

Support Vector Machine Classification of High-Risk University Students Using Academic Stress, Sleep Quality, Self-Efficacy, and Depression

Monica. Moreno¹, José Carlos. Barbosa^{2*}

¹ Faculty of Behavioral Sciences, Universidad de La Sabana, Cundinamarca 111321, Colombia

² Department of Psychosocial Studies, Faculty of Psychology, Universidad del Valle, Campus Universitario Melendez, Cali 760042, Colombia

* Corresponding author email address: josebarbosa@correounivalle.edu.co

E d i t o r	R e v i e w e r s
Eduin Alexander Prada-Rodriguez  Department of Psychology, Universidad de Investigación y Desarrollo (UDI), Bucaramanga, Santander, Colombia eprada1@udi.edu.co	Reviewer 1: Oriana Piskorz-Ryń  Faculty of Health Sciences, University of Caldas, Street 6623b-03, Manizales 170004, Caldas, Colombia. Email: oriana_piskorz-ryne@gmail.com Reviewer 2: Diana Rodrigues da Silva  Insight - Piaget Research Center for Human and Ecological Development, Escola Higher de Education Jean Piaget, Almada, Portugal. Email: DianadaSilva@ese.ipv.pt

1. Round 1

1.1. Reviewer 1

Reviewer:

In the introduction paragraph discussing academic stress, the sentence “Academic stress is one of the most frequently studied determinants of student vulnerability” is appropriate, but the construct is treated mainly as a general psychological burden. Please clarify whether academic stress is conceptualized as a predictor of psychological risk, an academic-environmental stressor, or part of the high-risk classification criterion. This distinction is important because academic stress appears both as an independent predictor and as one of the criteria used to define the outcome variable, which may create conceptual overlap.

In the paragraph on sleep quality, the manuscript states that “poor sleep should be interpreted as an early warning indicator rather than as a routine feature of student life.” This is a valuable argument; however, the introduction should more explicitly explain why sleep quality is suitable for machine-learning classification. Please add a short rationale indicating that sleep

quality may contribute nonlinear and interaction effects with depression and academic stress, which justifies its inclusion in an SVM model rather than only in conventional correlational analysis.

In the paragraph defining the dependent variable, the manuscript states that “Risk classification was established by combining clinically validated cut-off scores for depressive symptoms with indicators of severe academic stress and poor sleep quality.” This is a critical point and requires substantial elaboration. Please specify the exact cut-off scores used for depression, academic stress, and sleep quality, explain whether students needed to meet all criteria or any combination of criteria, and justify this rule using prior literature or clinical reasoning.

The manuscript should address potential circularity in the modeling framework. Since academic stress, sleep quality, and depression are used both to define high-risk status and to predict high-risk classification, the model may partly learn the rule used to create the outcome rather than discover an independent predictive pattern. Please revise the Methods and Discussion to acknowledge this limitation and consider defining the outcome using an external criterion, such as clinically significant depression, counseling referral status, academic probation, or composite risk validated independently from the predictors.

Authors revised and uploaded the document.

1.2. Reviewer 2

Reviewer:

In the introduction paragraph on self-efficacy, the manuscript states that “self-efficacy is a protective psychological factor that may buffer the negative effects of academic stress, sleep disturbance, and depressive symptoms.” This claim is theoretically sound, but the manuscript should explain whether self-efficacy is expected to have a direct protective effect, a moderating effect, or both. Since the reported SVM model uses self-efficacy as a direct predictor, the theoretical framing should align with the analytical approach or the authors should explicitly state that moderation was not formally tested.

In the final introduction paragraph, the manuscript argues that “Support Vector Machine modeling offers an analytically rigorous approach for identifying students whose combined psychological profile indicates elevated risk.” This statement should be expanded with a clearer methodological justification for choosing SVM over other classification algorithms. Please specify whether SVM was selected because of its suitability for small-to-medium feature spaces, nonlinear decision boundaries, robustness in binary classification, or prior performance in psychological classification tasks.

In the Methods and Materials section, the paragraph beginning “This study employed a quantitative, cross-sectional predictive research design” provides an adequate overall design, but the manuscript should clarify whether the study is purely predictive or also explanatory. Because SHAP analysis is later used to interpret feature importance, the authors should acknowledge that model explainability provides post hoc interpretive insight but does not establish causality or confirm theoretical mechanisms.

In the Study Design and Participants paragraph, the manuscript states that the study was conducted “between September 2025 and February 2026 across five major public and private universities located in Bogotá, Medellín, Cali, Barranquilla, and Bucaramanga, Colombia.” Please provide more detail on how universities were selected and whether institutional differences were considered. Since public and private universities may differ in socioeconomic composition, academic demands, support services, and student mental health resources, ignoring institutional clustering may affect the interpretation of the results.

In the participant description, the manuscript reports that “A total of 1,248 undergraduate students participated in the study.” The exact sample size is commendable, but the manuscript should add a sample size justification. Please state whether the sample size was determined through power analysis, machine-learning learning-curve considerations, expected prevalence of high-risk cases, or practical recruitment constraints. For predictive modeling studies, sample adequacy should be justified not only statistically but also in relation to the number of predictors and the outcome class distribution.

In the same section, the manuscript states that participants were recruited using “a stratified random sampling strategy.” This needs methodological clarification. Please specify the strata used for sampling, such as university, academic discipline, year of study, gender, or city. Also indicate how students were randomly selected within each stratum. Without this information,

the claim of stratified random sampling may appear insufficiently supported and may be interpreted as convenience sampling with stratification after recruitment.

Authors revised and uploaded the document.

2. Revised

Editor's decision after revisions: Accepted.

Editor in Chief's decision: Accepted.