JOURNAL OF
EMERGING INVESTIGATORS

Article

Early detection of student burnout using data science: a study of behavioral and psychological indicators

Eshaal Baber[1], **Muhammad Talha Malik**[1,2]
[1] Army Public School International Studies, Rawalpindi, Punjab, Pakistan
[2] Quaid-e-Azam University, Islamabad, Islamabad Capital Territory, Pakistan

SUMMARY

While burnout is being increasingly noted in educational contexts, the existing research primarily focuses on professional settings. Rising academic demands have exposed students to high workloads and psychological distress. Academic burnout is defined as exhaustion, disengagement, and low feelings of professional competence, which negatively affect students' well-being, motivation, and subsequent career prospects. In this study, we aimed to examine lifestyle and psychological predictors to help explain burnout among high school and university students in Pakistan, as well as the potential application of machine learning in early detection. We hypothesized that sleep, screen time, and mental exhaustion would each be robust predictors of the risk of burnout, and that predictive modeling has the capacity to classify at-risk students accurately. We used an online survey with 74 high school and university students (ages 12-25) to measure demographics, sleep duration, screen exposure, coping strategies, motivation, and fatigue. We analyzed the data using statistical analyses, and applied the scikit-learn machine learning module in Python for predictive modeling. Overall, sleep duration was negatively related to burnout, while mental fatigue was the strongest positive predictor of burnout. Screen time exposure only had a weak, non-significant association with burnout once other variables were accounted for. Additionally, there were some gender differences, with males exhibiting lower levels of burnout, although effect sizes were modest. Lastly, machine learning predictive models were accurate at classifying risk for students, with particular success from the Random Forest algorithm, demonstrating the potential for predictive analytics in the well-being of students.

INTRODUCTION

The term "burnout" was originally described as a state of mental exhaustion, a detached attitude toward others, and a diminished sense of accomplishment generally associated with occupations that require continuous engagement with individuals (1). In recent years, the subject of burnout has been expanded into the academic space and has become concerningly prevalent in teens. Students engage in structured and intentional activities, which include attending classes, studying for exams, and meeting deadlines (2). Academic burnout is most often expressed through exhaustion from study requirements, detachment or cynicism toward classroom tasks, and diminished competence (3). Research on student burnout has shown that it leads to negative outcomes such as academic inefficacy, absenteeism, diminished motivation, and, in serious cases, dropout (4). Furthermore, personal attributes like perfectionism can put individuals at greater risk of experiencing burnout, while engagement, described as vigor, dedication, and absorption, can act as a protective influence against burnout (5).

Burnout in academic life has implications beyond the immediate academic challenges faced by students, and also has ramifications for mental health outcomes and career trajectory. For instance, research on medical students and residents has demonstrated that burnout correlates with unprofessional behavior (6). Despite schools' and health centers' best efforts to alleviate burnout with wellness programs and work-hour restrictions, high level of burnout persists among students and trainees, as reported in a systematic review of U.S and Canadian medical students (6). These students and trainees often resort to an ineffective survival mode in which they struggle to manage their daily academic and personal responsibilities, and rely on maladaptive coping strategies (6). Therefore, understanding the complexity of student burnout is essential, so that appropriate interventions can be identified, thereby ensuring that current and future student academic success align with well-being.

Most current research describes the symptoms and causes of burnout, with little focus on predictive models or early warning systems that could provide early detection of burnout. This study aimed to move beyond description by applying data analysis and machine learning to predict early signs of student burnout.

Since exhaustion is a key component of burnout, it is reasonable to expect that factors directly correlated with exhaustion, such as lack of sleep, would raise the risk for burnout. Mental fatigue also involves cognitive overload and a decline in mental resources, similar to how burnout leads to reduction in emotional and motivational functioning (7). Based on this reasoning, we hypothesized that reduced sleep duration and higher self-reported mental fatigue would significantly predict burnout risk among students. Our findings support this, showing that both shorter sleep and elevated fatigue scores were strongly associated with higher burnout risk.

We further hypothesized that these specific predictors, when combined in a machine learning-based model, would allow reliable identification of individuals at risk of developing burnout. Consistent with this, our predictive model classified student burnout risk with relatively high accuracy. This highlights the potential of such approaches to enable early detection systems and timely interventions that support student well-being.

https://doi.org/10.59720/25-269

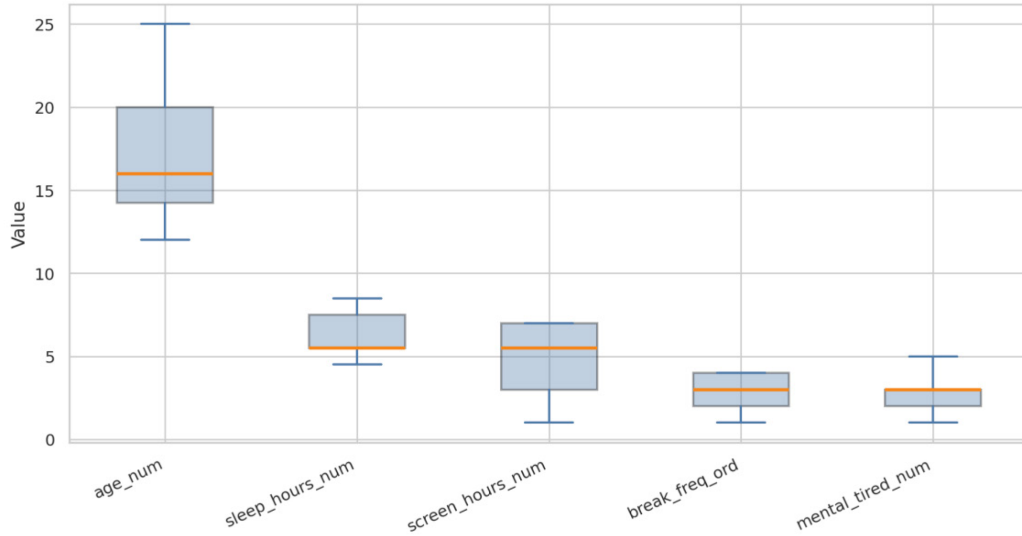

Figure 1: Descriptive statistics of survey responses. Boxplot visualization of summary statistics for survey variables: age *(age_num),* sleep duration in hours *(sleep_hours_num),* daily screen time in hours *(screen_hours_num),* break frequency *(break_freq_ord),* and mental tiredness level *(mental_tired_num).* The x-axis lists the variables, and the y-axis represents their values in original units. For each variable, the central line indicates the median, the box shows the interquartile range (Q1–Q3), and whiskers represent the minimum and maximum values.

RESULTS

Burnout Prevalence

To assess burnout prevalence in a student population, we conducted a survey from Sept 2–11, 2025, following school reopening, when students were readjusting to routines. It asked about general habits and feelings on average, not over a specific time period, to capture typical behaviors. To assess the prevalence of burnout among students and identify typical patterns of psychological strain, we constructed a composite burnout index by z-scoring and averaging four indicators: (inverted) mental tiredness (i.e., the scale was inverted so that higher values indicate greater fatigue), difficulty focusing, feeling overwhelmed, and irritability. We classified participants into low, medium, and high burnout risk based on their index scores (see Methods for cutoff definitions). Participants were between 12 and 25 years of age (Mean (M) = 17.14, Standard Deviation (SD) = 3.47). Of the sampled population, 62.2% were female (n = 46) and 37.8% were male (n = 28), with 66.2% high school students (n = 49) and 33.8% university students (n = 25). The average sleep was 6.42 ±1.35 hours per night, and 41% of respondents reported sleeping fewer than 6 hours per night, well below the recommended 8–10 hours sleep time for adolescents and young adults (8) (**Figure 1**). In contrast, participants reported 4.77 ±2.13 hours of screen exposure, and 42% of respondents indicated they rarely or never took breaks from study or work (**Figure 1**). Mental tiredness was concentrated around the midpoint of the 1–5 scale (M = 2.59, SD = 1.18), with the modal score of 3 **(Figure 1)**. The overall distribution of burnout risk appeared to place a large number of students into moderate to high risk of burnout, particularly those reporting shorter sleep duration, greater screen exposure, and higher levels of mental fatigue **(Figure**

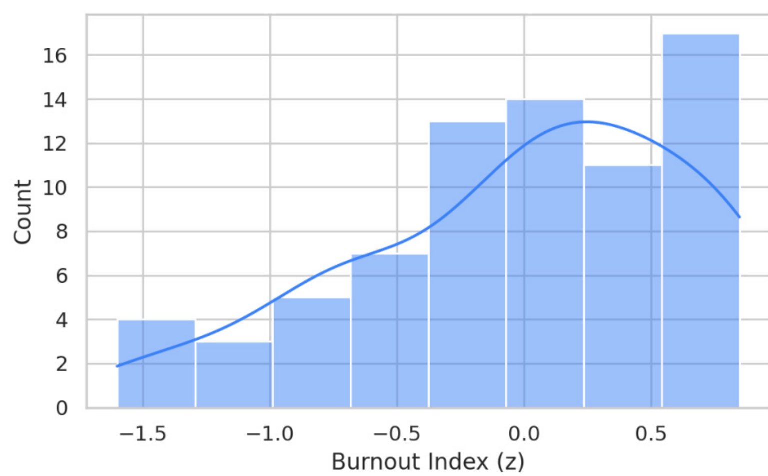

Figure 2: Distribution of the Burnout Index. Composite measure of mental fatigue and overwhelm shown as a frequency count of Burnout Index scores (N = 74 student responses). The x-axis represents the resulting Burnout Index score, where higher values indicate greater burnout risk. The y-axis represents the count of students falling into each score bin, with the highest bar (greatest value) representing 17 students. The blue line superimposed over the bars is the kernel density estimate curve, which provides a smooth estimate of the data's overall distribution.

JOURNAL OF
EMERGING INVESTIGATORS
https://doi.org/10.59720/25-269

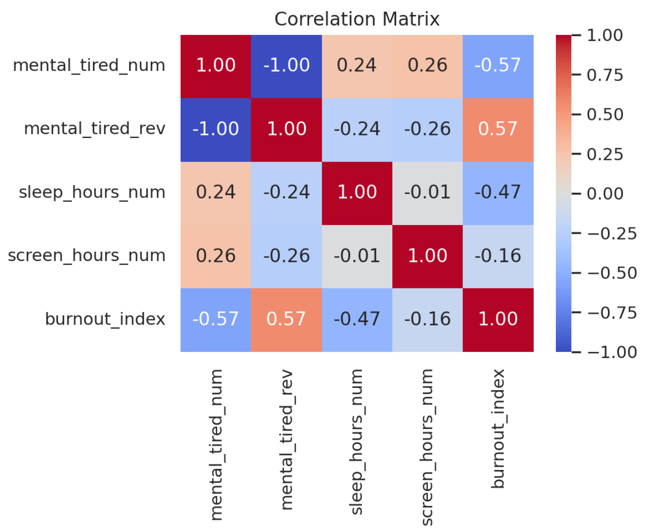

Figure 3: Correlation matrix of burnout predictors. The heatmap displays the Pearson correlation coefficients (r) between the Burnout Index and key behavioral variables. Variables are defined as follows: mental tiredness (mental_tired_num; reversed scale: mental_tired_rev), sleep duration in hours (sleep_hours_num), daily screen time in hours (screen_hours_num), and burnout index (burnout_index). Color intensity and hue indicate the strength and direction of relationships, with red representing positive correlations (max = 1.00) and blue representing negative correlations (min = -1.00). The strongest negative correlation was observed between sleep duration and the Burnout Index (r = –0.47**). A strong positive correlation was observed with fatigue (r = 0.57***). Screen exposure showed a weak, non-significant correlation (r = 0.16, ns). None of the other relationships were statistically significant. ns, not significant; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

2).

We found that approximately 80% of students were above the mean burnout level, indicating moderate to high burnout risk based on the classification criteria defined in this study (see Methods). The descriptive summary of the sample showed that students reported relatively short average sleep (<6 hours/night) and high screen exposure (>5 hours/day)—two behaviors often associated with elevated stress levels **(Figure 1)** (9). We applied the thresholds of short sleep (<6 hours/night) and high screen exposure (>5 hours/day) uniformly across age groups to maintain comparability.

Correlation Analyses

To assess the strength and direction of the linear relationship between key predictors and burnout, we performed a Pearson correlation analysis on continuous variables (sleep time, mental tiredness, screen exposure). Our analyses revealed that sleep duration had a moderate negative correlation with burnout (r = -0.47, $p < 0.01$). In other words, students reporting a longer duration of sleep also reported a lower level of burnout. This association was consistently found across all types of analyses, including descriptive statistics **(Figure 1)** and correlation matrix **(Figure 3)**. Lastly, mental fatigue also demonstrated a moderate positive correlation with burnout (r ≈ 0.57, $p < 0.001$). Conversely, exposure to screen had a weak and non-significant relationship (r = 0.16, ns).

Student Motivation and Burnout

We evaluated motivations and coping strategies to identify patterns that could help explain differences in burnout among students beyond their sleep and screen habits **(Figure 4)**. In addition to sleep and screen time, students reported varied motivations and coping strategies **(Figure 4)**. We examined these items descriptively to identify behavioral and emotional

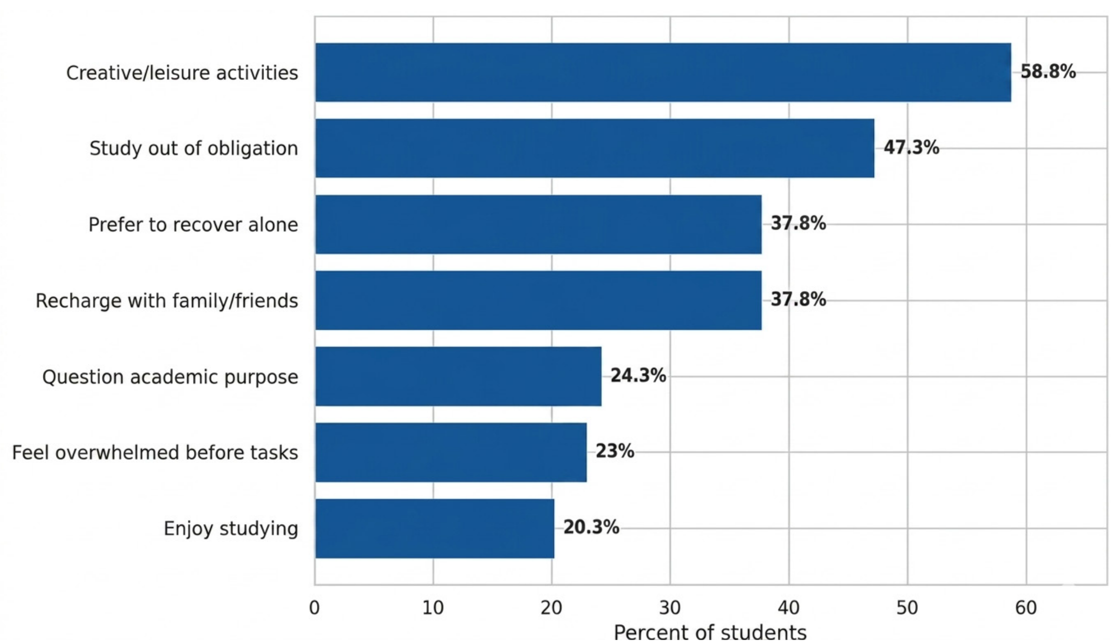

Figure 4: Student motivation and coping responses (N=74). Bar chart showing the percentage of students selecting each response option for survey items related to motivation and coping. The x-axis percentage of respondents (0–100%), and the y-axis represents the survey items. Bar height corresponds to the proportion of participants selecting each response. The baseline represents 0%.

https://doi.org/10.59720/25-269

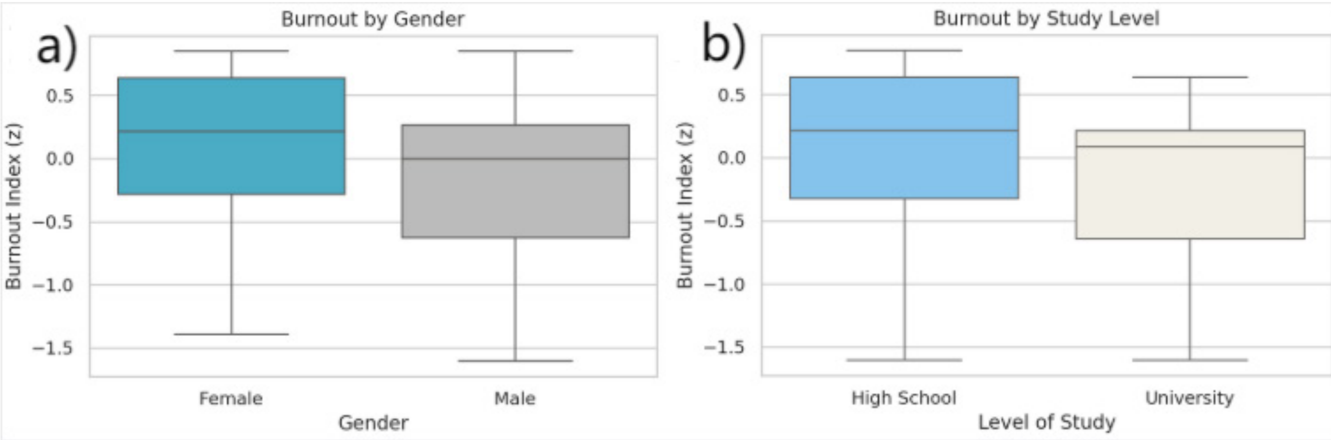

Figure 5: Burnout Index scores by gender and study level. Box plots showing the distribution of Burnout Index scores (N = 74). The y-axis represents the Burnout Index score, where higher values indicate greater burnout risk. Within each plot, the horizontal line inside the box marks the median score (the 50th percentile); the top and bottom edges of the box indicate the quartiles (the 75th and 25th percentiles); and the whiskers show the typical range of the data. a) Burnout Index scores by gender. No statistical difference between females and males was detected with an Independent Samples T-test (p = 0.078). b) Burnout Index scores by study level. There was no statistical difference between high school and university students (p = 0.070).

patterns that might explain differences in burnout beyond the statistical predictors. We presented survey questions related to motivation and coping as multiple-choice items, asking respondents to select what best applied to them (see appendix). Almost half (47.3%) of the respondents claimed to study primarily out of obligation, while only 20.3% enjoyed studying **(Figure 4)**. When investigating measures of coping, we found that a majority (58.8%) of the students used creative or leisure-based activities (music, art, or crafts) for coping purposes **(Figure 4)**. Also, some respondents indicated that spending time with family or friends helps them recharge (37.8%), while an equal number of respondents prefer to recharge alone (37.8%) **(Figure 4)**.

Group Comparisons

To verify whether burnout patterns differed between demographic groups, we used independent samples t-tests to compare means. Females displayed higher burnout than males (M = 3.21, SD = 0.72 vs. M = 2.88, SD = 0.79, respectively) **(Figure 5a)**, although this difference was not statistically significant ($t \approx 1.80$, p = 0.078). Similarly, high school students reported being more burned out than university students (M = 3.27, SD = 0.76 vs. M = 2.91, SD = 0.72, respectively; $t \approx 1.85$, p = 0.070) **(Figure 5b)**.

Extracurricular Engagement

As extracurricular activities are often associated with reduced stress, we examined whether different types of participation were associated with variation in burnout levels (10). We categorized extracurricular participation into active (e.g., sports), passive/creative (e.g., music, crafts), and none/mixed groups. We based the classification on keyword matching in free-text responses. We used a one-way ANOVA to see whether burnout levels varied across these three categories. The descriptive means indicated that those who were active participants had somewhat less burnout overall, however there were no significant differences between groups ($F \approx 0.93$, p = .403, $\eta^2 \approx 0.048$), and post-hoc Tukey tests confirmed the non-significant findings, as there were no reliable pairwise differences **(Figure 6)**.

Multivariable Regression

To determine key predictors of burnout, we performed a

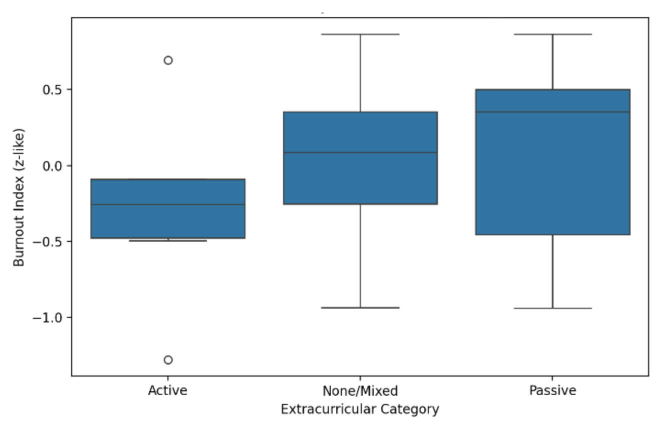

Figure 6: Burnout index scores by extracurricular category. Box plot displaying the distribution of the Burnout Index scores (line indicates median value, box limits indicate the quartiles, and whiskers indicate the range of non-outlier values within 1.5 × the interquartile range (IQR)) across three student groups: active (e.g., sports, n=6), passive/creative (e.g., music, n=10), and none/mixed (n=24). The y-axis represents the Burnout Index score, where higher values indicate greater burnout risk. Group differences were analyzed using an ANOVA test. Individual circles represent outliers (scores falling far outside the typical range). The ANOVA confirmed that there were no statistically significant differences in burnout risk across the categories (p=0.403).

https://doi.org/10.59720/25-269

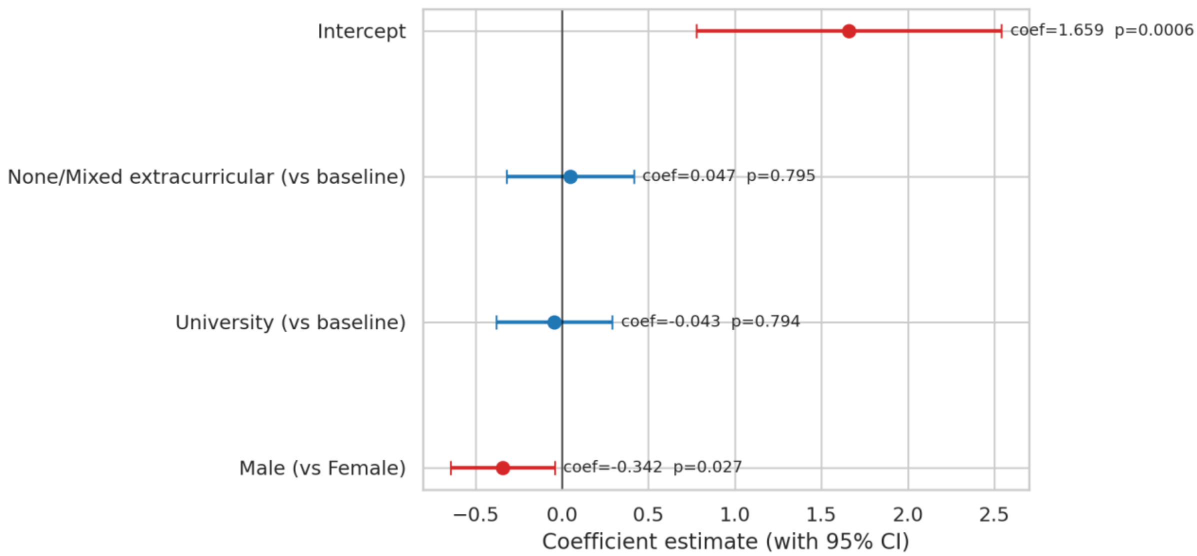

Figure 7: Regression model coefficients predicting burnout. Forest plot displaying regression coefficients for predictors of the Burnout Index. The y-axis lists predictor variables, and the x-axis represents coefficient values. Each point indicates the estimated coefficient, and horizontal lines represent 95% confidence intervals. The vertical reference line at zero indicates no effect; coefficients to the right are positive and to the left are negative.

multivariable linear regression using sleep duration, screen time, mental tiredness, gender, study level, and extracurricular participation as predictors for the burnout index. We conducted this analysis to examine how these variables worked together, since factors like sleep, mental tiredness, and demographics often overlap in their effects. Regression allowed evaluation of each predictor while controlling for the others, offering a clearer picture of which factors independently contribute to burnout. The model explained a large amount of variance in burnout ($R^2 \approx 0.616$; Adjusted $R^2 \approx 0.532$). Upon examination of the predictors, only gender was statistically significant, with males scoring lower in burnout ($\beta = -0.342$, $p = 0.027$) **(Figure 7)**. The predictors of study level (high school versus university), extracurricular categories (active, passive, or none/mixed), and screen exposure were not statistically significant after adjusting for potential confounding factors (all $p > 0.05$) **(Figure 7)**. The other noteworthy aspect of the results is that sleep and mental tiredness were significant variables in the model, corroborating the findings in the bivariate analysis and supporting our hypothesis **(Figure 7)**.

Predictive Modeling

To evaluate whether the collected variables could reliably identify students at risk of burnout, we leveraged survey responses to develop a predictive model for student burnout risk in Python using scikit-learn (11). We evaluated four complementary algorithms—Logistic Regression, support vector machine (SVM), Gradient Boosting, and Random Forest—as they are suitable for survey data containing both continuous and categorical variables. Comparing the models allowed us to identify which model was the best predictor for burnout risk. The SVM (Linear) model achieved an accuracy of 0.757, F1-score of 0.795, precision of 0.761, recall of 0.833, and CV mean of 0.676. Logistic Regression obtained an accuracy of 0.730, F1-score of 0.773, precision of 0.739, recall of 0.810, and CV mean of 0.704. Random Forest achieved an accuracy of 0.986, F1-score of 0.988, precision of 0.977, recall of 1.000, and CV mean of 0.716. Gradient Boosting also showed strong performance with accuracy of 0.986, F1-score of 0.988, precision of 0.977, recall of 1.000, and CV mean of 0.690. The highest predictive accuracy across all four models was produced by the Random Forest model with cross-validation stability of 0.716 . The Random Forest model identified low sleep duration, high screen time, and high mental fatigue as key predictors of burnout risk. This supports our hypotheses that both lower sleep duration and higher levels of self-reported mental fatigue would result in a significant risk for burnout among high school and university students. Applying the model indicated that mental fatigue emerged as the strongest predictor of burnout, contributing approximately 24.48% to the model's decision-making, followed by sleep duration (12.92%) and screen time (9.05%). screen exposure and sleep patterns.

Discussion

Upon evaluating our results, we concluded that student burnout is a complex phenomenon influenced by academic and lifestyle-related factors. We found significant correlations between less sleep and higher levels of mental fatigue with burnout, supporting our hypothesis. Prior studies often tested single predictors of burnout, while ours examined multiple measures of burnout risk, including academic pressure, peer relationship quality, engagement, and coping strategies. However, due to our small sample size, we must interpret these results with caution, as some connections between these relationships may weaken, exaggerate, or go undetected, limiting the interpretability of our findings. Therefore, future studies with larger and more diverse samples could yield a more reliable association of these factors to burnout.

One possible reason for the weaker effect of screen use with burnout is that screen exposure may be causing an indirect effect, influencing behaviors such as sleep and tiredness, rather than being an independent predictor of burnout. For example, students may stay up later on phones or computers,

https://doi.org/10.59720/25-269

reducing sleep and increasing fatigue, which in turn contributes to burnout. Regardless, this difference is important because it shows that tiredness and disengagement correlate with a lack of rest, not mere exposure to screens.

The regression model identified differences between genders; males had less burnout, perhaps indicative of gendered differences in coping, or construal of stress, which is consistent with our survey responses showing males reported using more active stress management strategies (e.g., physical activity) compared with females in our sample **(Figure 7)**. However, study level (high school vs. university) did not emerge as a significant predictor in the multivariable regression model after adjusting for other covariates, including gender **(Figure 7)**. This suggests that while younger students descriptively reported higher burnout **(Figure 5b)**, this effect may be explained by associated behaviors—such as shorter sleep duration—rather than study level itself (12). This could indicate that although younger students may feel more strain, their risk for burnout may reflect behaviors associated with young age like less sleep, rather than simply being younger (12). One explanation for these inconsistent results could include the smaller number of male participants, which may have prevented significant differentiations in burnout. Future studies could use balanced samples of gender and study level to better understand these trends.

Extracurricular involvement did not show a statistically significant protective effect, although descriptive trends suggested that active participation may serve as a buffer against stress, and that the lack of extracurricular participation tended to have higher burnout scores (**Figure 6**).

This apparent discrepancy concerning extracurricular involvement raises an important limitation with the study: the size of our sample was relatively small and unevenly distributed across groups, and this likely minimized the power to detect smaller effects. It is also possible that extracurricular involvement provides only temporary long-term protection, which is supported by self-reports in the current study showing that activities such as prayer, music, or artistic hobbies relieved stress in the short term, but did not necessarily prevent its re-appearance and extra fatigue. Nonetheless, larger future studies with more balanced samples could ascertain whether extracurricular activities, along with their types or intensities provide a lasting effect to reduce burnout. In addition to lifestyle factors, research shows that motivation and coping strategies influence burnout, supporting our study which highlights the importance of psychological factors (13). Based on these descriptive findings, both at-risk (e.g., obligation-driven motivation, pre-task fatigue) and protective (e.g., creative or social coping strategies) patterns were observed (Figure 4). These behaviors conceptually align with components of the burnout index, which includes measures such as fatigue, difficulty focusing, and feeling overwhelmed (see Methods), suggesting their potential relevance to variation in burnout levels within the sample.

Our findings extend this work by showing that burnout is not only physical but also emotional and motivational, consistent with theories emphasizing its multilayered nature **(Figure 4)**. Nevertheless, one limitation of our study is that we relied on self-reports and a cross-sectional design, which prevents us from drawing conclusions about cause-and-effect relationships. Future studies may use longitudinal designs or experimental designs to study which coping strategies are more effective in the long term, and across different student populations. This work highlights the necessity of enhancing intrinsic motivation and adaptive coping, rather than relying on short-term respite.

To extend beyond traditional analyses, we evaluated the feasibility of predicting burnout using machine learning. Among all the models, Random Forest achieved the best cross-validation accuracy (0.714) and highest overall performance on the test set. Yet, we should view these results cautiously because it can be a sign of overfitting. Interpreting our findings was complicated by the Random Forest's ensemble structure, where predictions are made through many decision trees, making the decision process less transparent and requiring careful tuning of tree parameters, which added complexity to model selection. In the future, we would expand on this work by building advanced predictive models and performing extensive generalization tests on significantly larger and more representative student samples to validate the high accuracy of the exploratory Random Forest model.

Overall, our findings indicate that sleep deprivation, psychological strain, and lifestyle habits strongly correlate with each other and are good predictors of burnout risk in students, with sleep deprivation and psychological fatigue being the strongest predictors, as we hypothesized. In addition, we found that we could accurately identify students at risk of burnout using quantitative data, which suggests the potential for using similar data tools for schools to detect risk earlier. However, applying these methods in real educational settings would be difficult, as it would require consent from students, families, and educational institutions, raising concerns regarding privacy and data security. This significantly limits the practical effectiveness of our research in improving student wellbeing.
Improving sleep habits, balancing workloads, moderating screen use, and staying involved in healthy routines may help students manage stress. Similarly, supportive conversations at home and adaptive coping behaviors such as prayer, music, journaling, or talking with peers may support students emotionally. Future research will aim to expand upon this study by examining a larger and more demographically diverse population, examining the impact of intervention over time, and building advanced predictive models to help schools in promoting student well-being. With continued development, such tools could help institutions address the underlying causes of burnout rather than responding only after students begin to struggle. Furthermore, our findings demonstrate that burnout arises from multiple interacting factors. By applying quantitative and data science methods, we can better characterize the prevalence and severity of burnout within student populations and evaluate how different behavioral and psychological factors contribute to risk.

MATERIALS AND METHODS

Study Design and Data Collection

This study employed an online, cross-sectional design using a structured questionnaire (to see questions asked, see appendix) developed in Google Forms (14). The survey was conducted for 10 days and the link was distributed to high school and university students through class group leaders and extracurricular groups such as Model United Nations debate groups, ensuring diverse participation (full list provided in appendix). The respondents were aged between 12 and 25 years. The survey was conducted in English, and participation

https://doi.org/10.59720/25-269

was completely voluntary. Written consent was mentioned at the top of the survey, and students opted to fill it out willingly.

Measures and Composite Burnout Index

The questionnaire collected basic demographic information (age, gender, and academic level) and information regarding lifestyle factors (sleep duration, screen time, and break frequency). It also asked questions related to psychological strain, such as mental tiredness, difficulty focusing, feeling overwhelmed, irritability, and participation in relaxing or extracurricular activities. Response formats varied by item: sleep deprivation, difficulty focusing, feeling overwhelmed, irritability, gender, and participation in activities were measured using binary yes/no responses; sleep hours, screen time, and break frequency were measured using multiple-choice categorical options; mental tiredness was assessed using a Likert-type scale ranging from "never" to "always"; and the final section used a checklist format where students selected statements that best described their study motivation, energy levels, and coping behaviors.

A composite burnout index was constructed by combining four indicators of psychological strain: reversed mental tiredness, difficulty focusing, feeling overwhelmed, and irritability. This approach was adopted in place of standardized instruments such as the widely used Maslach Burnout Inventory (1), as many students found it lengthy and were reluctant to complete it. Mental tiredness was originally measured on a 1–5 scale, where 1 indicated the most tired and 5 the least tired. To make the direction consistent with burnout (higher = worse), this item was reverse-scored using the transformation:

$$mental\_tired\_rev = 6 - mental\_tired\_num \quad \text{(Eqn 1)}$$

where *mental_tired_num* is the original mental tiredness score and *mental_tired_rev* is the reverse-scored value of mental tiredness. The remaining three items (difficulty focusing, overwhelm, irritability) were originally coded as Yes/No responses and were converted to binary variables (1 = Yes, 0 = No). To ensure that each component contributed equally despite being on different scales, every variable was z-scored using:

$$z_i = \frac{x_{i-x}}{SD} \quad \text{(Eqn 2)}$$

where z_i is the resulting standardized score, x_i is the observed value for a participant, $\bar{x}$ is the sample mean of that variable and *SD* is its standard deviation.

After standardization, the four z-scores were averaged to create the final Composite Burnout Index:

$$Burnout\ Index = \frac{z(mental\_tired\_rev)+z(difficulty\_focusing)+z(overwhelmed)+z(irritable)}{4} \quad \text{(Eqn 3)}$$

where *z (mental_tired_rev)* is the standardized score of reversed mental tiredness, *z(difficulty_focusing)* is the standardized binary indicator of difficulty focusing, *z(overwhelmed)* is the standardized binary indicator of feeling overwhelmed, and *z(irritable)* is the standardized binary indicator of irritability.

Higher values indicate higher burnout relative to the rest of the sample. For interpretation, participants were classified into burnout risk categories: low, medium, and high, based on their index scores: values above 0 (the sample mean) were labeled high-risk, –0.5 and 0 were labeled medium risk and scores below –0.5 were classified as low risk. The mean-based threshold was chosen because z-scores center the data around zero, making 0 a natural and easily interpretable cutoff for "higher than average" burnout in a moderate-sized sample.

Data Processing and Statistical Analysis

Survey responses were exported from Google Forms into spreadsheets and imported into Jamovi v2.6.44 (15) and Python v3.13.5 (16) for analysis in Visual Studio Code v1.103.2 (17). Categorical variables (gender, study level, extracurricular activity) were numerically encoded, and binary responses were mapped to 0/1. Continuous predictor variables, including sleep hours, screen time, break frequency, and reversed mental tiredness, were standardized by calculating z-scores. Reliability of multi-item constructs was measured using Cronbach's alpha ($\alpha \geq 0.70$ is acceptable) (18).

All analyses were first conducted in Jamovi and then repeated in Python in VS Code to ensure consistency of results. Bivariate relationships were examined using Pearson correlations between the composite burnout index (response variable) and continuous predictors: sleep hours, screen time, break frequency, and reversed mental tiredness. Heatmaps were generated in Python to visualize these associations. Group comparisons were performed using independent samples t-tests for gender (male vs. female) and study level (high school vs. university), and one-way ANOVA with Tukey post-hoc tests for extracurricular activity categories (active, passive/creative, none/mixed). These comparisons assessed whether the mean burnout index differed significantly across groups, with effect sizes (η^2) provided. Multivariable linear regression was used to examine the combined predictive power of demographics (age, gender, study level), lifestyle factors (sleep hours, screen time, break frequency), and psychological factors (reversed mental tiredness) on the composite burnout index, with model fit assessed using both R^2 and adjusted R^2. All figures including correlation matrices, distribution plots of predictor variables, and visualizations of group comparisons, were created using Python libraries (Matplotlib, Seaborn) in VS Code.

Predictive Modeling

Predictive modeling was performed in Python. The core library utilized for this project was scikit-learn for machine learning. The entire dataset was used for model training and evaluation, and then we tuned it with a 5-fold cross validation in the training dataset. We evaluated four algorithms, Logistic Regression, Random Forest, Gradient Boosting, and SVM. Logistic Regression provides an interpretable baseline; Random Forest captures non-linear interactions; Gradient Boosting sequentially corrects errors; and SVM is effective in high-dimensional data. To assess linear and nonlinear effects, models with different structural assumptions were compared under the same dataset and evaluation framework. Linear relationships between predictors and burnout were examined using Logistic Regression and a linear-kernel Support Vector Machine, both of which assume that the outcome can be separated using a linear combination of input variables. In contrast, nonlinear effects were evaluated using Random Forest and Gradient

https://doi.org/10.59720/25-269

Boosting classifiers, which capture complex interactions between variables by constructing multiple decision trees and combining their outputs. All models were trained on the same standardized feature set, and performance was compared using accuracy, precision, recall, F1-score, and cross-validation mean accuracy.

Received: September 26, 2025
Accepted: January 07, 2026
Published: July 13, 2026

REFERENCES

1. Maslach, Christina, and Susan E. Jackson. “The Measurement of Experienced Burnout.” *Journal of Occupational Behavior*, vol. 2, no. 2, 1981, pp. 99–113. https://doi.org/10.1002/job.4030020205
2. Law, Daniel W. “Exhaustion in University Students and the Effect of Coursework Involvement.” *Journal of American College Health*, vol. 55, no. 4, 2007, pp. 239–245. https://doi.org/10.3200/JACH.55.4.239-245
3. Yang, Hsiu Ju. “Factors Affecting Student Burnout and Academic Achievement in Multiple Enrollment Programs in Taiwan’s Technical–Vocational Colleges.” *International Journal of Educational Development*, vol. 24, no. 3, 2004, pp. 283–301. https://doi.org/10.1016/j.ijedudev.2003.12.001
4. Rahmati, Zahra. “The Study of Academic Burnout in Students with High and Low Level of Self-Efficacy.” *Procedia - Social and Behavioral Sciences*, vol. 171, 2015, pp. 49–55. https://doi.org/10.1016/j.sbspro.2015.01.087
5. Zhang, Yu, et al. “Perfectionism, Academic Burnout and Engagement among Chinese College Students: A Structural Equation Modeling Analysis.” *Personality and Individual Differences*, vol. 43, no. 6, 2007, pp. 1529–1540. https://doi.org/10.1016/j.paid.2007.04.010
6. Dyrbye, Liselotte N., et al. “Systematic Review of Depression, Anxiety, and Other Indicators of Psychological Distress among U.S. and Canadian Medical Students.” *Academic Medicine*, vol. 81, no. 4, 2006, pp. 354–373. https://doi.org/10.1097/00001888-200604000-00009
7. Maslach, Christina, Wilmar B. Schaufeli, and Michael P. Leiter. “Job Burnout.” *Annual Review of Psychology*, vol. 52, no. 1, 2001, pp. 397–422. https://doi.org/10.1146/annurev.psych.52.1.397
8. Paruthi, S., Brooks, L. J., D’Ambrosio, C., Hall, W. A., Kotagal, S., Lloyd, R. M., Malow, B. A., Maski, K., Nichols, C., Quan, S. F., Rosen, C. L., Troester, M. M., & Wise, M. S. “Recommended Amount of Sleep for Pediatric Populations: A Consensus Statement of the American Academy of Sleep Medicine.” *Journal of Clinical Sleep Medicine*, vol. 12, no. 6, 2016, pp. 785–786. https://doi.org/10.5664/jcsm.5866
9. Zakiei, Abolfazl, et al. “The Mechanism of Sleep Quality’s Effect on Academic Burnout: Examining the Mediating Role of Perceived Stress.” *BMC Medical Education*, vol. 25, no. 1018, 2025. https://doi.org/10.1186/s12909-025-07617-6
10. Feldman, A. F., & Matjasko, J. L. (2005). *The role of school-based extracurricular activities in adolescent development: A comprehensive review.* Review of Educational Research, 75(2), 159–210. https://doi.org/10.3102/00346543075002159
11. Pedregosa, F., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., … Duchesnay, É. “Scikit-learn: Machine Learning in Python.” *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html
12. Wang, X., et al. “Burnout and Depression in College Students.” *Psychiatry Research*, vol. 329, 2024, p. 115508. https://doi.org/10.1016/j.psychres.2024.115508.
13. Li, W., J. Chen, M. Li, A. P. Smith, and J. Fan. “The Effect of Exercise on Academic Fatigue and Sleep Quality among University Students.” *Frontiers in Psychology*, vol. 13, 2022, p. 1025280. https://doi.org/10.3389/fpsyg.2022.1025280.
14. Google LLC. “Google Forms.” *Google Workspace*. https://www.google.com/forms. Accessed 28 Aug 2025.
15. The Jamovi Project. “Jamovi (Version 2.6.44).” *Jamovi*. https://www.jamovi.org. Accessed 3 Sep 2025.
16. Python Software Foundation. “Python (Version 3.13.5).” *Python.org*. https://www.python.org. Accessed 3 Sep 2025.
17. Microsoft Corporation. “Visual Studio Code (Version 1.103.2).” *Visual Studio Code*. https://code.visualstudio.com. Accessed 3 Sep 2025.
18. Sijtsma, Klaas. “On the Use, the Misuse, and the Very Limited Usefulness of Cronbach’s Alpha.” *Psychometrika*, vol. 74, no. 1, Dec. 2008, pp. 107–120. https://doi.org/10.1007/s11336-008-9101-0

Appendix

Github Link

https://github.com/eshaalb1/Predictive-modeling-and-Statistical-Analysis

Participating Schools and Student Groups

The survey link was distributed to:

1. **Army Public School**
2. Quaid-e-Azam University
3. Roots International
4. DMUN Online Debate Group

Survey Questionnaire

1. Full Name *
2. Age *
3. Gender *
• Male
• Female
4. Grade/Level of Study *
• High school
• University
5. Are you sleep deprived? *
• Yes
• No
6. How many hours do you sleep on average? *
• Less than 5
• 5–6
• 7–8

• More than 8
7. How many hours do you spend on screen (phone/laptop) per day? *
• Less than 2
• 2–4
• 5–6
• More than 6
8. How often do you take breaks during study or work? *
• Every 30 min
• Every hour
• Rarely
• Not at all
9. How often do you feel mentally tired or drained? *
• 1 - Always
• 2
• 3
• 4
• 5 - Never
10. Do you experience difficulty focusing, even when trying? *
• Yes
• No
11. Do you feel overwhelmed by your daily tasks? *

• Yes
• No
12. Do you get irritated or frustrated more easily lately? *
• Yes
• No
13. Do you engage in any relaxing or extracurricular activities? (e.g., sports, music) *
14. Which of the following statements best describe you? (Select all that apply) *
• I usually enjoy studying and feel motivated to learn.
• I study mostly because I have to, not because I want to.
• I often feel tired even before starting my academic tasks.
• Spending time with family/friends helps me recharge.
• I prefer being alone to recover my energy.
• I enjoy social gatherings and feel energized after them.
• Creative or leisure activities (e.g., art, music, sports) help me reduce stress.
• I often question the value or purpose of my academic work.