H.J., Kim, J.H., Jang, J.H., Jeong, B.: Deep graph neural network-based prediction of acute suicidal ideation in young adults. Scientific Reports 11(1), 15828 (2021)
9. Downs, N.S., Alderman, T., Schneiber, K., Swerdlow, N.R.: Treat and teach our students well: College mental health and collaborative campus communities. Psychiatric Services 67(9), 957–963 (2016)
10. Emran, A.N.W., Islam, A.B.M.A.A.: Mental health across contexts: a cross-dataset study covering medical students, quarantined individuals, and psychiatric disordered subjects. Humanit Soc Sci Commun 12, 1157 (2025)
11. Farooq, S.A., Konda, O., Kunwar, A., Rajeev, N.: Anxiety prediction and analysis- a machine learning based approach. In: 4th International Conference for Emerging Technology (INCET). IEEE (2023)
12. Fki, Z., Ammar, B., Ayed, M.B.: Towards automated optimization of residual convolutional neural networks for electrocardiogram classification. Cognitive Computation 16, 1334–1344 (2024)
13. Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S., Traum, D., Rizzo, S., Morency, L.P.: Daic-woz: Distress analysis interview corpus. Dataset Publication 1(1), 1–10 (2017), multimodal interview data (audio, video, text) for depression detection
14. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: NeurIPS (2017)
15. Hernández-Lorenzo, L., Hoffmann, M., Scheibling, E., List, M., Matías-Guiu, J.A., Ayala, J.L.: On the limits of graph neural networks for the early diagnosis of alzheimer's disease. Scientific Reports 12(1), 17632 (2022)

16. Huang, Y., Chung, A.C.S.: Disease prediction with edge-variational graph convolutional networks. Medical Image Analysis 77, 102375 (2022)
17. Koppe, G., Meyer-Lindenberg, A., Durstewitz, D.: Deep learning for small and big data in psychiatry. Neuropsychopharmacology 46(1), 176–190 (2021)
18. Korpershoek, H., Canrinus, E.T., Fokkens-Bruinsma, M., de Boer, H.: The relationships between school belonging and students' motivational, social-emotional, behavioural, and academic outcomes in secondary education: a meta-analytic review. Research Papers in Education 35(6), 641–680 (2019)
19. Lei, D., Qin, K., Pinaya, W.H.L., Young, J., Van Amelsvoort, T., Marcelis, M., Donohoe, G., Mothersill, D.O., Corvin, A., Vieira, S., Lui, S., Scarpazza, C., Arango, C., Bullmore, E., Gong, Q., McGuire, P., Mechelli, A.: Graph convolutional networks reveal network-level functional dysconnectivity in schizophrenia. Schizophrenia Bulletin 48(4), 881–892 (2022)
20. Leigh-Hunt, N., Bagguley, D., Bash, K., Turner, V., Turnbull, S., Valtorta, N., Caan, W.: An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health 152, 157–171 (2017)
21. Lewis, R., Ghandeharioun, A., Fedor, S., Pedrelli, P., Picard, R.W., Mischoulon, D.: Mixed effects random forests for personalised predictions of clinical depression severity. arXiv abs/2301.09815 (2023)
22. Li, Z., Li, W., Wei, Y., Gui, G., Zhang, R., Liu, H., Chen, Y., Jiang, Y.: Deep learning based automatic diagnosis of first-episode psychosis, bipolar disorder and healthy controls. Computerized Medical Imaging and Graphics 89, 101882 (2021)
23. Losada, D.E., Crestani, F., Parapar, J.: erisk dataset: Early risk prediction on the internet. Dataset Publication 1(1), 1–10 (2017), social media data for early detection of depression and suicidal ideation
24. Lovibond, P.F., Lovibond, S.H.: Depression Anxiety Stress Scales (DASS-21). Psychology Foundation of Australia, Sydney, Australia (1995)
25. Lyu, H.: Application of machine learning on depression prediction and analysis. Applied and Computational Engineering 5(1), 712–719 (2023)
26. Nadeem, M., Rashid, J., Moon, H., Dosset, A.: Machine learning for mental health: A systematic study of seven approaches for detecting mental disorders. In: 2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC). pp. 1–6 (2023)
27. Namdari, S.: Mental health data. Dataset Publication 1(1), 1–5 (2023), dataset including mental health indicators for the general population
28. Radhika, C.: Prediction of mental health instability using machine learning and deep learning algorithms. Journal of Computer Science and Applications 15(1), 47–58 (2023)
29. Rashid, B., Calhoun, V.: Towards a brain-based predictome of mental illness. Human Brain Mapping 41(12), 3468–3535 (2020)
30. Razavi, M., Ziyadidegan, S., Jahromi, R., Kazeminasab, S., Baharlouei, E., Janfaza, V., Mahmoudzadeh, A., Sasangohar, F.: Machine learning, deep learning and data preprocessing techniques for detection, prediction, and monitoring of stress and stress-related mental disorders: A scoping review (2023), preprint, JMIR Publications Inc.
31. Saeed, Q.B., Ahmed, I.: Early detection of mental health issues using social media posts (2025), arXiv preprint arXiv:2503.07653
32. Song, H., Shin, J., Lee, H., Park, J.: A simple and flexible modeling for mental disorder detection by learning from clinical questionnaires. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics (2023)
33. Soria, C., Arroyo, Y., Torres, A.M., Redondo, M.Á., Basar, C., Mateo, J.: Method for classifying schizophrenia patients based on machine learning. Journal of Clinical Medicine 12(13), 4375 (2023)

34. Tank, C., Mehta, S., Pol, S., Katoch, V., Anand, A., Jaiswal, R., Shah, R.R.: Su-roberta: A semi-supervised approach to predicting suicide risk through social media using base language models (2024), arXiv preprint arXiv:2412.01353
35. Venkatapathy, S., Votinov, M., Wagels, L., Kim, S., Lee, M., Habel, U., Ra, I.H., Jo, H.G.: Ensemble graph neural network model for classification of major depressive disorder using whole-brain functional connectivity. Frontiers in Psychiatry 14, 1125339 (2023)
36. Wang, R., Chen, Y., Chen, B., Li, T., Harari, G., Campbell, A.: StudentLife Dataset: Assessing Mental Health, Academic Performance and Behavioural Trends of College Students via Smartphones. Dartmouth College, Hanover, NH, USA (2014), smartphone sensor and survey data on student mental well-being
37. Xia, Y., Liu, L., Dong, T., Chen, J., Cheng, Y., Tang, L.: A depression detection model based on multimodal graph neural network. Multimedia Tools and Applications 83(23), 63379–63395 (2024)
38. Yang, M.B., Liu, H.Y., Wang, Z.Y., Zhang, K.C., Xi, M., Zhang, W.X., Si, S.B.: A cross-task fnirs framework for multi-class mental health classification using adaptive stacking of temporal and network topological features. Biomedical Signal Processing and Control 112, 108559 (2026)
39. Zhang, Z., Meng, Q., Jin, L., Wang, H., Hou, H.: A novel eeg-based graph convolution network for depression detection: incorporating secondary subject partitioning and attention mechanism. Expert Systems with Applications 239, 122356 (2024)

Seifeddine Mechti received his Ph.D. in Computer Science from the University of Tunis, where he was a member of the LARODEC Laboratory (Operational Research, Decision Making and Process Control Laboratory). He is also affiliated with the MIRACL Laboratory and the Arabic Natural Language Processing Research Group (ANLP-RG) at the University of Sfax, Tunisia. He is currently a faculty member at the Faculty of Economic Science and Management of Sfax, University of Sfax, Tunisia. His research interests include natural language processing, author profiling, author attribution, authorship identification, sentiment analysis, and machine learning.

Anis Saddoud is a researcher and faculty member at the High Institute of Sport and Physical Education of Sfax, University of Sfax, Sfax 3000, Tunisia. He is affiliated with the Research Laboratory on Education, Motricity, Sport and Health (EM2S, LR19JS01). His research focuses on sports science, exercise physiology, sleep deprivation, cognitive performance, and decision-making in elite athletes. He has contributed to several publications in sports medicine and exercise science journals.

Azza Allouch was born in Sfax, Tunisia, in 1991. She received her master's degree from the National School of Electronics and Telecommunication of Sfax in 2016. She pursued her Ph.D. in cooperation with the Faculty of Mathematical, Physical and Natural Sciences of Tunis (FST), University of El Manar, Tunisia, and the Robotics and Internet-of-Things (RIOTU) Laboratory at Prince Sultan University, Riyadh, Saudi Arabia. Her research interests include machine learning, blockchain, cybersecurity, and unmanned aerial vehicles (UAVs).

Salma Sassi received her Ph.D. in Computer Science from INSA Lyon, France, and Manouba University, Tunisia, in 2009. She is an assistant professor of Computer Science and a researcher affiliated with the Department of Computer Engineering at Ecole Centrale

d'Electronique (ECE), Lyon, France. Her research interests include knowledge representation and visualization, semantic annotation and indexing, ontology enrichment, healthcare information systems, and big data. She has collaborated extensively with the LIUPPA laboratory at the University of Pau and the Adour Region, France.

Richard Chbeir received his Ph.D. in Computer Science from the Institut National des Sciences Appliquées de Lyon (INSA Lyon), France, in 2001, and his Habilitation degree in 2010 from the University of Burgundy, France. He is currently a Full Professor in the Computer Science Department at the University of Pau and the Adour Region (UPPA), France, where he leads the LIUPPA laboratory (Computer Science Laboratory, UR 3000) and directs the Semantics & Privacy in Digital Ecosystems Research group (SPiDER). He is also the head of the OpenCEMS industrial chair and the creator of the MEDES Conference. His research covers data management, information retrieval, data semantics, access control, and digital ecosystems.

Hajer Brahmi is a researcher affiliated with the REGIM-Lab (REsearch Groups in Intelligent Machines) at the National Engineering School of Sfax (ENIS), University of Sfax, Sfax 3038, Tunisia, and with the Department of Management IT at the Higher Business School of Sfax (ESCS), University of Sfax, Tunisia. Her research interests lie in intelligent systems, machine learning, and information management.

Boudour Ammar is a researcher and faculty member at the REGIM-Lab (REsearch Groups in Intelligent Machines), National Engineering School of Sfax (ENIS), University of Sfax, Sfax 3038, Tunisia. Her research focuses on intelligent machines, computer vision, pattern recognition, and machine learning. She has been involved in multiple national and international research projects related to artificial intelligence and image processing.

Mehdi Chlif is a researcher jointly affiliated with the Faculty of Mathematical, Physical and Natural Sciences of Tunis, University of El Manar, Tunisia, and with the Research Laboratory on Sports Performance Optimisation at the National Center of Medicine and Science in Sports (NCMSS), Tunis, Tunisia. His research interests span sports physiology, exercise science, respiratory function, sleep physiology, and performance optimization in athletes. He has co-authored numerous publications in peer-reviewed international journals in the fields of sports medicine and exercise science.

Received: July 25, 2026; Accepted: October 6, 2026.