



(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 6, June 2025

⊕ www.ijirset.com ⊠ ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406168|

Lifestyle-Based Student Stress Detection with Real-Time ML Recommendations

Anindita Ghara¹, Arijit Khan², Pranab Mahata³

MCA Student, Department of Computer Applications, Dr. B.C. Roy Engineering College, Durgapur,

West Bengal, India^{1,2,3}

ABSTRACT: Stress has become a critical concern among students, adversely affecting their academic performance and mental health. This study presents a machine learning-based system for early detection of student stress levels using lifestyle-related data, including sleep duration, study hours, physical activity, social interaction, and GPA. A structured dataset comprising 2,000 student records was preprocessed using Isolation Forest for outlier removal, Robust Scaler for normalization, and SMOTE for addressing class imbalance. Four supervised learning algorithms—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression—were implemented and evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Among them, the SVM with an RBF kernel achieved the highest performance, with an accuracy of 96.84%. Novel engineered features, such as the study-to-sleep ratio and stress risk score, significantly enhanced predictive capability. The proposed model is integrated into a real-time Streamlit web application that provides students with stress predictions and personalized lifestyle recommendations. This research highlights the potential of interpretable AI solutions to support student mental wellbeing through proactive, data-driven feedback.

KEYWORDS: Student Stress Detection, Lifestyle Analytics, Machine Learning, Support Vector Machine, Real-Time Prediction, Streamlit Application, Behavioral Modeling, Personalized Recommendations, Educational AI, Mental Health.

I. INTRODUCTION

In recent years, student mental health has emerged as a critical area of concern within academic institutions worldwide. The increasing academic pressure, competitive environments, erratic daily routines, and a lack of emotional support have significantly contributed to elevated stress levels among students. Chronic stress not only hampers academic performance but also leads to long-term psychological and physiological issues, including anxiety, burnout, and depression. As students navigate through their academic careers, the ability to recognize and manage stress becomes essential for maintaining both well-being and academic success.

Despite growing awareness, most current approaches to assessing student stress rely on self-reported surveys, institutional counseling, or subjective judgment, which are often reactive rather than preventive. These traditional methods lack real-time insights, personalization, and scalability— resulting in delayed interventions and overlooked stress symptoms. There exists a pressing need for intelligent, data-driven systems capable of detecting stress levels early and recommending actionable changes.

The advent of machine learning (ML) presents a transformative opportunity to tackle this issue. By analyzing lifestylerelated data such as sleep duration, study hours, physical activity, and social interactions, ML models can uncover hidden patterns associated with varying stress levels. The availability of structured behavioral datasets allows for the development of predictive systems that are objective, scalable, and adaptable across academic settings.

This research introduces a novel approach to student stress detection through supervised machine learning models trained on a structured lifestyle dataset. The system not only predicts the stress level of students (Low, Moderate, High) with high accuracy but also offers real-time, personalized recommendations using a lightweight Streamlit-based web application. This integration of behavioral analytics with intelligent prediction and feedback represents a step toward AI-powered mental wellness support in education.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025 |DOI: 10.15680/IJIRSET.2025.1406168|

II. RELATED WORK

The assessment of psychological stress, particularly among students, has been a longstanding focus of research in psychology and education. Traditional approaches to stress detection often rely on standardized surveys and questionnaires, such as the Perceived Stress Scale (PSS) or the Student Academic Stress Scale. While these tools provide valuable insights, they are limited by subjectivity, non-continuous assessment, and the inability to provide real-time feedback. Moreover, students may underreport or misjudge their stress levels due to stigma or lack of awareness, leading to delayed intervention.

With the advancement of computational technologies, machine learning (ML) has emerged as a powerful tool in mental health analytics. Several recent studies have applied ML techniques to classify stress levels using both physiological signals (e.g., heart rate, EEG, GSR) and behavioral patterns. For example, Shahapur et al. employed Random Forest and Gradient Boosting classifiers on a dataset of over 6,000 students, achieving over 95% accuracy in stress classification based on features like sleep quality and academic performance. Similarly, Arya et al. utilized Artificial Neural Networks (ANNs) to predict stress during academic periods with a test accuracy exceeding 89%.

However, many of these studies either use complex physiological data requiring specialized sensors or depend heavily on retrospective self-reporting. In contrast, our approach focuses on lifestyle-based behavioral data—such as daily hggstudy hours, sleep duration, GPA, and physical activity—which are more readily available and non-intrusive. By leveraging structured tabular data and engineered features like the study-to-sleep ratio and stress risk score, we demonstrate that psychological stress can be effectively predicted using commonly tracked student behaviors.

Furthermore, while most existing systems focus solely on prediction accuracy, our system extends the utility by integrating a real-time recommendation engine within a Streamlit application. This provides personalized, actionable feedback to students, transforming the system from a diagnostic tool into a behavioral assistant. Thus, our work bridges the gap between traditional survey-based stress assessment and AI-enabled real-time mental health support in academic settings, intelligent, data-driven systems capable of detecting stress levels early and recommending actionable changes.

The advent of machine learning (ML) presents a transformative opportunity to tackle this issue. By analyzing lifestylerelated data such as sleep duration, study hours, physical activity, and social interactions, ML models can uncover hidden patterns associated with varying stress levels. The availability of structured behavioral datasets allows for the development of predictive systems that are objective, scalable, and adaptable across academic settings.

This research introduces a novel approach to student stress detection through supervised machine learning models trained on a structured lifestyle dataset. The system not only predicts the stress level of students (Low, Moderate, High) with high accuracy but also offers real-time, personalized recommendations using a lightweight Streamlit-based web application. This integration of behavioral analytics with intelligent prediction and feedback represents a step toward AI-powered mental wellness support in education.

III. METHODOLOGY

This section outlines the methodology adopted for predicting student stress levels using machine learning. It includes details on the dataset, preprocessing techniques, feature engineering, modeling approaches, and evaluation metrics.

Data Acquisition Layer Data Pre-Processing Layer Engineering Layer Engineering Layer Machine Machine Layer Machine Machine Layer Machine Machine

Fig1- System Architecture of Stress Prediction and Personalized recommendations.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406168|

A. Dataset Acquisition Layer -

The dataset used in this study was obtained from Kaggle and consists of 2,000 instances of student lifestyle data, structured in a tabular format.

. It includes the following features:

- Sleep_Hours_Per_Day
- Study_Hours_Per_Day
- Physical_Activity_Hours_Per_Day
- Social_Hours_Per_Day
- Extracurricular_Hours_Per_Day
- GPA (Grade Point Average)

- Stress_Level (target variable: Low, Moderate, High)

- Student ID (excluded from modeling)

B. Data Pre-processing Layer -

To prepare the data for modeling, the following steps were applied:

1. Handling Missing and Duplicate Data-

No missing or duplicate entries were found in the dataset.

2. Outlier Detection-

The Isolation Forest algorithm was used with a contamination rate of 5% to detect and remove 100 outliers, reducing the dataset to 1,900 records.

3. Label Encoding-

The target variable Stress_Level was encoded as integers: High = 0, Low = 1, Moderate = 2.

4. Feature Scaling-

Robust Scaler was used to normalize numerical features. It centers data using the median and scales based on the interquartile range, making it ideal for skewed data with outliers.

5. Class Imbalance Handling-

SMOTE (Synthetic Minority Over-sampling Technique) was applied to balance the distribution of Low, Moderate, and High stress categories.

C. Feature Engineering Layer -

To enhance learning, the following new features were created:

1.Study-to-Sleep Ratio: Captures the imbalance between study time and rest.

-Formula :-

Study-to-Sleep Ratio = Study Hours Per Day Sleep Hours Per Day + €

-Where:

• Study Hours per Day is the daily average time a student spends studying.

• Sleep Hours per Day is the daily average time a student sleeps.

• \in is a small constant (1×10⁻⁵) added to avoid division by zero.

This ratio aims to capture behavioral imbalance that may contribute to psychological stress.

2. Total Active Time: Aggregates time spent on non-academic activities.

-Formula :-

Total Active Time=Extracurricular Hours per Day+Physical Activity Hours per Day+Social Hours per Day

-Where:

- Extracurricular Hours per Day represents time spent on non-academic skill development (e.g., music, arts).
- Physical Activity Hours per Day includes exercise, sports, or fitness-related activities.
- Social Hours per Day reflects time spent interacting socially, either in person or virtually.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406168|

×(4-GPA)

3.Stress Risk Score: A composite score derived as Study-to-Sleep Ratio × (4 - GPA).

- Formula :-

Stress Risk Score =

- Where -

- $\varepsilon = 10^{-5}$ is a small constant added to avoid division by zero.
- GPA is normalized on a 4-point scale.
- A higher score reflects potential overexertion with lower academic return, suggesting elevated stress.
- This composite metric was introduced only if the dataset did not already contain a Stress_Risk_Score column.

After created new features ,check the correlation analysis of each features -

Through correlation analysis and model interpretability techniques, the following features were found to be highly influential in predicting student stress levels:

- Sleep Hours per Day: Negatively correlated with stress. Students sleeping less tended to have higher stress levels.
- Study-to-Sleep Ratio: High ratio (more study, less sleep) strongly indicated high stress.
- Social Hours: Students engaging in regular social interaction showed lower stress levels.
- Physical Activity Hours: Regular physical activity was associated with reduced stress levels.
- Total Active Time (study + extracurricular + physical): When excessive, it correlated with increased stress.

These insights suggest that balanced time management—especially between study, rest, and physical/social activities—is critical for maintaining low stress levels



Fig 2- Correlation analysis heatmap analysis

D. Machine Learning Modeling Layer -

Four supervised machine learning classifiers were used to model stress levels:

-Support Vector Machine (SVM with RBF Kernel): Selected for its ability to model complex non- linear boundaries. It delivered the best performance across all evaluation metrics.

- K-Nearest Neighbors (KNN): Used as a simple, interpretable classifier based on distance similarity.

- Naive Bayes (Gaussian): Included for its simplicity and computational efficiency.
- Logistic Regression: A linear model used for baseline comparison.

The dataset was split using stratified sampling (80% training, 20% testing) to preserve class proportions.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406168|

E. Evaluation Layer - Models were evaluated using:

- Accuracy: Overall proportion of correctly predicted instances.

- Precision, Recall, F1-score: Evaluated per class to measure exactness and completeness.
- Confusion Matrix: Reveals types of classification errors.

- ROC-AUC Curve (One-vs-Rest): Used to visualize and compare classifier performance for each stress category.

Among the evaluated models, SVM (RBF Kernel) achieved the highest accuracy (96.84%), followed by KNN (93.16%), Naive Bayes (90.00%), and Logistic Regression (83.95%).

IV. RESULTS AND DISCUSSION

This section presents the performance evaluation of the four machine learning models used for student stress level classification and interprets the results using visual metrics and feature impact analysis.

A. Model Performance Comparison

Each of the four classifiers—Support Vector Machine (SVM with RBF kernel), K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression—was evaluated using an 80/20 stratified train-test split. The SVM model significantly outperformed the others in terms of accuracy and class-wise balance.

Model Classification Report Comparison (%):

Model	Accuracy	Precision	Recall	F1-Score
SVM (RBF Kernel)	96.84%	96.87%	96.84%	96.83%
K-Nearest Neighbors	93.16%	93.12%	93.16%	93.13%
Naive Bayes	90.00%	90.51%	90.00%	89.81%
Logistic Regression	83.95%	84.00%	83.95%	83.68%

Model Performance Comparison Graph-





B. Confusion Matrix and ROC Curves

Confusion matrices were used to visualize prediction accuracy for each class. The SVM classifier exhibited strong diagonal dominance, indicating excellent class-wise accuracy for Low, Moderate, and High stress levels.

The ROC-AUC curves (One-vs-Rest) provided an additional layer of evaluation, showing the model's ability to separate each class. The SVM achieved an average AUC of **0.98**, demonstrating high discriminatory power across all stress levels.

IJIRSET©2025





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406168|

Confusion Matrices of best model performance(SVM) graph-



Fig 4- Confusion Matrices of best model performance(SVM) graph

ROC-Curve for best model(SVM) model graph-



Fig 5- Roc-curve for svm model(best performance model)

C. Feature Importance and Interpretation

Although SVM is not inherently interpretable, **permutation feature importance** was used to estimate the contribution of each feature. The most influential features included:

- Study-to-Sleep Ratio
- GPA
- Stress Risk Score
- Total Active Time
- Study Hours Per Day

These findings align with psychological literature that associates high academic pressure and poor lifestyle balance with elevated stress levels.



Volume 14, Issue 6, June 2025



[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|





D. Limitations and Real-World Applicability

A key limitation of this work is the reliance on self-reported data, which may contain bias or inaccuracies. Additionally, the dataset may not generalize across diverse student populations without further validation.

Despite these limitations, the model's integration into a Streamlit web application demonstrates real- world applicability. Students can enter their lifestyle habits and receive real-time predictions along with personalized recommendations. This approach transforms the model into a practical AI-driven assistant for promoting student mental well-being.

V. REAL-TIME STRESS PREDICTION AND PERSONALIZED RECOMMENDATION SYSTEM

To enhance the practical applicability of the proposed machine learning system, we integrated the trained Support Vector Machine (SVM) model into a real-time prediction platform using the Streamlit framework. This interface enables students to input their daily lifestyle habits and receive instant feedback on their predicted stress levels along with personalized recommendations.

System Workflow -



Fig 7- System Workflow of Real-Time Stress Prediction and Personalized Recommendation System





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406168|

A. User Input -

The web application accepts user input for key features including:

- Sleep Hours per Day
- Study Hours per Day
- GPA
- Physical Activity Hours Per Day
- Social Hours Per Day
- Extracurricular Hours Per Day

B. Pre-Processing Module -

These inputs are passed through the same preprocessing as used during model training:

Robust scaling

The data is then fed into the trained SVM model, which classifies the stress level as Low, Moderate, or High. The system returns this prediction in real-time.

C. Feature Engineering Module -

After the pre-processing module inputs undergo the same feature engineering steps used during model training to ensure consistency and accuracy. Specifically:

- Derived Features are computed:
 - Study-to-Sleep Ratio: Represents the balance between study time and rest.
 - Total Active Time: Sum of study, physical, and social hours per day.
 - Stress Risk Score: A heuristic indicator calculated from behavioral metrics.

The prediction is generated and displayed in real time.

D. Personalized Recommendation Engine -

Based on the predicted stress level, a rule-based recommendation system provides actionable advice tailored to the user's lifestyle pattern. The logic is defined as follows:

- Low Stress: Encouragement to maintain current routine with suggestions for continued balance (e.g., keep up regular exercise, consistent sleep).
- Moderate Stress: Alerts the student about potential risk and recommends specific behavioral adjustments, such as increasing sleep or reducing study overload.
- High Stress: Flags critical risk and advises immediate interventions, including more rest, physical activity, reduced workload, and seeking peer or professional support.

These recommendations are context-aware and directly mapped to the user's actual input, thereby enabling meaningful and timely behavioral changes.

E. Output Module -

The Output Module acts as the system's final interaction point, delivering interpretable, actionable feedback based on the predicted stress level. After the Support Vector Machine (SVM) model classifies the input into Low, Moderate, or High stress, this module presents:

- Predicted Stress Level: Clearly labeled and color-coded for immediate understanding.
- **Personalized Recommendations**: Generated through rule-based logic, tailored to the predicted stress class and the user's input lifestyle profile.

Derived Metrics Visualization:

- Study-to-Sleep Ratio: Indicates academic-rest balance.
- Total Active Time: Reflects cumulative time spent on daily engagements.
- Stress Risk Score: A composite metric quantifying stress susceptibility.

These outputs are presented via a responsive Streamlit interface, using intuitive visuals such as colored indicators, metric boxes, and expandable panels. The design ensures that users receive not only a classification result but also meaningful, contextual insights to promote behavioral awareness and stress management.

F. Streamlit User Interface and Deployment -

The Streamlit-based interface is lightweight, responsive, and user-friendly. It can be easily deployed on a local machine or hosted online for institutional access. This real-time system bridges the gap between static predictive models and





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025 |DOI: 10.15680/IJIRSET.2025.1406168|

proactive mental health support tools.

D localhost:8501			
	Check Your Stress Level		
	Study Hours per Day		
	0.00		12.00
	Extracurricular Hours per Day		
		4.48	
	0.00		6.00
	Sleep Hours per Day		
		9.97	
	0.00		12.00
	Social Hours per Day		
		6.40	
	0.00		8.00
	Physical Activity Hours per Day		
		3.30	
	0.00		5.00
	CDA .		
	GPA		
	0.00		4.00

Fig7 - User input the value according to their daily lifestyle habits via streamlit UI

🞯 Predicted Stress Level: Moderate	Predict Stress Level		
	oredicted Stress Level: Mo	oderate	

Fig8- Show the User's Stress level according to their given daily lifestyle habits value via streamlit UI

4	Recommendations:
<u></u>	You're in a Moderate Stress zone.
•	Try balancing academics and social life.
•	Add short physical breaks while studying.
•	Try planning your time weekly to reduce anxiety.

Fig9- Show the User's personalized recommendation according to their stress level via streamlit UI





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406168|



Fig10 - Show the user's derived features according to their input value via streamlit UI

VI. CONCLUSION

This study presents a lifestyle-based machine learning approach for detecting student stress levels with high accuracy. By analyzing behavioral factors such as sleep duration, study hours, physical activity, and GPA, the model effectively classifies stress into three levels—Low, Moderate, and High. Among the evaluated classifiers, the Support Vector Machine (SVM) with an RBF kernel demonstrated superior performance, achieving an accuracy of 96.84% and strong generalization across all categories.

The integration of the trained model into a real-time Streamlit application highlights the practical utility of this research. The system not only predicts stress levels but also delivers personalized lifestyle recommendations, making it a useful tool for early intervention in educational environments. Such systems can assist academic counselors and students alike in promoting mental wellness and improving academic outcomes.

Future work will focus on expanding the dataset with more diverse demographics, incorporating emotional or psychological signals (e.g., mood, anxiety), and integrating explainable AI (XAI) techniques such as SHAP or LIME to improve interpretability. Additional deployment avenues, such as mobile applications or institutional dashboards, can further enhance accessibility and impact.

REFERENCES

[1]. www.kaggle.com/datasets/steve1215rogg/student-lifestyle-dataset

[2] 373858231 University student's mental stressdetection using machine learning

[3].382943391_Predicting_the_stress_level_of_students_using_Supervised_Machine_Learning_and_Artificial_neural_network_ANN

[4]. V5ISSUE2/IJRPR22974.pdf

[5].305829v1.full.pdf

[6].Stress-detection-using-machine- learning-and- deep-learning

[7]. Decoding-minds-estimation-of-stress-level-in-students-using- machine learning

[8]. IJST-2023-2951%20.pdf

[9].MalikSS,KhanA.Anxiety,DepressionandStresspredictionamongCollegeStudentsusingMachineLearningAlgorithms. In:2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT). IEEE. 2023.

[10]. RadhikaC,ShraddhaN,VaishnaviP,Shirisha K. Prediction of Mental Health Instability using Machine Learning and Deep Learning Algorithms. Journal of Computer Science and Applications. 2023

[11].Predictive Analysis Of Student Stress Level Using Machine Learning

[12]. Lamb R, Almusharraf N, Choi I, Firestone J, Kavner A, Owens T, et al. Machine learning prediction of mental health strategy selection in school aged children using neurocognitive data. Computers in Human Behavior.

[13]. https://project19.streamlit.app









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com