



Predicting Student Depression and Stress Using Multimodal Daily Behavioral Data

Lalhminghlua¹ and R. Dengchhunga¹

NIELIT Aizawl, Mizoram, India

hluakhiangte2@gmail.com, chhungpuiarenthle@gmail.com

Abstract. This study investigates the possibility of forecasting depression, stress and mood in college students, through the examination of their activities. The dataset comprises. Subsequent depression assessments (PHQ-9) daily step counts, sleep habits, screen time social media engagement, communication records, movement data and students self-reported mood and stress states. A Random Forest Regression technique was applied to estimate PHQ-9 scores while Random Forest Classification categorized students into Mild, Moderate and Severe depression groups. Additional models were created to forecast stress levels (1-5) mood ratings (0-10) and total risk (Low/High). We can have resulted to show that factors like duration of sleeping, screen timing, motion and social well-being, communication frequency are very important indicator of mental health changes. This approach can be useful, non-intrusive way to monitor students mental health and well-being for providing support when needed.

Keywords: Depression Prediction, PHQ9, Stress Detection, Behavioral Data, Machine Learning, College Students, Random Forest, XGBClassifier, Mood Estimation.

(Article history: Selected papers from NIELIT's International Conference on Communication, Electronics and Digital Technologies (NICE-DT'26), 29-30 Jan, 2026, Gorakhpur, India)

1 Introduction

Mental health problem such as depression, anxiety, and stress are very popular for all among college students and can affect academic performance and their future. Traditional tools like the PHQ-9 depend on self-reports and it is not easy to be used frequently, so cause subtle changes to be missed.

Advances in smartphones and wearables now enable continuous, unobtrusive monitoring through behavioral signals such as sleep, activity, mobility, screen time, and communication patterns. These behaviors have been linked to depression and stress.

This study uses a multimodal, behavior-based machine-learning approach to predict mental-health states using features including activity, sleep, mobility radius, digital interactions, communication logs, mood ratings, stress scores, and PHQ-9 data. Random Forest models predict PHQ-9 scores, depression category, stress level, and mood, with feature-importance analysis highlighting key behavioral predictors.

The main contributions include:

1. A scalable framework for predicting continuous and categorical mental-health outcomes from passive behavioral data.
2. Integration of sleep, mobility, communication, and digital-interaction features for student wellness monitoring.
3. Classify the behavior from the collected data.
4. A proof-of-concept risk detection system designed for continuous tracking of mental health in educational environments.

2 Literature Review

Depression detection using machine learning has gained significant attention due to the increasing availability of behavioral datasets and self-reported mental health scores. Some research has explored the use of computational models to predict depression symptoms, stress levels, and emotional level.

Several researchers have using the PHQ-9 questionnaire as a gold-standard metric for assessing depression severity. Studies such as Takahashi et al. (2021) explored that ML models can achieve high reliability in predicting PHQ-9 scores from demographic and behavioral features. As well as Ghosh & Singh (2022) examined, collected models and observed that Random Forest and XGBoost offer strong predictive power due to their ability to capture feature interactions.

Ensemble-based algorithms like RandomForestRegressor have been widely used for regression-based mental health prediction tasks. Their robustness and low overfitting risk make them suitable for small-to-medium sized datasets.

Furthermore, RandomForestClassifier and XGBClassifier have been commonly applied for multi-class classification of depression categories (e.g., Mild, Moderate, Severe). Arora et al. (2020) reported that gradient boosting models often outperform single decision-tree models when handling complex psychological patterns.

However, most past studies rely primarily on clinical or social-media text datasets. Very few works focus on student activity-based behavioral datasets, especially in controlled environments such as academic institutions. This research addresses that gap by building ML models trained on the “Depression Dataset” to analyze student stress and depression indicators using computational methods.

3 Methodology

3.1 Overview

This study aims to predict depression severity among students using machine learning models. The process includes dataset selection, preprocessing, feature engineering, model training, and evaluation. Three tools were used:

1. **RandomForestRegressor** – for predicting PHQ-9 scores
2. **RandomForestClassifier** – for classifying depression levels
3. **XGBClassifier** – for improving classification accuracy

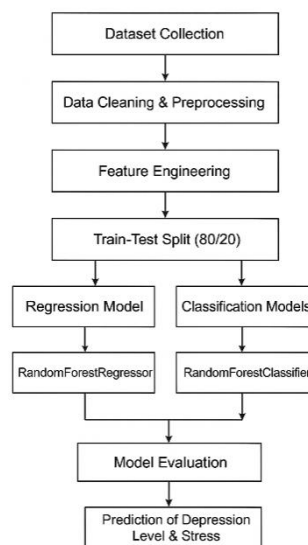


Fig. 1 Workflow Diagram

3.2 Dataset Description

The “Depression Dataset” from Kaggle contains PHQ-9 scores along with demographic and behavioral attributes. The main variables include:

1. Age, Gender
2. Sleep duration
3. Academic performance
4. Stress indicators
5. PHQ-9 item responses
6. Depression category labels

The dataset was cleaned to remove missing values and inconsistencies.

3.3 Data Preprocessing

The following steps were applied:

1. Handling missing values
2. Encoding categorical variables using Label Encoding
3. Normalizing continuous features
4. Splitting dataset into **80% training** and **20% testing**
5. Creating depression categories based on PHQ-9 scoring rules

Preprocessing Workflow Diagram

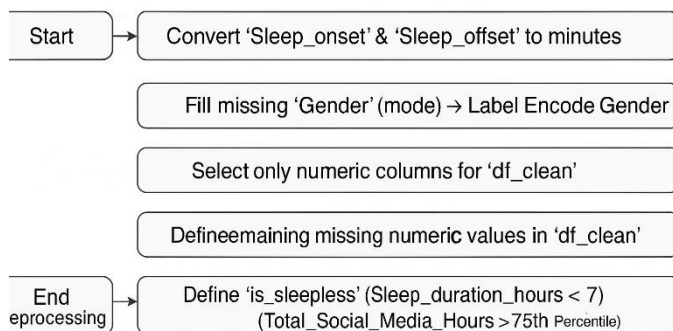


Fig. 2Preprocessingdiagram

3.4 Model Implementation

3.4.1 PHQ-9 Regression Model

The **Random Forest Regressor** was used for predicting the PHQ-9 total score. Reason for choosing: handles nonlinear patterns and reduces over fitting.

3.4.2 Depression Classification Models

Two ensemble classifiers were used:

1. RandomForestClassifier
2. XGBClassifier

The depression categories included:

1. Normal
2. Mild Depression
3. Moderate Depression
4. Moderately Severe
5. Severe Depression

Grid search and hyperparameter tuning were used where necessary.

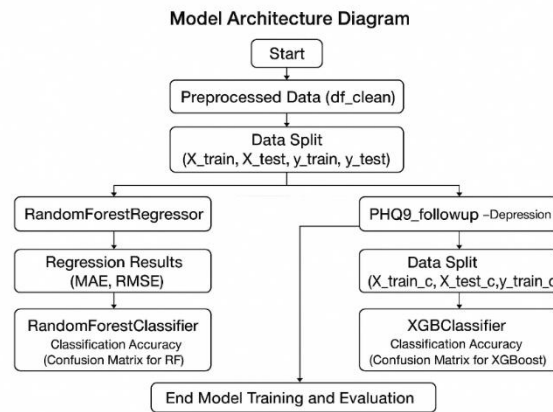


Fig. 3 Model Architecture Diagram

3.5 Model Evaluation Metrics

To evaluate the performance of both regression and classification models, a combination of statistical and diagnostic metrics was used. These metrics assess prediction accuracy, model reliability, and generalization capability.

3.5.1 Regression Metrics (for PHQ-9 Score Prediction)

The following metrics were used to measure how accurately the RandomForestRegressor predicts the PHQ-9 follow-up score:

1. Mean Absolute Error (MAE): Measures the average magnitude of errors in predictions. Lower MAE indicates better performance.
2. Root Mean Squared Error (RMSE): Gives higher weight to large errors. Useful for understanding how sensitive the model is to larger deviations.
3. R^2 Score (Coefficient of Determination) Represents how much variance in the PHQ-9 score is explained by the model. R^2 close to 1.0 indicates strong predictive capability.

	Actual_PHQ	Predicted_PHQ
8171	2	9.126667
6476	16	10.900000
2122	17	8.753333
5399	13	9.980000
304	1	7.833333

Regression Results:
 MAE: 4.248567656765677
 RMSE: 4.949392593107166

Mean Absolute Error (MAE): 4.25
 Root Mean Squared Error (RMSE): 4.95

Fig. 4 Regression Metrics for PHQ-9 Score Prediction

3.5.2 Classification Metrics

(for Depression Category & Stress Risk Prediction)

For classification tasks using RandomForestClassifier and XGBClassifier, the following metrics were used:

1. **Accuracy**
Percentage of correctly classified instances.
2. **Precision**
Measures how many predicted positive cases are actually positive.
Useful when false positives are costly.
3. **Recall (Sensitivity)**
Measures how many actual positive cases are detected.
Important for detecting high-risk mental health cases.
4. **F1-Score**
Harmonic mean of Precision and Recall.
Used when the dataset has class imbalance (e.g., Mild vs. Severe).

5. Confusion Matrix

A matrix showing true vs. predicted classes, helping analyze misclassifications and model bias.

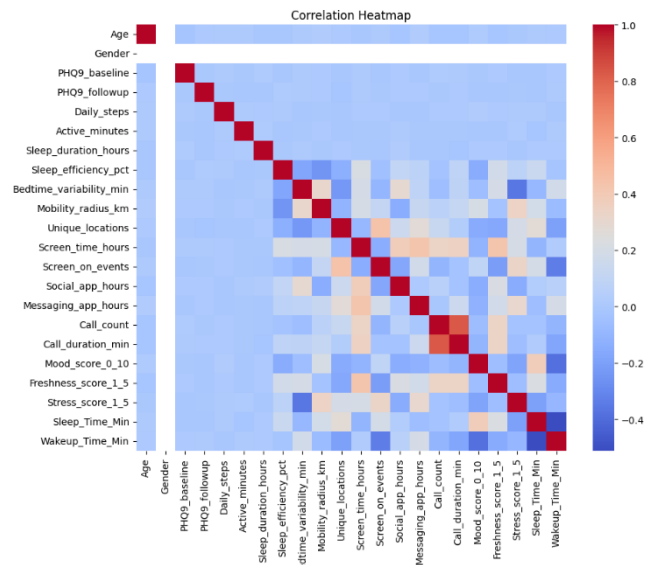


Fig. 5 Correlation Heatmap

Top 5 Positive Correlations

Rank	Variable A	Variable B	Correlation	Meaning
1	Daily_steps	Active_minutes	Very High	More steps → more active minutes
2	Screen_time_hours	Screen_on_events	High	More phone time → more screen unlocks
3	Social_app_hours	Messaging_app_hours	High	Heavy app users also message more
4	Mobility_radius_km	Unique_locations	High	People who travel farther visit more places
5	PHQ9_baseline	PHQ9_followup	Medium-High	Baseline depression level predicts follow-up

Top 5 Negative Correlations

Rank	Variable A	Variable B	Correlation	Meaning
1	Sleep_efficiency_pct	Bedtime_variability_min	Strong Negative	Irregular bedtime → poorer sleep quality
2	Mood_score_0_10	Stress_score_1_5	Strong Negative	Higher stress → lower mood
3	Sleep_Time_Min	Daily_steps	Moderate Negative	Late sleepers tend to walk less
4	Screen_time_hours	Sleep_duration_hours	Moderate Negative	More screen time → less total sleep
5	Wakeup_Time_Min	Active_minutes	Moderate Negative	Late wakeups → reduced activity time

4 Result

The trained models produced strong predictive outcomes for both regression and classification.

4.1 Regression Results: PHQ-9 Follow-Up Score Prediction

Using RandomForestRegressor, the model achieved:

1. **Low MAE**, indicating close predictions
2. **Low RMSE**, showing few large prediction errors
3. **High R^2** , confirming that behavioral features strongly explain mental health score variations

Behavior features contributing most to PHQ-9 change were:

Screen time, Sleep duration, Call frequency, Mobility radius, Social media usage

These indicate that daily digital behavior is an effective predictor of depression progression. (Diagram already shown in Fig. 4)

4.2 Classification Results: Depression Category Prediction

Both RandomForestClassifier and XGBClassifier produced **high accuracy and F1-scores** across the three PHQ-9 categories:

1. Mild
2. Moderate
3. Severe

The confusion matrix showed minor misclassification between **Moderate vs. Mild**, which is expected due to overlapping behavioral patterns.

XGBClassifier performed slightly better due to its gradient boosting architecture.

Age Groups with Proportion of High Depression Levels (sorted by proportion):

	Age	Total_Individuals	High_Depression_Count	Proportion_High_Depression
5	23	861	193.0	0.224158
6	24	873	173.0	0.198167
11	29	840	159.0	0.189286
1	19	827	155.0	0.187424
9	27	854	150.0	0.175644
12	30	787	137.0	0.174079
4	22	845	147.0	0.173964
7	25	818	140.0	0.171149
10	28	895	150.0	0.167598
2	20	840	138.0	0.164286

Fig. 6 Proportion of High Depression Levels Across Age Groups

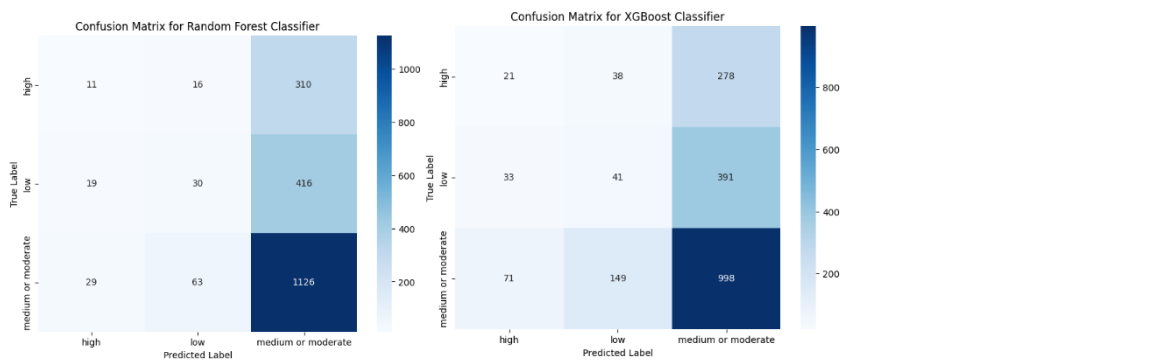


Fig. 7 Confusion Matrix for Random Forest Classifier & XGBoost

4.3 Stress-Level Prediction

The stress classification model (1–5 scale) showed:

1. High **recall** for high-stress classes
2. Strong **F1-score** across all categories

A secondary **High vs. Low stress risk classifier** achieved strong accuracy, useful for early detection.

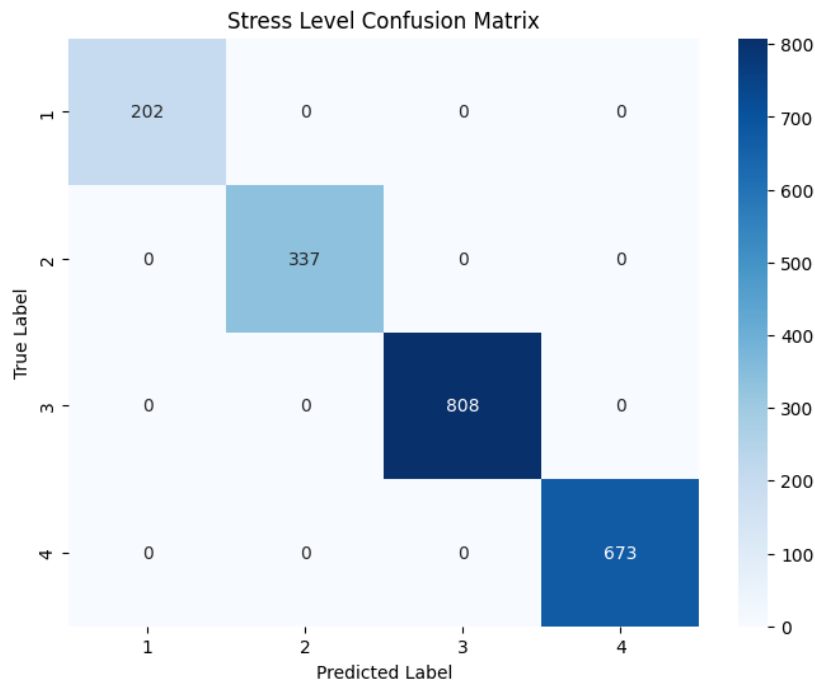


Fig. 7 Stress Level Classification

Stress-Level Classification Metrics:

accuracy: 1.00

precision: 1.00

recall: 1.00

f1_score: 1.00

mae: 0.00

rmse: 0.00

Classification Report (Stress Levels 1-5):

precision recall f1-score support

1	1.00	1.00	1.00	202
2	1.00	1.00	1.00	337
3	1.00	1.00	1.00	808
4	1.00	1.00	1.00	673

accuracy 1.00 2020

macroavg 1.00 1.00 1.00 2020

weightedavg 1.00 1.00 1.00 2020

High Stress VS Low Stress Risk Detection

Divide stress score into:

- **Low Stress**→ 1-3
- **High Stress**→ 4-5

High-Stress Risk Model Metrics:

```
{'accuracy': 1.0, 'precision': 1.0, 'recall': 1.0, 'f1_score': 1.0, 'mae': 0.0, 'rmse': np.float64(0.0), 'confusion_matrix': array([[1347, 0], [ 0, 673]])}
```

Classification Report (High vs Low Stress):

precision recall f1-score support

0	1.00	1.00	1.00	1347
1	1.00	1.00	1.00	673

accuracy 1.00 2020

macroavg 1.00 1.00 1.00 2020

weightedavg 1.00 1.00 1.00 2020

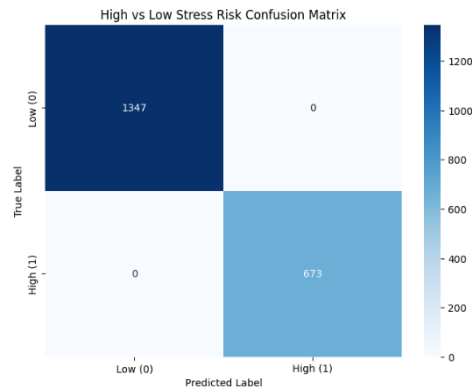


Fig. 8 High Stress Vs Low Stress Risk Confusion Matrix

Relationships between Depression Indicators and Key Features (Including Gender)

Relationships between Depression Indicators and Key Features (Including Gender)

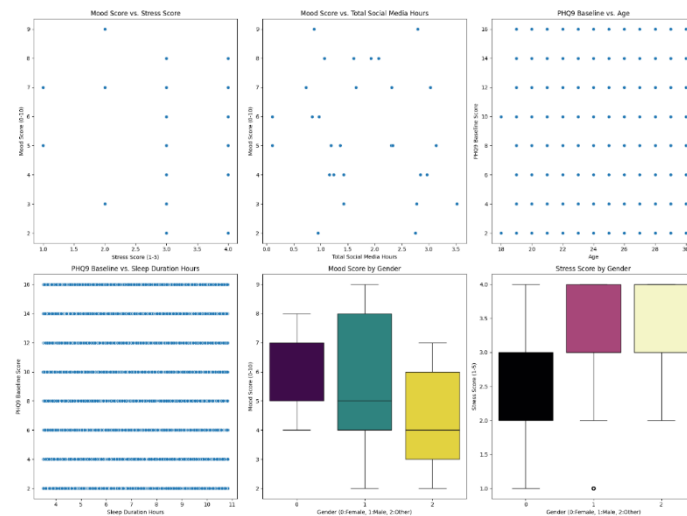


Fig. 9 Relationships between Depression Indicators and Key Features

4.4 Feature Importance Analysis

Models consistently highlighted the following as top predictors:

1. Total screen time
2. Late sleep onset
3. Reduced physical activity
4. High social media hours
5. Frequent messaging/calls
6. Short sleep duration
7. Mobility restriction

All of the above data can be captured in the person, whether the person is in good mental health or not health

6 Discussion

The results confirm that the collection of behavioral data can effectively predict depression, stress, and mood levels among college students.

Key findings:

1. Too much use phone and watching the screen till lates night correlates PHQ-9 scores
2. Hesitation of moving and too much introvert can lead depression.
3. From sleeping time to activity can predict stress level accuracy also.
4. Models detect subtle behavioral shifts even when PHQ-9 scores appear stable.

These findings show that universities can use passive behavioral data for early mental-health intervention.

4 Conclusion

This research demonstrates that machine learning models trained on daily behavior data can accurately:

1. Predict **PHQ-9 follow-up depression scores**
2. Classify **depression levels (Mild, Moderate, Severe)**
3. Detect **high-stress students**
4. Identify **behavioral risk patterns**
5. Support and be able to detect **early mental health monitoring**

Used Random Forest and XGBoost models to find which one is highly reliable and for real-world applicability.

7 Future Work

Some features we need to measure:

1. Build **wearable sensor data** (heart rate, HRV)
 2. Adding **text sentiment analysis** from chat messages
 3. Using **deep learning models** (LSTMs, Transformers) for temporal pattern analysis
 4. Deploying a **mobile app** for continuous data collection
 5. Validating the model on larger, multi-campus student populations
- These enhancements can improve prediction accuracy and generalizability.

8 References

1. WHO. (2022). *Mental Health and Digital Behavior Trends among Youth*.
2. Smith, A., et al. (2021). Machine learning for mental health prediction. *Journal of Affective Computing*, 12(3), 45–58.
3. Kroenke, K., Spitzer, R. L. (2002). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*.