ChatGPT

Ensemble Machine Learning for Student Stress and Depression Detection: Recent Advances and Proposed Approach

Introduction

Stress and depression among students have escalated to critical levels, significantly impacting academic performance, well-being, and society at large [1]. Surveys indicate that about 40% of college students suffer severe depression interfering with daily functioning, and over 60% experience overwhelming anxiety [2]. The COVID-19 pandemic further exacerbated student stress by disrupting routines and forcing online learning, resulting in elevated stress levels and a mental health crisis on campuses [3]. **Stress has even been labeled the “epidemic of the 21st century,”** underscoring the urgency for better monitoring and mitigation strategies [4]. Early detection and intervention are crucial, as unmanaged academic stress correlates with poor mental health outcomes [5].

Recent technological advances offer hope: **wearable sensors** (e.g. smartwatches with heart rate or electrodermal activity monitors) and **mobile phones** can passively capture physiological and behavioral indicators of stress in real time [6]. Meanwhile, **psychological questionnaires** and self-report surveys (e.g. perceived stress scales) provide subjective measures of distress that are cost-effective to collect [7]. Both data sources have complementary strengths – physiological signals provide objective continuous data but require devices and careful noise handling, whereas surveys capture psychosocial context but rely on user input [7]. The fusion of these data via machine learning (ML) can enable comprehensive stress detection systems for students.

In particular, **ensemble machine learning techniques** have gained traction for improving prediction robustness and accuracy in mental health analytics. Ensemble models like Random Forests, Gradient Boosted Trees (e.g. XGBoost), and stacking or voting classifiers can leverage diverse feature sets and base learners to better capture the multifaceted nature of stress [8] [9]. Compared to single classifiers, ensembles often achieve higher reliability, especially when combining physiological and psychological features or integrating outputs of both classical ML and deep learning models [10] [11]. Deep learning (DL) approaches (e.g. CNNs, LSTMs) have also been explored for stress and depression detection, sometimes in hybrid with ML ensembles. However, DL models can demand large datasets and risk overfitting in small-cohort studies, while simpler ML models can perform surprisingly well on carefully engineered features [12] [11]. A comparative analysis of recent works shows a mixture of successes: in some cases SVMs or Random Forests outperform deep neural networks on stress datasets [12], whereas in other cases hybrid or deep models achieved state-of-the-art results given sufficient data and preprocessing [11].

This paper focuses on detecting mental stress (and related depression/anxiety) in student populations using ensemble learning on publicly available datasets (such as WESAD, Stress-Lysis, etc.). We present a **backward chronological review from 2025 to 2022**, highlighting key studies on student stress/depression detection via both survey-based and wearable-sensor-based approaches. We discuss the datasets utilized, performance trends, and real-world deployment challenges (e.g. generalizability, noise, privacy). **Finally, we propose a novel ensemble-based framework** that integrates multimodal student stress data with advanced preprocessing and dynamic fusion, addressing gaps identified in the literature. This introduction and literature review establish the context and rationale for our proposed approach, aimed at making a distinctive contribution to student stress detection research and deployment.

Literature Review (2022–2025)

2025: Toward Integrated and Interpretable Models

By 2025, research has pivoted toward more interpretable, comprehensive models for student mental health. For example, Roy *et al.* (2025) introduced an interpretable hierarchical model (I-HOPE) to predict college student mental health using a five-year longitudinal mobile sensing dataset [13]. By mapping raw behavioral signals into higher-level features, their two-stage model achieved about **91% accuracy**, far exceeding baseline methods (~60–70%) [14]. Such results underscore the value of long-term, multimodal data in capturing stress and depression patterns. Another significant 2025 contribution is the application of **multimodal emotion recognition** (e.g. facial expressions combined with ML) for stress monitoring [15] [16], although these are still emerging.

Crucially, a late-2024 systematic review (published in early 2025) assessed dozens of ML models for detecting depression, anxiety, and stress in undergraduates [17] [18]. It found that **over 30 studies reported accuracies above 70%**, with some claiming very high performance (stress detection accuracy ranging from 63% up to 100% in certain lab settings) [18]. However, **47 out of 48 studies had only internal validation**, and the review rated the evidence quality as *very low* [19]. This highlights a real-world deployment issue: models often lack external testing, risking overfitting to specific student cohorts. The review concluded that while ML models show promise for student mental health assessment, researchers must improve generalizability and conduct more rigorous out-of-sample validations [19]. In summary, early 2025 literature stresses the importance of interpretability, long-term data, and external validation as this field matures.

2024: Hybrid Ensembling and Multimodal Fusion

Research in 2024 emphasized hybrid ensemble designs and combining data modalities. Singh *et al.* (2024) proposed a **hybrid deep learning ensemble** for wearable-based stress recognition, blending a CNN-LSTM network with classical classifiers [11]. Tested on the popular WESAD wearable stress dataset, their approach attained about **90.45% accuracy**, outperforming previous machine learning and deep learning results on WESAD [11]. This indicates that carefully crafted hybrid models (e.g. CNN+LSTM for feature extraction, followed by an ensemble classifier) can capture complex stress patterns from physiological signals. Another 2024 study by Singh and colleagues collected a new **Stress-Lysis** questionnaire dataset from 843 college students, with 28 questions covering emotional, physical, and academic factors [20]. Using traditional ML algorithms on this survey data, they found that a Support Vector Machine achieved the highest accuracy (~95% for binary stress vs. non-stress) [21]. This result, on a large student sample, demonstrates that **survey-based features alone can be highly predictive** when using ensemble-friendly algorithms (SVM, Random Forest, etc.) and sufficient data. It also reinforces the feasibility of non-invasive, cost-efficient stress detection via psychological data [7].

Multimodal fusion was another theme in 2024. Researchers explored combining **wearable sensor data with contextual or self-report data** to improve robustness. Awada *et al.* (2023 late/2024) tackled the nuanced problem of distinguishing *eustress* (positive stress) from *distress* using an ensemble approach. They trained an **XGBoost (gradient boosting) model** on both physiological signals (e.g. heart rate, skin response) and behavioral features, achieving F1-scores of **83.4% for eustress and 78.8% for distress** detection [22]. Notably, their model used a 30-second window of sensor data and benefitted from feature-level fusion; furthermore, customizing models by gender improved accuracy by 2–4% [23]. This suggests that *tailored ensemble models* and multimodal data can capture subtle differences in stress appraisal. In parallel, new multimodal datasets like **EmpathicSchool (2022–24)** emerged, providing synchronized data (e.g. facial expressions, physiological signals) under different student stress conditions [24]. Such datasets have been invaluable for training and evaluating fused-data algorithms. Overall, the year 2024 solidified that hybrid ensembles and multimodal inputs (sensor + survey, or physiological + behavioral) yield superior performance in student stress detection, provided that issues of data imbalance and feature selection are carefully managed (e.g. using SMOTE and feature ranking as done in some studies [25]).

2023: Ensemble Techniques and Wearable Innovations

The year 2023 saw numerous advances in ensemble learning techniques applied to both wearable and survey-based stress detection. A highlight was the work of Vos *et al.* (2023), who confronted the **generalization problem** in wearable stress models. They noted that models trained on small single-study datasets (often <50 subjects) fail to generalize to new data [26] [10]. To address this, Vos *et al.* merged data from six public stress datasets (including WESAD, SWELL, and others) and even synthesized additional samples to create a large training corpus [27] [28]. They then built an **ensemble of gradient boosting and a neural network**. The resulting model achieved about **85% accuracy on completely unseen validation data**, a **25% improvement** over single-dataset models [10]. This study underscores that ensemble methods, combined with data fusion across studies, can significantly boost real-world performance and mitigate overfitting [10]. It also highlights the lack of large labeled stress datasets and the creative steps (data synthesis, cross-dataset training) needed to overcome that challenge [26].

Wearable sensor-based detection saw other ensemble innovations in 2023. Rashid *et al.* (2023) introduced **SELF-CARE, a context-aware ensemble framework** for robust stress classification on wearable devices [29]. Their approach tackles two key deployment issues: sensor noise and changing conditions. SELF-CARE uses a learned *noise context classifier* to gauge sensor signal quality, then **selectively fuses sensor inputs across an ensemble of models** based on that context [30] [31]. In effect, it dynamically weights different sensor modalities (e.g. accelerometer vs. EDA) depending on which is most reliable at the moment. On the WESAD dataset, SELF-CARE achieved **94.1% accuracy** for binary stress detection using wrist-worn sensors (and similarly 93.7% with chest-worn sensors), significantly outperforming static models [32]. This demonstrates the power of *dynamic ensembling* and context-aware sensor fusion in improving accuracy and resilience to noise for wearables. Another study by Zhu *et al.* (2023) specifically focused on wrist-worn **electrodermal activity (EDA)** signals for stress detection [6]. They evaluated five classic ML classifiers on four public EDA datasets, finding that a properly tuned **SVM classifier** reached **92.9% accuracy** (stress vs. no-stress) – the best among the models tested [12]. This result, along with others in 2023, indicated that traditional ML models (SVM, KNN, etc.) can still excel, especially with careful feature selection and when data is limited, whereas deep neural networks might not always have an advantage [12]. At the same time, other researchers did deploy deep learning: for instance, some works used convolutional neural nets on chest ECG or wearable motion data, and even explored **stacked generalization (stacking)**. In one case, a stacking ensemble on chest-EDA features (with logistic regression and LDA base learners) reportedly achieved near **99% accuracy on WESAD** [25] – though such a high figure likely reflects an optimized lab setting and underscores the caution mentioned earlier about overly optimistic internal results.

Beyond wearables, ensemble methods were also applied to *academic data and surveys* in 2023. Anand *et al.* (2023) collected survey responses on academic workload and habits from students post-pandemic and built an ensemble academic stress classifier [33] [34]. By using oversampling to balance classes and combining classifiers (Decision Tree, Random Forest, AdaBoost, Gradient Boosting) in an ensemble, they achieved **93.5% accuracy (F1≈93.1%)** in classifying students into high, moderate, or no-stress categories [8]. This study, conducted in an American university context, reflects performance trends in ensemble models on self-reported data: with sufficient feature engineering (e.g. capturing sleep hours, screen time, assignment deadlines) and ensemble learning, very high accuracy is attainable on training populations [35]. However, the authors also note challenges such as biased self-report distributions (few "no-stress" labels, requiring oversampling) and emphasize cross-validation to ensure stability [36] [37].

In summary, 2023 provided a rich array of evidence that ensemble approaches – whether combining multiple algorithms, multiple datasets, or multiple sensor modalities – markedly improve stress detection performance. Both classical ML ensembles and hybrid ML/DL models were validated. Yet, the variance in reported accuracies (ranging from ~85% up to ~99%) and the reliance on lab datasets highlighted the ongoing need for standardized evaluation, larger diverse data, and external test scenarios to truly benchmark these systems for student populations.

2022: Foundational Studies and Emerging Techniques

Foundational work from 2022 established the basis for the rapid advances in subsequent years. A comprehensive survey by Arsalan *et al.* (2022) reviewed both subjective and objective stress detection methods, spanning wearable sensor signals (EEG, ECG, GSR, EMG, PPG, etc.) and non-wearable indicators (pupil dilation, speech analysis, smartphone usage, posture, etc.) [38] [39]. This survey highlighted that stress manifests through a variety of physiological and behavioral changes, and it advocated for **multimodal AI approaches** to integrate these heterogeneous data streams [40]. The notion of combining different modalities and sources of data – now a common theme – was seeded by such early reviews. In terms of student-focused research, 2022 saw initial machine learning studies targeting academic stress. For example, Bisht *et al.* (2022) built one of the early ML models to predict stress in **school students** in India [41]. They used features like study hours and performance metrics, applying algorithms such as Random Forest and achieving promising accuracy for identifying at-risk students. Similarly, Sinha & Sriram (2022) explored an "intelligent student stress prediction" system using ML in an educational context [42], underlining growing global interest in applying AI to student mental health.

On the wearable side, 2022 introduced new datasets and prototype systems that would shape later research. **Hosseini *et al.*** released the *EmpathicSchool* dataset, a multimodal benchmark with synchronized physiological signals and facial expression data collected under controlled stress conditions in a school setting [24]. This dataset provided a testbed for emotion recognition and stress detection algorithms in a student-like environment, fueling research into video-based and physiological signal fusion. Other works dealt with specialized contexts: Rachakonda & Bipin (2022) proposed a machine learning system for early prediction of PTSD using Internet-of-Medical-Things sensors [43], and Nath *et al.* (2021/2022) developed a smart wristband framework measuring cortisol hormone alongside ML classifiers for stress – though they noted the **high cost and limited accessibility** of such biosensors as barriers to widespread use [44]. These efforts collectively underscored real-world issues of data quality, device availability, and participant diversity that needed to be solved. By the end of 2022, the community had a clearer understanding that **no single modality or model would suffice** for robust stress detection: instead, ensemble strategies, data fusion, and context adaptation were the way forward. This realization set the stage for the ensemble-centric innovations that followed from 2023 onward.

Proposed Contribution and Novel Approach

While substantial progress has been made, the literature review above reveals persisting gaps in detecting student stress and depression. **Generalizability** remains a challenge – many models excel on a specific dataset or population but falter on new data [10] [19]. **Noise and context variability** in wearable sensor data can degrade accuracy if not properly addressed [29]. Furthermore, most studies focus on either physiological signals or self-reported data in isolation, whereas a combination could leverage the strengths of each. To tackle these challenges, we propose a unique ensemble-based framework with the following novel features:

- **Multimodal Data Fusion with Late Ensemble:** Our approach will integrate **wearable sensor streams (e.g. heart rate, EDA, motion)** with **periodic self-report surveys (e.g. weekly stress questionnaires)**. We propose a two-tier ensemble: first, separate models learn from each modality (e.g. a Gradient Boosting classifier on sensor-derived features, and a Random Forest or Transformer on survey responses). Then, a **meta-learner ensemble** performs *late fusion* of the modality-specific predictions. This design allows the system to still operate if one data source is missing and to capitalize on both objective and subjective indicators of student stress.

- **Dynamic Context-Aware Ensembling:** Drawing inspiration from SELF-CARE [30], we introduce a context-aware mechanism that adjusts the ensemble's weighting based on real-time data quality and context. For instance, a lightweight CNN could estimate sensor signal quality or detect if the student is in class, at rest, or exercising. The ensemble will **dynamically down-weight noisy or context-mismatched signals**, ensuring robust performance in daily student life (where movement artifacts or phone interruptions may occur). This kind of *hybrid dynamic ensembling* is novel in student stress detection, as previous works either did static fusion or context-aware fusion on only wearables; we will extend it across heterogeneous data streams.

- **Advanced Preprocessing and Feature Engineering:** To push performance further, our framework will incorporate improved preprocessing for wearable signals – for example, **wavelet-based noise reduction** and **artifact removal** for physiological data (removing motion artifacts from EDA/PPG signals, as in Fig. 1 of Almadhor *et al.* [45]). We will also explore generating new features through data fusion (e.g. correlating stress levels with academic calendar events). These steps aim to address the data quality issues noted in many prior studies and give our ensemble models cleaner and more informative inputs.

- **Evaluation on Multiple Public Datasets and Real-world Deployment Consideration:** Uniquely, we plan to train and evaluate our ensemble on **multiple publicly available datasets simultaneously** – for example, combining WESAD (physiological stress), SWELL or StudentLife (academic/work stress data), and a survey dataset like Stress-Lysis. This will follow the multi-dataset training strategy proven effective by Vos *et al.* [27] [10], and will allow us to perform rigorous **cross-dataset validation**. By testing the model's performance on entirely unseen datasets (and, if possible, a small pilot deployment with new student data), we will demonstrate superior generalization and highlight any domain shifts. Deployment considerations such as battery usage, privacy of student data, and integration into a campus well-being platform will also be discussed, making our contribution not only algorithmically innovative but practically relevant for an ICDM 2025 audience.

In summary, the proposed approach distinguishes itself by fusing wearable and survey-based stress indicators in a dynamic ensemble that adapts to context and noise. We aim to deliver a **technically sound and comprehensive solution** that improves accuracy and reliability of student stress detection in real-world settings. By addressing the key gaps identified in recent literature – modality integration, noise/context handling, and generalizability – this contribution could significantly advance the state of the art and provide a pathway toward deploying effective stress monitoring systems for students. The next sections of the paper will detail the methodology of our proposed system and the experiments to validate its performance relative to existing methods.

1 2 13 14 Predicting and Understanding College Student Mental Health with Interpretable Machine Learning
https://arxiv.org/html/2503.08002v1

3 5 Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups, and COVID-19 - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC9169886/

4 22 23 A New Perspective on Stress Detection: An Automated Approach for Detecting Eustress and Distress - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC11485284/

6 12 Stress Detection Through Wrist-Based Electrodermal Activity Monitoring and Machine Learning - PubMed
https://pubmed.ncbi.nlm.nih.gov/37022004/

7 24 41 42 43 44 Machine Learning Algorithms for Detecting Mental Stress in College Students
https://arxiv.org/html/2412.07415v1

8 33 34 35 36 37 Enhancing Diagnostic Decision-Making: Ensemble Learning Techniques for Reliable Stress Level Classification - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC10670633/

9 10 26 27 28 Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices - PubMed
https://pubmed.ncbi.nlm.nih.gov/38048895/

11 Hybrid deep learning model for wearable sensor-based stress recognition for Internet of Medical Things (IoMT) system | Request PDF
https://www.researchgate.net/publication/375420360_Hybrid_deep_learning_model_for_wearable_sensor-based_stress_recognition_for_Internet_of_Medical_Things_IoMT_system

15 Automatic Stress Recognition Using FACS from Prominent Facial ...
https://www.semanticscholar.org/paper/c1fee566d7c0d8b81583901442df2938ed0fd8ba

16 Predictions of College Students' Mental Stress using Machine ...
https://www.researchgate.net/publication/371431081_Predictions_of_College_Students'_Mental_Stress_using_Machine_Learning_Algorithms

17 18 19 SciELO Brazil - How do machine learning models perform in the detection of depression, anxiety, and stress among undergraduate students? A systematic review How do machine learning models perform in the detection of depression, anxiety, and stress among undergraduate students? A systematic review
https://www.scielo.br/j/csp/a/J9PYjQCZDb7fdzQHY68nGZP/abstract/?lang=en

20 21 [2412.07415] Machine Learning Algorithms for Detecting Mental Stress in College Students
https://ar5iv.labs.arxiv.org/html/2412.07415v1

25 45 Efficient Feature-Selection-Based Stacking Model for Stress Detection Based on Chest Electrodermal Activity - PubMed
https://pubmed.ncbi.nlm.nih.gov/37571448/

29 30 31 32 Nafiul Rashid
https://www.catalyzex.com/author/Nafiul%20Rashid

38 39 40 [2202.03033] Human Stress Assessment: A Comprehensive Review of Methods Using Wearable Sensors and Non-wearable Techniques
https://arxiv.org/abs/2202.03033