Journal of Mental Health

https://doi.org/10.63808/jmh.v2i1.432 Volume 2 , Issue 1

Research Article

Construction of an Early Warning Model for Academic Burnout among High School Students Based on Learning Behaviour Data

Tiange Xiang[1,*]

[1]School of Social Work and Health Management, Xihua University, Chengdu 610039, China

[*]**Correspondence:** Tiange Xiang, xtg0222@163.com

Abstract

The issue of academic burnout among adolescents has gained increased prominence in the field of educational psychology; nevertheless, current measurement instruments and predictive models remain limited due to their use of cross-sectional self-reports and black box modeling pipelines which do not allow for any actionable solutions in the school environment. This paper suggests an explainable machine-learning approach to identifying cases of academic burnout among high school students based on their observable learning behaviors. More specifically, by employing a publicly accessible dataset that includes information about 649 adolescents and contains no validated burnout measurement, and considering the multi-dimensional construct of burnout, the modeling pipeline involves a composite proxy label which implies converging data on at least two aspects of burnout, predictors which consist of observable behaviors, three supervised learning algorithms (logistic regression, random forest, and extreme gradient boosting), and the SHapley Additive exPlanations method which is implemented at both population and individual levels. Crucially, all variables used to construct the proxy label, together with any feature derived from them, were excluded from the predictor matrix to prevent target leakage. Once leakage was removed, the models performed only modestly (held-out ROC-AUC ≈ 0.74 and PR-AUC ≈ 0.52–0.56 for the tree-based models), far below the near-ceiling values obtained when label-defining variables are retained, indicating that such high performance is largely circular. The strongest leakage-free predictors were early-term grades, alcohol use, free time, and age rather than attendance and study time. Thus, observable behaviors carry only a weak-to-moderate, non-causal signal for a proxy burnout label, and a deployable screening tool would require an independent dataset containing a measured burnout instrument such as the MBI-SS.

Keywords

Journal of Mental Health
WISDOM ACADEMIC https://doi.org/10.63808/jmh.v2i1.432 Volume 2 , Issue 1

Academic burnout, Adolescents, Educational psychology, Explainable machine learning, SHapley Additive exPlanations

1. Introduction

1.1. Background and Significance

It seems that the issue of academic burnout is relevant to contemporary adolescence psychology in the context of their learning experience. Over the past few years, the phenomenon of academic burnout has become quite prevalent in China, with more than 70 percent of adolescents admitting to experiencing this problem, which makes this issue relevant to the population's mental health issues rather than personal coping strategies (Gao, 2023). Taking into consideration the developmental nature of such experience, the importance of the constant interaction of exhaustion, cynicism concerning the activities in schools, and a decrease in one's sense of competence is essential because it determines how adolescents develop their self-concept in the context of learning process during their development period. From the point of view of academic performance, the concept of burnout is better understood as a multidimensional response to prolonged stress associated with studies, where emotions, motivation, and competence are connected through reciprocal links but not accumulative. Such a point of view is justified based on the findings related to the influence of traumatic experiences on adolescents' psychological well-being. First of all, it is necessary to note that burnout mediates the effect of trauma on people's psychological well-being in case their core beliefs were disturbed during their time in secondary school (Zeng et al., 2022). Furthermore, the study on school-related issues has shown that burnout is highly associated with the attachment of adolescents to their schools, creating a vicious circle of factors impossible to overcome with regular pedagogical interventions (Zhao et al., 2024). In other words, it is possible to conclude that burnout experienced by high school students is a syndrome characterized by certain behaviors. Such an understanding of the issue is crucial from the point of view of its mental health-related consequences since burnout increases existential anxiety and post-traumatic stress symptoms (Tomaszek & Muchacka-Cymerman, 2022).

1.2. Research Gap and Innovation

Taking into account the significance of the problem of academic burnout stated in the previous section, it becomes necessary to analyze the methodological tools utilized in investigating the issue. Prior studies in the domain have mostly focused on conducting cross-sectional surveys; although the data obtained through such research is informative in terms of determining the prevalence of the phenomenon and related variables, the method used does not allow predicting the occurrence of academic burnout among college students. In systematic reviews of the literature on predictive learning analytics, it is possible to see numerous innovations in predicting performance and student drop-out rate; yet, when it comes to psychological aspects of students' education, there seems to be very little research focused on predicting the onset of psychological phenomena, with most of it being centered on predicting performance outcomes (Sghir et al., 2023). This aspect of predictive learning analytics poses a challenge for burnout prevention efforts since performance outcomes usually appear after the decline in students' psychological wellbeing, meaning that it is difficult to predict burnout from declining grades. Moreover, another limitation of the existing state of affairs in the field of predictive learning analytics is related to the interpretability of the findings obtained. The increasing use of machine learning approaches makes it necessary to pay attention to the fact that there may be accurate predictive models lacking any explanation of their findings, which may lead to problems when it comes to decision-making based on predictive analytics results (Khosravi et al., 2022). Additionally, one should mention the sensitivity of self-report measures to situational factors, which was proven by recent psychometric research (Salmela-Aro et al., 2022). The combination of the above-mentioned limitations forms a coherent domain of innovation in predictive learning analytics. Specifically, the domain in question is characterized by the absence of predictive research on psychological burnout, low interpretability of predictive algorithms, and the dependence on self-report measures. The current paper seeks to fill in the gap through the utilization of learning behavior data along with an explainable machine learning algorithm.

1.3. Research Objectives

Building upon the methodology introduced in the previous section, this study aims to achieve the following three related goals. First, to develop a leakage-aware warning system for academic burnout using learning behaviors shown by high school students instead of just one questionnaire measurement, excluding every variable used to define the outcome from the predictor set and thereby reducing dependency on self-reports which may vary depending on the state. Second, to test how well a linear model performs compared to two tree-based models in realistic class imbalance situations, and to quantify how much of the apparent performance is attributable to target leakage by progressively removing label-defining and grade-related variables, in order to identify which modeling framework provides the best trade-off between sensitivity and specificity for the purpose of early-warning. Finally, to conduct post hoc explainability analysis at both the population and individual levels in order to evaluate the psychological plausibility, and the limits, of the behavioral processes used to predict outcomes.

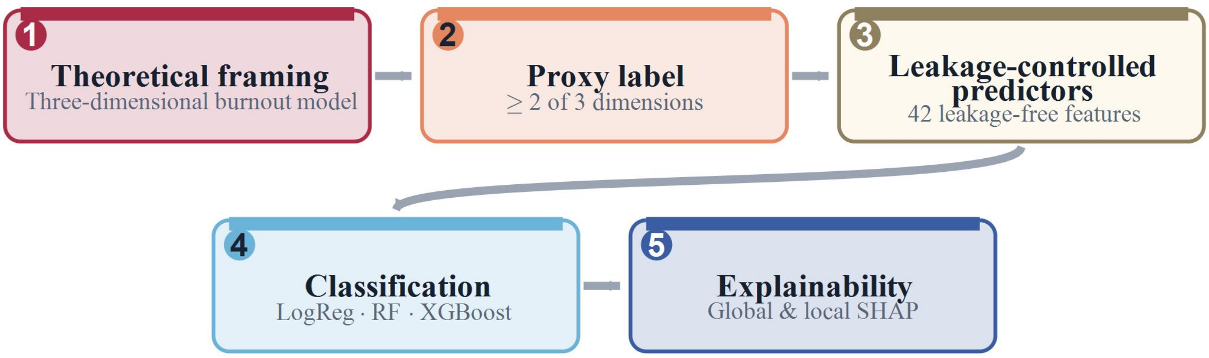
2. Methods

2.1. Research Design and Data Source

The three-dimensional construct of academic burnout, which consists of emotional exhaustion, cynical attitude toward learning, and perceived incompetence, has been widely recognized in educational psychology as one of the most comprehensive definitions (Schaufeli et al., 2002). It was developed through cross-cultural validation of the construct in different student samples. Thus, the three-dimensional model will serve as a conceptual framework for the study. Instead of developing an original survey, the researcher has chosen to use the publicly available UCI Student Performance Dataset (available at https://archive.ics.uci.edu/dataset/320/student+performance). It includes behavioral, demographic, and academic data of adolescents enrolled in two Portuguese public secondary schools (Cortez & Silva, 2008). The sample consists of 649 respondents between 15 and 22 years old (M = 16.74, SD = 1.22). Among the sample, 383 adolescents are females (59.0%), while 266 are males (41.0%). As for the place of residence, 452 respondents come from cities (69.6%) and 197 come from rural areas (30.4%). Academic performance is assessed during three grading periods, G1, G2, and G3, on a scale of 0-20 according to the Portuguese grading system. The dataset includes a relatively representative sample of adolescents, who are frequently considered as an underrepresented group in academic burnout research. Being freely available, the dataset allows ensuring the transparency and replicability of the research methodology. The absence of burnout values will require operationalization of academic burnout as a proxy variable.

Figure 1

Leakage-aware analytical pipeline



This can be seen in Figure 1, which shows an intentionally linear process at the macro-level but still retains its recursive nature at a smaller level. All of the decisions made as a result of modelling in the latter stages are based on the theory developed in the first stage, when machine learning models in educational psychology remain interpretable rather than predictive.

2.2. Feature Engineering and Variable Construction

Measurement validity related to translating a latent psychological construct into operational indicators has been an issue in the literature on measures of burnout, particularly in terms of the discriminant validity of the three constructs as distinct subscales among adolescents

(Aguayo-Estremera et al., 2023). The current research addresses this issue by requiring that at least two constructs converge before considering the students under investigation as those at high burnout risk, thereby overcoming measurement errors associated with measuring each dimension independently. High absenteeism represented emotional exhaustion since the students had absences amounting to or above the 75th percentile, which is typical of emotionally exhausted people who avoid their responsibilities. Cynicism was measured based on low study hours per week (≤ 1 on the four-point scale) or absence of aspirations to further education. Low efficacy was represented by poor academic performance, such as failing courses (≥ 1) or scoring less than ten marks overall. Because the dataset contains no validated burnout instrument, this composite is a behavioral proxy rather than a measured burnout score, a limitation addressed in the Discussion.

In designing the predictor matrix, attention was focused on learning-behavioral predictors, in line with the key assumption of the research that burnout indicators can be found in learning behavior rather than self-reported affective states. Apart from learning-behavioral predictors available in the original data set (internet availability, school support, and extracurricular activities), six predictors were engineered, one of which, the study-to-leisure index (which embeds weekly study time), was subsequently excluded as leakage, leaving five retained: the early-term grade change (G1-G2) and its magnitude, which together encode the direction and size of the first-to-second period trajectory and are typically observed among students with cynical burnout; the averaged parental education predictor, created from two correlated background predictors; the combined alcohol predictor, which considers both weekday and weekend alcohol use as a behavioral indicator of distress; and a combined social support predictor. Crucially, all predictors used to construct the label were excluded from the predictor matrix to prevent target leakage: total absences (exhaustion), study hours per week and higher-education aspiration (cynicism), and past failures and the final grade G3 (efficacy), together with any feature derived from them such as the study-to-leisure index. After one-hot encoding of categorical variables, the full candidate pool comprised 46 columns (20 numeric and 26 indicator columns from 17 categorical variables); removing the four leakage-related variables left 42 predictors in the primary matrix, and a stricter configuration that additionally removed all grade-related variables retained 38. The predictors were classified into four psychological categories, as shown in Table 1.

Table 1

Predictor domains

Conceptual Domain	Representative Variables	Psychological Interpretation
Demographic	age, sex, address, school	Developmental and contextual background

Familial	Medu, Fedu, famrel, famsup, parental_education_avg†	Parental capital and proximal support resources
Behavioural	activities, freetime, goout, alcohol_combined†, social_support†	Observable engagement and self-regulation patterns
Early-academic	G1, G2, grade_decline†, grade_volatility†	Performance trajectory and aspiration toward learning

Note. †Denotes engineered features. Variables used to construct the proxy label (absences, studytime, higher, G3, past failures) and any feature derived from them (study_leisure_balance) were excluded from the predictor matrix to prevent target leakage.

The structure of the organization mentioned in Table 1 holds great significance because it denotes the conceptual clusters that will help in treating feature-level attributions as information relating to family, behavior, or academic achievement, rather than just as numbers.

2.3. Model Construction and SHAP Interpretation

Three machine learning classifiers that were implemented and compared included logistic regression (linear classifier), random forest (bagged tree learner), and XGBoost (gradient boosted tree learner). Stratified random sampling yielded a training dataset, which consisted of 80% of data (n = 519) and a held-out test set (20%, n = 130). Continuous variables were standardized via Z-score normalization for logistic regression, while tree-based models did not require feature scaling due to invariance to monotonic transformations. Class imbalance problem in the training labels (24.0% high-burnout-risk) entailed application of SMOTE only to training split, whereas the test split was left intact to avoid overoptimistic performance estimation without any gain in generalization. Hyperparameters were optimized via five-fold cross-validation and grid search, area under the ROC curve (AUC) being chosen as the optimization metric. Test results were reported by metrics of accuracy, precision, recall, F1, ROC-AUC, and PR-AUC, and, following reviewer guidance, were extended to include sensitivity, specificity, balanced accuracy, the Matthews correlation coefficient, calibration (Brier score), and confusion matrices for all models. Robustness was assessed with repeated stratified five-fold cross-validation (ten repeats) and 95% confidence intervals from 1,000-sample bootstrapping of the held-out predictions, and, because the default 0.50 threshold is suboptimal under imbalance, an F1-optimal threshold was additionally derived on the training folds (never on the held-out test set); the searched hyperparameter grids and selected values are reported in the main text: for the primary leakage-controlled models, logistic regression used C = 0.1, random forest used 200 trees with a maximum depth of 6, and XGBoost used 200 trees with a maximum depth of 3 and a learning rate of 0.05. All analyses were carried out in Python 3.12 with scikit-learn 1.8.0, XGBoost 3.3.0, SHAP 0.52.0, imbalanced-learn 0.14.2, NumPy 2.4.4, and pandas 3.0.2, under a fixed random seed to support reproducibility.

Accuracy is not enough to make sense of the psychological value of predictions, as two models with comparable AUC can exhibit very different patterns of interaction between features, thus implying totally different intervention strategies. Thus, the SHAP framework was chosen as the interpretability technique, since it offers a unified attribution method that operates in one additive model and satisfies several axiomatic criteria for attribution (local accuracy, missingness, and consistency) making it possible to give interpretation both at the level of each student individually and aggregation of all students together (Lundberg & Lee, 2017). In the SHAP framework, each prediction is explained via the additive feature attribution model:

$$g(z') = \varphi_0 + \Sigma\ (i = 1 \text{ to } M)\ \varphi_i\, z'_i$$

Where $g(z')$ is the model used to predict the target variable, φ_0 represents the average prediction without the inclusion of feature information, φ_i represents the SHAP value of the *i*th feature, $zi' \in \{0,1\}$ indicates the presence/absence of the *i*th feature, and M is the total number of input features. The exact values of the SHAP values for tree-based models were determined through TreeSHAP. Global ranking was performed through the mean of absolute SHAP values of all observations in the test set, while SHAP waterfall plots illustrated the contribution of behavioral features in individual prediction.

3. Results

3.1. Sample Description and Burnout Distribution

The sample size of the preprocessed data amounted to 649 teenagers. As for the three proxies for dimensions defined in Section 2.2, the operationalized positive rates amounted to 26.3%, 36.5%, and 23.3% in the case of emotional exhaustion, cynicism, and reduced efficacy respectively, meaning that there is no dominating dimension in terms of constructing the compound label and overlapping of the three criteria is partially observed. Applying the two-of-three convergence rule, 156 students (24.0%) were allocated to the high-risk group in terms of burnout, while 493 students (76.0%) were assigned to the low-risk one. The achieved positive rate of 24% is the positive rate of a constructed proxy label rather than a measured prevalence of clinically defined burnout; although broadly within the range reported for adolescent academic burnout, this correspondence is indicative only, and the imbalance is realistic enough for modeling with stratified sampling/SMOTE usage. Notably, cynicism showed the highest positive rate, while reduced efficacy turned out to be the most conservative criterion, implying a higher deterioration rate of motivation compared to competence.

3.2. Model Performance Comparison

The three classifiers were assessed using an extended set of metrics, and all these metrics together provide a measure for the discriminative capability, accuracy per class, and ranking when there is an imbalance in the dataset. As the dataset already had a class imbalance, recall and PR-AUC got greater

Journal of Mental Health
WISDOM ACADEMIC https://doi.org/10.63808/jmh.v2i1.432 Volume 2 , Issue 1

importance than accuracy because of the cost of failure of predicting a student at high risk of burnout being much higher than that of the occurrence of false positives. After excluding every label-defining variable and its derivatives, the performance of the classifiers on the test set is presented in Table 2, with extended metrics, confidence intervals, and confusion matrices in Table 3.

Table 2

Held-out classification performance

Classifier	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Logistic Regression	0.723	0.447	0.677	0.538	0.742	0.459
Random Forest	0.762	0.500	0.484	0.492	0.741	0.558
XGBoost	0.785	0.552	0.516	0.533	0.738	0.517

Table 3

Extended evaluation metrics

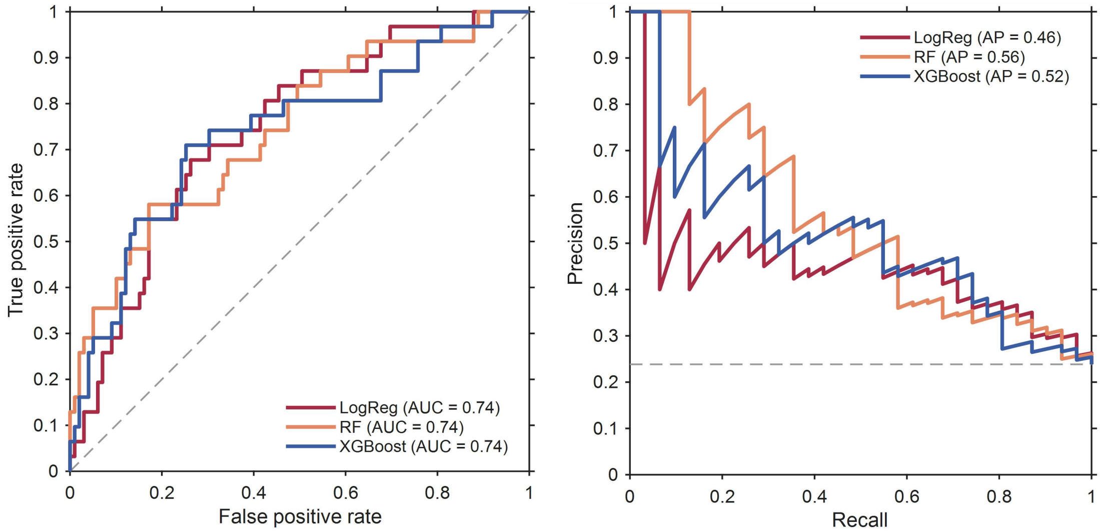
Classifier	Sensitivity	Specificity	Balanced Acc.	MCC	Brier	ROC-AUC (95% CI)
Logistic Regression	0.677	0.737	0.707	0.368	0.205	0.742 (0.64–0.83)
Random Forest	0.484	0.848	0.666	0.336	0.167	0.741 (0.64–0.84)
XGBoost	0.516	0.869	0.692	0.394	0.169	0.738 (0.62–0.84)

Note. Confusion matrices (true negatives, false positives, false negatives, true positives): logistic regression (73, 26, 10, 21); random forest (84, 15, 16, 15); XGBoost (86, 13, 15, 16). Repeated stratified five-fold cross-validation (ten repeats) ROC-AUC: logistic regression 0.769 ± 0.048, random forest 0.796 ± 0.042, XGBoost 0.759 ± 0.046. PR-AUC 95% bootstrap confidence intervals: logistic regression 0.459 (0.31–0.64), random forest 0.558 (0.40–0.73), XGBoost 0.517 (0.34–0.69).

From Table 2, it is observed that, once label-defining variables were removed, the three classifiers performed only modestly and comparably. Logistic regression achieved accuracy (0.723) and recall (0.677) with low precision (0.447), which indicates that linear decision boundaries might overestimate positive samples. Random forest classifier achieved precision (0.500) and the highest PR-AUC (0.558); however, its recall was lower (0.484). XGBoost achieved the most balanced results, which include accuracy (0.785), precision (0.552), recall (0.516), F1 (0.533), ROC-AUC (0.738), and PR-AUC (0.517), together with the highest specificity (0.869) and Matthews correlation (0.394). These values are far below those obtained when label-defining variables are retained (ROC-AUC ≈ 0.94–0.96; PR-AUC ≈ 0.87–0.88), confirming that the previously high performance was largely an artefact of target leakage. Hence, XGBoost classifier was selected for interpretability analysis. Because the default 0.50 threshold is suboptimal under imbalance, an F1-optimal threshold (≈ 0.28 for XGBoost, selected on the training folds) raises recall to roughly 0.71 at the cost of precision (≈ 0.45). Curves depicting ROC and PR-AUC scores of all three classifiers are shown in Figure 2 below.

Figure 2

ROC and precision–recall curves



It is evident from Figure 2 that the three classifiers show similar and only moderate area under the curve performance in the ROC plot; however, it is clear from the precision-recall plot that precision falls well below the near-ceiling values produced by leakage-prone models once the recall values are stringent. This is particularly relevant in early warning scenarios where recall is vital but high false positives are undesirable.

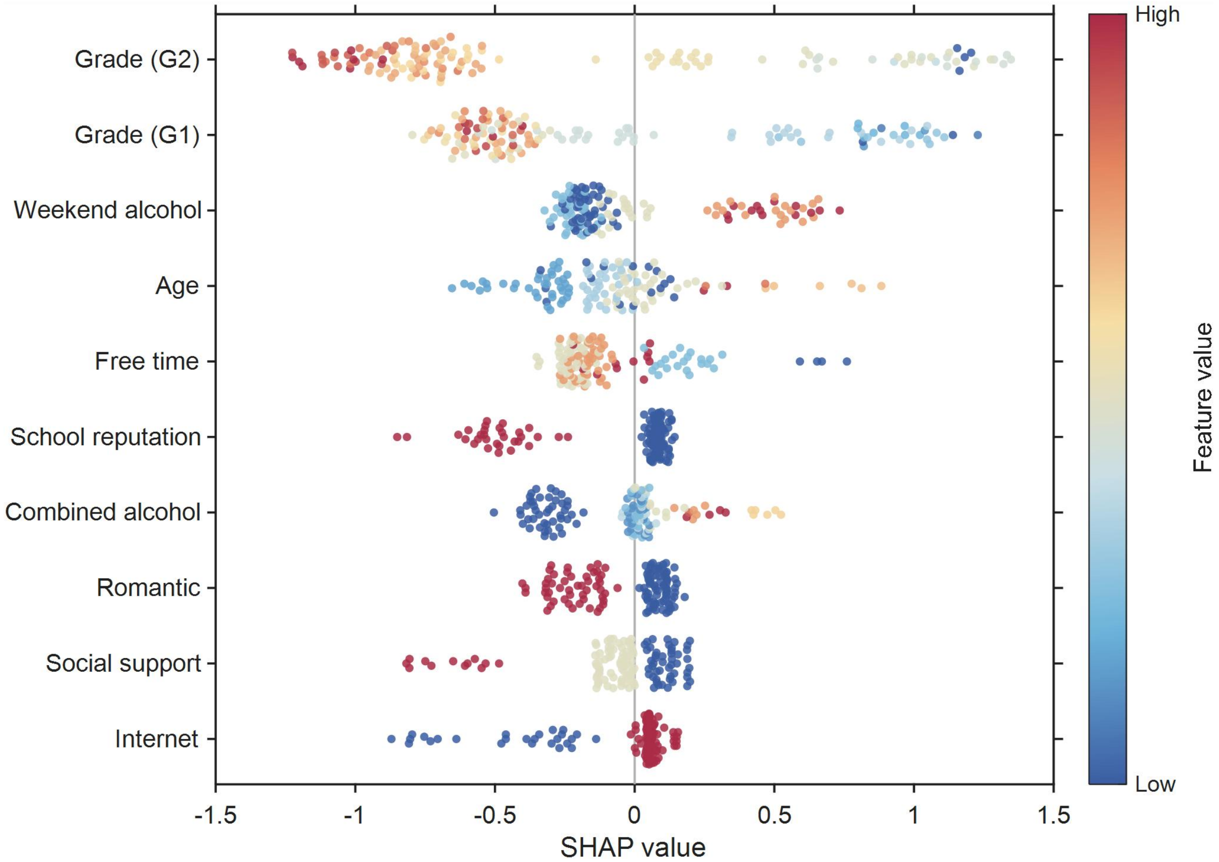
3.3. Global SHAP Interpretation

To ensure that the outcome of the predictive accuracy is psychologically significant, SHAP values were computed for the test set based on the leakage-controlled model built in section 3.2 using XGBoost. The ranking of the mean absolute SHAP values was done for the predictor vector, whereby the second-period grade G2 turned out to be the most significant predictor (mean |SHAP| = 0.786), followed by the first-period grade G1 (0.553), weekend alcohol use (0.247), age of the student (0.206), free time (0.199), choosing the school for its reputation (0.182), combined alcohol use (0.166), romance (0.140), social support (0.138), and internet access (0.137). Absenteeism and study time, which dominated the leakage-prone model, are absent here by design because they were used to construct the label. Although the ranking of the features gives an indication of the significance of the features, it does not reveal anything about the nature of their impact on prediction, which is essential for psychological research. This is shown graphically in figure 3.

Figure 3

Journal of Mental Health
https://doi.org/10.63808/jmh.v2i1.432 Volume 2 , Issue 1

SHAP feature attributions



As seen from Figure 3, lower early-term grades and higher alcohol use are associated with predictions that point toward the category of high burnout risk. Conversely, higher values of G1 and G2, together with stronger social support, are associated with predictions toward the category of low risk. The stated patterns are psychologically plausible, but their magnitudes are small, and SHAP attributions describe predictive contributions within this proxy-label model rather than causal effects on burnout.

3.4. Local Explanation Cases

However, global rankings assess the model's performance in a population context, without addressing individual-level problems which have been leading to the development of early warning approaches. To address this issue, three representative examples were chosen from the test set and analyzed in terms of additive feature contributions via SHAP. The chosen examples consist of: one example of a clearly high-risk individual (probability 0.965, true label 1); one example of a clearly low-risk individual (probability 0.006, true label 0); and one example of a borderline individual (probability 0.513, true label 1). High-risk individual: this prediction is mostly driven by very low early grades (G1 = 5, G2 = 0) and older age. Low-risk individual: this example is mostly defined by good grades in the first stages of the term and available social support, which provides negative SHAP values. Borderline individual: the contributions of both positive and negative features do not define the final outcome, with low early grades pushing toward risk and internet access pulling away from it. The above-described analysis shows that the model's predictions can be considered as psychological portraits of each individual, though given the modest aggregate performance they are best read as hypotheses for follow-up.

4. Discussion

In this particular study, an interpretable, leakage-aware machine learning pipeline has been designed and stress-tested for screening burnout risk among high school students in terms of their learning behavior, addressing the methodological gaps mentioned above, namely the absence of a prediction model for burnout, the non-interpretability of the models used, and the use of self-reports. Its principal contribution is methodological: an initial leakage-prone formulation produced near-ceiling performance, and removing the label-defining variables reduced performance to a modest level, demonstrating how target leakage can manufacture an apparent behavioral signal. The results of this study fit into the general framework of computational educational psychology.

The leakage-controlled performance of the three classifiers is appreciably lower than the strong results often suggested by machine-learning studies of student burnout, which typically draw on richer psychological and lifestyle predictors than the behavioral fields available here (Pereira et al., 2025). Specifically, the XGBoost algorithm displayed only modest accuracy and F1 scores of 0.785 and 0.533, correspondingly, whereas the random forest classifier reached an ROC-AUC of only 0.741. Rather than a shortcoming, this represents a more honest estimate of how much observable behavior, without access to the labelling rule, can predict a proxy burnout label, with the three algorithms performing comparably and the tree-based models offering only marginally better precision–recall behavior than the linear model.

Secondly, in addition to the factors related to the model itself, the very structure of the leakage-free feature attributions can also be seen as overlapping with the literature. Based on the findings of SHAP analysis, early-term grades, alcohol use, free time, and age were ranked the highest, while absenteeism and weekly study time were deliberately excluded because they define the proxy label. This ranking of the features is broadly consistent with the results of studies in the domain of explainable machine learning, where engagement-related features proved important for predicting students' dropout (Krueger et al., 2023). Their apparent importance in leakage-prone models is therefore a labelling artefact rather than a discovery.

Because weekly study time and higher-education aspiration were used to construct the cynicism dimension, they could not serve as predictors, and the protective associations sometimes attributed to learning engagement (Wang et al., 2023) could not be tested directly in this design. What remains is a weaker, indirect signal carried by alcohol use, leisure patterns, and family resources, consistent with the view that disengagement is multiply determined but not strongly recoverable from observable behavior alone in the absence of a validated burnout measure.

The practical implication concerns the usage of machine learning techniques in school psychological counseling programs. The presented study tempers rather than supports immediate deployment: a leakage-controlled model built only from routinely collected attendance and academic behavioral characteristics did not reach the discrimination needed for confident individual screening, and Chen and Liu (2024) similarly caution that ensemble accuracy on academic data does not transfer automatically to psychological targets. SHAP local explanations could nonetheless support transparent, human-in-the-loop triage if the pipeline were first validated against a measured burnout instrument.

It is crucial to point out the existence of certain restrictions that should be considered when evaluating the results of the conducted research. First, the application of the open source data set created in a non-Chinese cultural environment makes generalization of the findings difficult as far as they concern the target group of secondary school students; future research will need to prove the viability of the pipeline on Chinese adolescents using a culturally appropriate dataset that includes a validated burnout instrument. Second, the outcome is a behaviorally-constructed proxy: the dataset contains no MBI-SS or equivalent, so the label could not be validated internally, and the study should be read as proof-of-concept rather than a validated screening system. Third, even after leakage control, the retained predictors remain statistically associated with the variables that define the label, so a residual, non-causal dependency cannot be fully excluded; the configuration that additionally removed all grade-related variables (ROC-AUC 0.67–0.71, with recall falling sharply (full metrics in Appendix)) marks the lower bound of the genuinely behavioral signal. Fourth, SHAP attributions are predictive, not causal. Finally, no independent external dataset was available; the two UCI subsamples share most students, so cross-course validation would itself be contaminated, and robustness rests on repeated cross-validation and bootstrapping rather than true external validation.

5. Conclusion

The purpose of this research paper is to design an interpretable machine-learning algorithm for screening academic burnout risk in high school students using observable learning behaviors. This study seeks to solve the problems associated with questionnaire methods and black-box machine learning models used in previous studies. In this regard, burnout was operationalized in an archival adolescent dataset as a multi-dimensional construct, with the label assigned when at least two of the three dimensions were satisfied. All label-defining variables and their derivatives were excluded from the feature space to prevent target leakage. Three supervised classifiers spanning linear and tree-based families were compared, and both global and local SHAP analyses were used for interpretability. Once leakage was controlled, the best-balanced but still modest performance was given by the XGBoost classifier, which had an accuracy and F1 score of 0.785 and 0.533, respectively, on the test set. Using SHAP values, early grades, alcohol use, free time, and age were identified as the leading features, whereas absenteeism and study time were excluded by design because they define the label.

Firstly, the value of the findings lies less in predictive strength than in what they reveal about method, and in their significance regarding the general problem of applying machine learning in the context of adolescent mental health. By making the labelling rule explicit, removing it from the feature space, and quantifying the resulting drop in performance, the authors present an empirically grounded proof-of-concept rather than a validated screening tool, whose deployment at school would require validation against a measured burnout instrument on an independent, culturally appropriate sample. The main value of this study is that it recognizes machine learning analysis of adolescent burnout as a form of psychological research, rather than an abstract classification framework.

Acknowledgments

The author declares that no specific acknowledgments are required.

Author contributions

Not applicable.

Funding

This research received no external funding.

Data availability

No data availability statement is required.

Declarations

Conflict of interest

The author declares that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Additional information

Received: 25 May 2026

Accepted: 1 June 2026

Published online: 22 September 2026

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Aguayo-Estremera, R., Cañadas, G. R., Ortega-Campos, E., Ariza, T., & De la Fuente-Solana, E. I. (2023). Validity evidence for the internal structure of the Maslach Burnout Inventory-Student Survey: A comparison between classical CFA model and the ESEM and the bifactor models. *Mathematics*, *11*(6), Article 1515. https://doi.org/10.3390/math11061515

[2] Chen, M., & Liu, Z. (2024). Predicting performance of students by optimizing tree components of random forest using genetic algorithm. *Heliyon*, *10*(12), Article e32570. https://doi.org/10.1016/j.heliyon.2024.e32570

[3] Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.), *Proceedings of the 5th Future Business Technology Conference (FUBUTEC 2008)* (pp. 5–12). EUROSIS-ETI.

[4] Gao, X. (2023). Academic stress and academic burnout in adolescents: A moderated mediating model. *Frontiers in Psychology*, *14*, Article 1133706. https://doi.org/10.3389/fpsyg.2023.1133706

[5] Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

[6] Krüger, J. G. C., de Souza Britto, A., Jr., & Barddal, J. P. (2023). An explainable machine learning approach for student dropout prediction. *Expert Systems with Applications*, *233*, Article 120933. https://doi.org/10.1016/j.eswa.2023.120933

[7] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates, Inc.

[8] Pereira, M. G., Santos, M., Magalhães, R., Rodrigues, C., Araújo, O., & Durães, D. (2025). Burnout risk profiles in psychology students: An exploratory study with machine learning. *Behavioral Sciences*, *15*(4), Article 505. https://doi.org/10.3390/bs15040505

[9] Salmela-Aro, K., Upadyaya, K., Ronkainen, I., & Hietajärvi, L. (2022). Study burnout and engagement during COVID-19 among university students: The role of demands, resources, and

psychological needs. *Journal of Happiness Studies*, *23*(6), 2685–2702. https://doi.org/10.1007/s10902-022-00518-1

[10] Schaufeli, W. B., Martínez, I. M., Marques Pinto, A., Salanova, M., & Bakker, A. B. (2002). Burnout and engagement in university students: A cross-national study. *Journal of Cross-Cultural Psychology*, *33*(5), 464–481. https://doi.org/10.1177/0022022102033005003

[11] Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in predictive learning analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, *28*(7), 8299–8333. https://doi.org/10.1007/s10639-022-11536-0

[12] Tomaszek, K., & Muchacka-Cymerman, A. (2022). Student burnout and PTSD symptoms: The role of existential anxiety and academic fears on students during the COVID-19 pandemic. *Depression Research and Treatment*, *2022*, Article 6979310. https://doi.org/10.1155/2022/6979310

[13] Wang, X., Qiao, Y., & Wang, S. (2023). Parental phubbing, problematic smartphone use, and adolescents' learning burnout: A cross-lagged panel analysis. *Journal of Affective Disorders*, *320*, 442–449. https://doi.org/10.1016/j.jad.2022.09.163

[14] Zeng, Z., Wang, X., Chen, Q., Gou, Y., & Yuan, X. (2022). Core belief challenge moderated the relationship between posttraumatic growth and adolescent academic burnout in Wenchuan area during the COVID-19 pandemic. *Frontiers in Psychology*, *13*, Article 1005176. https://doi.org/10.3389/fpsyg.2022.1005176

[15] Zhao, H., Han, M., Wang, Z., & Liu, B. (2024). School connectedness and academic burnout in middle school students: A multiple serial mediation model. *Behavioral Sciences*, *14*(11), Article 1077. https://doi.org/10.3390/bs14111077

Appendix

Grade-free configuration (all grade-related variables removed; 38 predictors): held-out performance

Classifier	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Logistic Regression	0.677	0.378	0.548	0.447	0.674	0.425
Random Forest	0.738	0.412	0.226	0.292	0.706	0.427
XGBoost	0.769	0.533	0.258	0.348	0.707	0.452

Note. Balanced accuracy / MCC: logistic regression 0.633 / 0.238; random forest 0.562 / 0.158; XGBoost 0.594 / 0.250. This configuration additionally removes G1, G2, and their derivatives, leaving 38 behavioural, demographic, familial, and contextual predictors.