

Privacy-Preserving Prediction of Psychological Distress Across Schools Using Academic Workload, Bullying, Sleep Quality, and Social Support: A Federated XGBoost Study

Luis. Montenegro¹, Eirik. Solheim^{2*}

¹ Department of Psychology, University of Panama, Panama City, Panama

² Department of Clinical Psychology, University of Bergen, Bergen, Norway

* Corresponding author email address: e.solheim@uib.no

E d i t o r	R e v i e w e r s
Seyed Ali Darbani  Assistant Professor, Department of Psychology and Counseling, South Tehran Branch, Islamic Azad University, Tehran, Iran Ali.darbani@iau.ac.ir	Reviewer 1: Parvaneh Mohammadkhani  Professor, Department of Clinical Psychology, University of Rehabilitation Sciences and Social Health, Tehran, Iran. Email: Pa.mohammadkhani@uswr.ac.ir Reviewer 2: Mehdi Rostami  Department of Psychology and Counseling, KMAN Research Institute, Richmond Hill, Ontario, Canada. Email: dr.mrostami@kmanresce.ca

1. Round 1

1.1. Reviewer 1

Reviewer:

The statement that “a total of 2,684 students received invitations, of whom 2,491 completed the survey” requires a more complete participant-flow description. The authors should report the number and reasons for nonparticipation before survey completion, the exact reasons for each of the 63 exclusions, and the distribution of exclusions across schools. A participant-flow diagram would improve transparency, particularly because data-quality exclusions such as “implausible response patterns” and “implausibly short completion times” may introduce subjective decisions. The thresholds and algorithms used to identify these cases should be prespecified and reported in sufficient detail to permit replication.

The ethical statement indicates that “ethical approval was obtained from the appropriate Norwegian research ethics and data-protection authorities,” but the specific approving body, approval number, and date are absent. These details are required in the final manuscript. The authors should also describe the consent procedure more precisely, especially for participants aged 16 and 17 years, because Norwegian rules regarding adolescent consent may depend on the nature of the research and the sensitivity of the data. The manuscript should clarify whether parental consent, parental notification, student assent, or

independent adolescent consent was used and should describe the protocol followed when students reported severe distress or possible risk of self-harm.

The operationalization of the primary outcome requires stronger methodological justification. The sentence “Students scoring 30 or higher were classified as experiencing high psychological distress” should be supported by evidence that this cutoff is valid for Norwegian adolescents or for a closely comparable population. The Kessler Psychological Distress Scale may be scored and categorized differently across studies, and cutoffs developed in adults or other countries may not provide equivalent sensitivity and specificity in Norwegian upper secondary students. The authors should report the source of the threshold, explain why a binary outcome was preferred over ordinal or continuous modeling, and include sensitivity analyses using alternative cutoffs and the continuous score.

The manuscript repeatedly characterizes the model as “privacy-preserving,” but the privacy guarantees are not formally demonstrated. Secure aggregation and encrypted transmission reduce exposure, yet they do not automatically prevent model inversion, gradient leakage, membership inference, or reconstruction attacks. The authors should clearly distinguish data localization, transport encryption, secure aggregation, and formal privacy protection. If differential privacy was not applied, the term “privacy-preserving” should be qualified. The paper should include a threat model defining the assumed adversary, trusted and untrusted parties, potential attack vectors, and residual privacy risks.

The creation of the centralized XGBoost benchmark requires additional ethical and methodological explanation. The sentence “a fully de-identified research copy of the data” appears to conflict with the study’s emphasis on avoiding centralized pooling. The authors should explain where the centralized dataset was stored, who had access, what de-identification procedure was applied, and why centralized pooling was ethically permissible for benchmarking but not intended for operational use. If centralized analysis was only simulated or conducted under restricted conditions, this should be stated clearly. Otherwise, the privacy rationale of the study may appear internally inconsistent.

The data-splitting strategy may lead to optimistic estimates because students from the same schools were represented in the training, validation, and test subsets. Although the leave-one-school-out analysis partially addresses institutional generalization, the primary test performance reflects within-school rather than entirely external validation. The authors should emphasize this distinction. It would be preferable to designate one or more schools as fully external test sites from the beginning or to use nested leave-one-school-out validation for hyperparameter tuning and evaluation. The current analysis should be described as internal-external cross-validation rather than definitive external validation.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

The description of academic workload is currently insufficient for psychometric evaluation. The manuscript refers to “an adapted Academic Workload Scale,” but the original source, adaptation process, scoring method, dimensional structure, and evidence of validity are not reported. The authors should explain which items were retained, removed, or modified, whether translation and back-translation were conducted, and whether cognitive interviews or expert review were performed. Because academic workload emerged as the strongest predictor, the measurement quality of this construct is central to the validity of the conclusions. Factorial validity and measurement invariance across schools, gender groups, and academic versus vocational programs should also be examined.

The bullying assessment paragraph states that selected items from the revised Olweus Bully/Victim Questionnaire were used. The use of selected items rather than the complete validated instrument requires justification. The authors should identify the exact items, response categories, reference period, scoring procedure, and criterion used to define recurrent victimization. Traditional bullying and cyberbullying should be conceptually separated because their frequency, context, and psychological consequences may differ. It should also be clarified whether the cyberbullying items came from the Olweus measure or another instrument, and whether the measure assessed perpetration, victimization, or both.

The sleep-quality measurement section includes the Pittsburgh Sleep Quality Index, bedtime, wake time, sleep duration, and social jetlag, but the derivation of these variables is not fully described. The authors should specify how sleep efficiency and social jetlag were calculated, how implausible sleep times were handled, and whether sleep duration was computed separately for school days and weekends. Because several sleep features are mathematically and conceptually related to the global Pittsburgh Sleep Quality Index, the authors should discuss the possibility of redundant predictors and explain whether highly correlated or overlapping features were retained intentionally. The potential for feature duplication to inflate apparent sleep importance should be evaluated.

The manuscript reports that internal consistency was “examined separately within each participating school and in the complete analytical sample,” yet only overall Cronbach’s alpha coefficients are presented. The authors should provide school-specific reliability ranges and preferably McDonald’s omega in addition to alpha, particularly for multidimensional instruments such as the Pittsburgh Sleep Quality Index. Internal consistency alone does not establish construct validity. The authors should also report confirmatory factor analyses or cite robust prior validation evidence for Norwegian-language versions of all measures used in the study.

The missing-data procedure requires revision. The sentence “missing continuous values were imputed locally using the median value calculated from the training portion of the relevant school dataset” indicates a simple single-imputation strategy that may underestimate uncertainty and distort associations when missingness is systematic. The authors should report the proportion of missing data for every variable and school, evaluate missingness patterns, and compare participants with and without missing responses. Multiple imputation or model-based procedures should be considered, or the authors should provide evidence that missingness was minimal enough for median and modal imputation to be unlikely to affect results. A sensitivity analysis comparing imputed and complete-case models would strengthen confidence in the findings.

The federated learning architecture is not described with sufficient technical specificity. The paragraph stating that “schools calculated local gradient and Hessian statistics” should identify the exact federated XGBoost implementation, software package, version, communication protocol, and aggregation algorithm. The authors should specify whether horizontal federated learning was used, whether trees were constructed synchronously, how split candidates were communicated, how many communication rounds occurred, and how institutional sample-size differences were handled. Without these details, the analysis cannot be independently reproduced or meaningfully compared with other federated XGBoost implementations.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.