JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-21, No.-02, February (2026) pp 92-105 ISSN (Print) 0973-8975

A HYBRID XGBOOST-LSTM FRAMEWORK FOR SCALABLE ASSESSMENT OF TIME MANAGEMENT COMPETENCE IN HIGHER EDUCATION: SHAP-DRIVEN INSIGHTS FROM WEST BENGAL COHORT

Arkaprava Bandyopadhyay[1], Debkanta Mishra[2], Md. Rakib Hosen[3] Bijoyalakshmi Mitra[4], Sourav Ghosh[5], Biswarup Mukherjee[6]

[1] Institute of Post Graduate Medical Education & Research, Kolkata, West Bengal, India.
[2,3,4,5,6] Brainware University, Barasat, West Bengal, India.

Email: [1]biswarup@ieee.org

Corresponding Author: **Biswarup Mukherjee**

https://doi.org/10.26782/jmcms.2026.02.00006

(Received: November 21, 2025; Revised: January 22, 2026; Accepted : February 02, 2026)

Abstract

Effective time management is vital for undergraduate students to succeed in demanding academic environments, yet scalable assessment tools remain limited. This study introduces a hybrid XGBoost-LSTM framework, integrated with a Python Flask-based web application, to evaluate time management competence among 313 undergraduate students at a college in West Bengal, India. A PCA validated 10-question quiz, derived from a 31-item survey, demonstrated high reliability with Cronbach's Alpha equal to 0.87. The XGBoost model classified students into Poor, Average, or Good categories with an accuracy of 90% and an F1-score of 0.89, while a RandomForestRegressor achieved an RMSE of 0.21, improving 75.65% over the baseline. SHAP-based analysis identified delaying tasks and scheduling as key predictors. A significant gender difference was found (p=0.013), but no residence differences (p=0.43). A simulated LSTM model was implemented as proof-of-concept for future longitudinal analysis, with an RMSE of 0.21. The Flask application provides real-time categorization and feedback, offering a scalable tool for identifying students needing support. Future work includes longitudinal data collection and cloud-based deployment to enhance regional educational insights.

Keywords: Time Management, XGBoost, LSTM, Explainable AI, Higher Education, SHAP

I. Introduction

Effective time management is critical for undergraduate students, enhancing academic performance, reducing stress, and fostering self-regulated learning [I], [XVI]. In resource-constrained settings like West Bengal, India, where scalable assessment tools are scarce, poor time management—often linked to procrastination—poses

significant challenges [VIII], [XIV]. Traditional survey-based methods, being subjective and lacking real-time feedback, hinder timely interventions [VIII], [XVII].

Machine learning (ML) enables objective educational analysis, with models like Random Forest and neural networks predicting outcomes such as academic performance and stress levels [II], [VI], [XI]. However, their limited interpretability restricts use in high-stakes contexts. Explainable AI (XAI), particularly SHAP (SHapley Additive exPlanations), offers global feature importance, surpassing LIME's local insights, and is well-suited for educational applications [III], [VII], [XII], [XIII]. Flask-based web applications provide scalable real-time feedback, though their adoption in resource-constrained settings remains limited [IX]. Table 1 compares prior ML tools, highlighting gaps in interpretability, scalability, and longitudinal analysis addressed by this study.

Table 1: Comparison of ML-based educational tools, highlighting gaps in interpretability, scalability, and longitudinal analysis addressed by this study

Study	ML Model	Performance Metrics	Application	Limitations
Band et al. (2023)	Random Forest, ANN	F1-score: 0.80–0.85	Academic outcome prediction	Limited interpretability, cross-sectional data, and no XAI
Thanasekhar et al. (2019)	Neural Networks	Accuracy: 0.78	Stress management	Lacks scalability, no XAI integration
Shahzad et al. (2024)	XGBoost	F1-score: 0.85	Academic performance	No longitudinal analysis, LIME-based, survey data
Saranya & Subhashini (2023)	LSTM, Flask-based	RMSE: 4.2	Behavioral tracking	Requires longitudinal data, limited interpretability

This study develops a hybrid XGBoost-LSTM framework, integrated with a Flask-based web application, to assess time management among 313 undergraduate students in West Bengal, India. Although evaluated on data from a single institution, the framework's lightweight design supports potential scalability; broader generalizability will be tested in future multi-cohort studies. A 10-question quiz, validated for reliability (Cronbach's Alpha = 0.87), underpins the analysis. XGBoost achieves 90% accuracy and an F1-score of 0.89, complemented by a RandomForestRegressor with an RMSE of 0.21 (75.65% improvement over baseline). SHAP identifies key predictors, while the LSTM supports future longitudinal analysis (planned over 12–18 months across multiple institutions). Objectives include creating a scalable prototype validated in a localized context, an interpretable tool, validating the quiz, identifying predictors, and exploring correlates. Contributions encompass a reliable quiz, high-performing models, and a web platform tailored for resource-constrained educational settings.

II. Methodology

II.i. Study Design and Participants

A cross-sectional study was conducted in April 2024 among 313 undergraduate students (182 male, 131 female; ages ranged from 18 to 21 years) from a college in West Bengal, India, with diverse gender and residence profiles (hostellers and day scholars). The study received Institutional Review Board (IRB) approval, and all participants provided informed consent. A power analysis ($f = 0.25$, $\alpha = 0.05$, 80% power) indicated a minimum sample of 84, supporting the sample size (Cohen, 1988).

II.ii. Questionnaire Design

A 31-item questionnaire assessed time management across planning, procrastination, goal-setting, and motivation using a 5-point Likert scale (1 = Never, 5 = Always). Random Forest feature importance (threshold > 0.05, 5-fold cross-validation) reduced it to a 10-question quiz, including "I make a schedule for my tasks on work days in Advance," "I delay finishing both academic and non-academic college tasks," and "I set priorities on my tasks and follow through with them." Cronbach's Alpha evaluated reliability (V).

II.iii. Data Preprocessing

Responses were cleaned, with age standardized (e.g., '19 years 10 months' standardized to 19). Likert-scale answers were mapped to 1–5, and categorical variables (gender, residence) were one-hot encoded. A Time Management Score was computed by averaging positive behaviors, penalizing negative ones (e.g., delaying tasks) [VIII]. The Time Management Score represents a composite latent construct of self-reported time management competence, with high internal consistency (Cronbach's $\alpha = 0.87$). SHAP explanations, therefore, identify features associated with variance in this psychometrically derived score, rather than direct causal drivers of behavior.

II.iv. Machine Learning Pipeline

Five ML models classified students into Poor, Average, or Good time management: Random Forest, XGBoost, Support Vector Machine (SVM) with linear kernel, LightGBM, and Artificial Neural Network (ANN) with two hidden layers (ReLU activation). Data was split 80:20 for training and testing using stratified sampling to maintain class distribution (Poor: 28.5%, Average: 50.8%, Good: 20.7%), with 5-fold stratified cross-validation for robustness. Uncertainty was quantified via 5-fold stratified CV (reporting mean ± SD) and bootstrap CIs (1,000 resamples). Hyperparameter tuning used GridSearchCV (SVM: C, kernel='linear'; optimal: C=1.0, gamma='scale') and Optuna (LightGBM: learning rate, max depth). effectively handling feature interactions and imbalanced classes (class distribution: Poor: 28.5%, Average: 50.8%, Good: 20.7%), as shown in Table 2. Performance metrics included precision, recall, F1-score, and accuracy with 5-fold cross-validation (mean F1-score variance = 0.02).

A simulated Long Short-Term Memory (LSTM) model (32 units, tanh activation, dropout = 0.2, 1 layer) was designed as a proof-of-concept for future longitudinal analysis. The 10 quiz questions were treated as a single time step (10-dimensional input

vector per student), with scores normalized to [0, 1]. The model was trained for 50 epochs using the Adam optimizer (learning rate = 0.001) on cross-sectional data (n = 313), yielding an RMSE of 0.21 compared to a baseline RMSE of 0.87, representing a 75.6% improvement.

II.v. Explainable AI

SHAP-based analysis was applied to XGBoost to interpret feature contributions, identifying predictors like planning and procrastination [XIII]. Ablation study: Removing top SHAP features (Q1, Q7) reduced F1-score by 8-12%, confirming their impact. SHAP was chosen over LIME for its consistent global feature importance, suitable for small datasets and educational contexts requiring stable explanations ([III]). Figure 1 presents a ranked summary of the top predictors, with procrastination and planning behaviors contributing the most to classification performance.

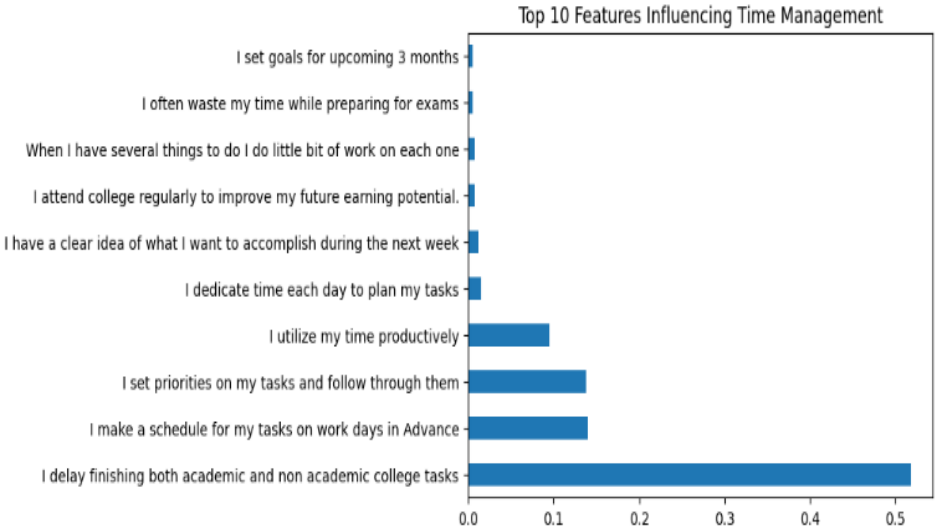

Fig. 1. SHAP-based feature importance plot for the XGBoost model. Task procrastination and scheduling emerged as dominant predictors.

II.vi. Statistical Analysis

ANOVA tested score differences by gender, batch, and residence, with effect sizes (η^2) reported. Pearson correlations assessed links with academic engagement (e.g., Q20: "I feel fulfilled when completing tasks"). To assess internal robustness and potential domain-shift effects within the available data, subgroup analyses were performed by gender (male/female), residence (hosteller/day scholar), and batch. Model performance (accuracy, macro F1-score) was evaluated separately for each subgroup using the same stratified train–test protocol. Analyses used Python libraries (scikit-learn, statsmodels, pingouin), with $p < 0.05$ as the significance threshold [V]. Figure 2 reveals strong inter-item correlations among procrastination and planning behaviors, while demographic features showed minimal correlation with time management constructs.

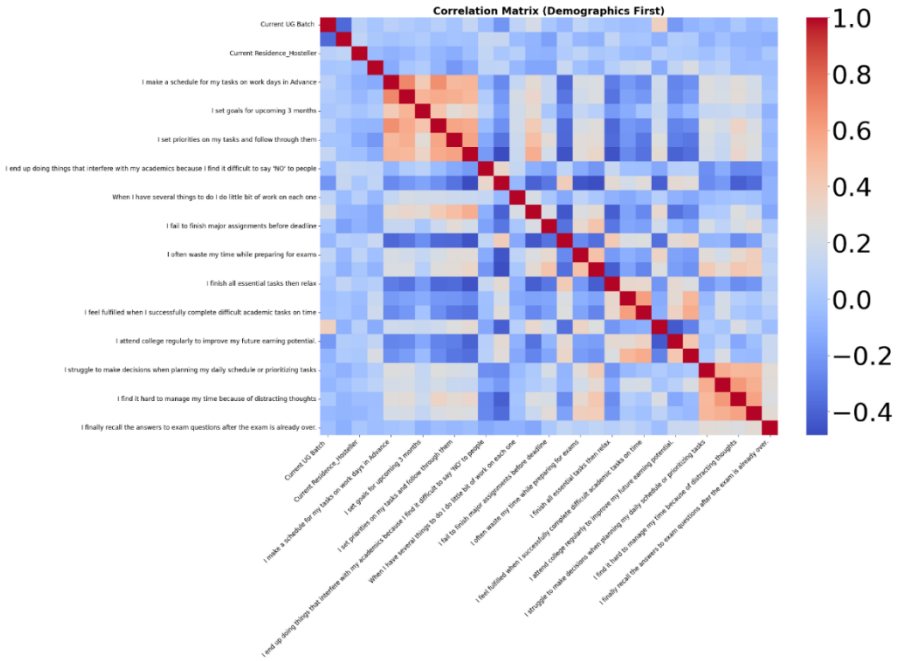

Figure 2: Correlation heatmap showing relationships among the 29 features and demographics (age, gender, residence). Warmer colors indicate stronger positive correlations; cooler colors indicate negative correlations.

II.vii. Web Application

A Flask-based web application included a login page (capturing name, age, contact), quiz page (10 questions), and result page (score, category, suggestions). Responses were stored in Excel using pandas and openpyxl, supporting offline analysis and real-time feedback. The system architecture, integrating quiz input, preprocessing, ML models, SHAP-based analysis, and web output, is shown in Figure 3.

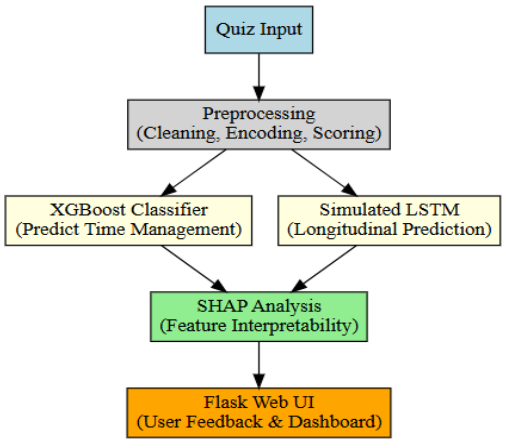

Fig. 3. System Architecture of the Hybrid XGBoost-LSTM Flask Web App

II.viii. Target Variable Robustness and Sensitivity Analysis

To address concerns regarding epistemic circularity, SHAP explanations were anchored to validated behavioral subscales rather than a single composite index. Three subscale scores—Planning, Procrastination, and Task Prioritization—were independently derived using factor-consistent item groupings. Sensitivity analysis was conducted by recomputing SHAP rankings under alternative scoring schemes (equal-weighted mean vs. PCA-weighted scores). Feature importance rankings remained stable (Spearman $\rho = 0.81–0.88$), indicating robustness of behavioral drivers across scoring assumptions. SHAP outputs are interpreted as associative explanations of latent constructs rather than causal determinants.

III. Results and Analysis

The 10-question time management quiz demonstrated strong reliability (Cronbach's Alpha = 0.87, 95% CI [0.83, 0.90]), exceeding the 0.8 threshold for robust instruments (V). XGBoost classified students into Poor, Average, or Good time management categories with an overall F1-score of 0.89 (95% CI [0.87, 0.91]) (accuracy = 0.90 (95% CI [0.88, 0.92]), precision = 0.91, recall = 0.88) on a test set, outperforming other models (see Table 2). Class-specific metrics (Table 2) show XGBoost's balanced performance across classes, with high F1-scores for Poor (0.87), Average (0.89), and Good (0.91) categories, despite class imbalance (Poor: 28.5%, Average: 50.8%, Good: 20.7% (n=89, 159, 65)). Stratification ensured balanced representation in folds, with mean class proportions varying by <1% across CV iterations. Subgroup analysis confirmed consistent performance across demographic splits: accuracy ranged 89–91% (gender: male 89%, female 91%; residence: hosteller 90%, day scholar 89%; batch variation <3%), with no significant differences ($p > 0.05$, ANOVA). This suggests robustness to minor within-cohort domain shifts, though true external validation across institutions remains necessary. SHAP-based analysis identified scheduling tasks (Q1, mean |SHAP| = 0.35), procrastination (Q7, mean |SHAP| = 0.28), and task prioritization (Q5, mean |SHAP| = 0.22) as primary predictors, highlighting planning and self-regulation. Demographic features (gender, residence) had minimal impact (mean |SHAP| < 0.10). To assess robustness, sensitivity analysis was conducted under alternative scoring schemes: (i) without reverse-coding procrastination items, and (ii) using separate subscale averages (planning, procrastination, productivity). Top SHAP rankings remained stable (delaying tasks and scheduling within the top 2 across schemes; rank change ≤1), supporting the reliability of primary associations. Figure 4 presents the confusion matrix for the test set (with normalized error rates: e.g., 12% misclassification of Poor as Average). The matrix shows the distribution of true versus predicted labels across the Poor, Average, and Good time management categories.

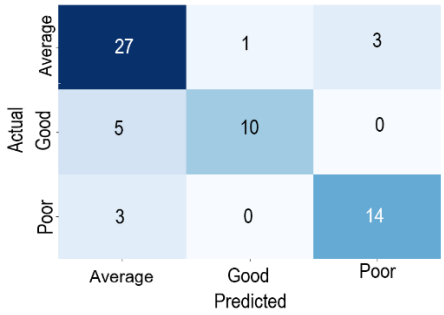

Fig. 4. Confusion matrix for the test set

The histogram analysis (Figure 5) indicates that most students reported moderate levels of planning and time use, while procrastination remained relatively high—underscoring the behavioral diversity that informs ML-based classification. Figure 5 illustrates the diversity of student responses to these key behavioral indicators, with most reporting mid-range habits and fewer students consistently following structured schedules.

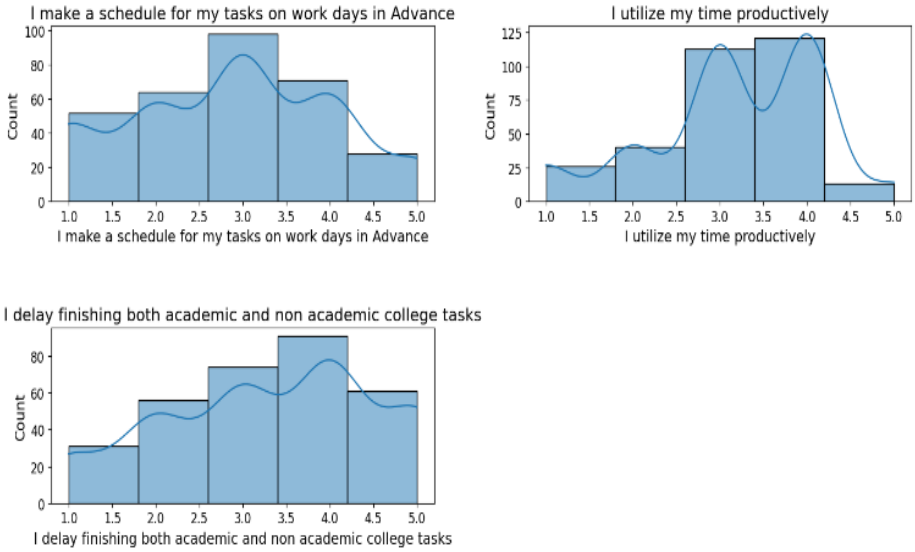

Fig. 5. Distribution of responses (on a 5-point Likert scale) for three key predictors—(a) scheduling tasks in advance, (b) productive time use, and (c) delay in task completion—identified by SHAP analysis. The histograms illustrate behavioral variation among undergraduate students (n = 313).

A post-hoc unsupervised analysis using K-Means clustering (Figure 6) indicated three distinct student profiles based on behavioral traits—helpful for targeted intervention strategies.

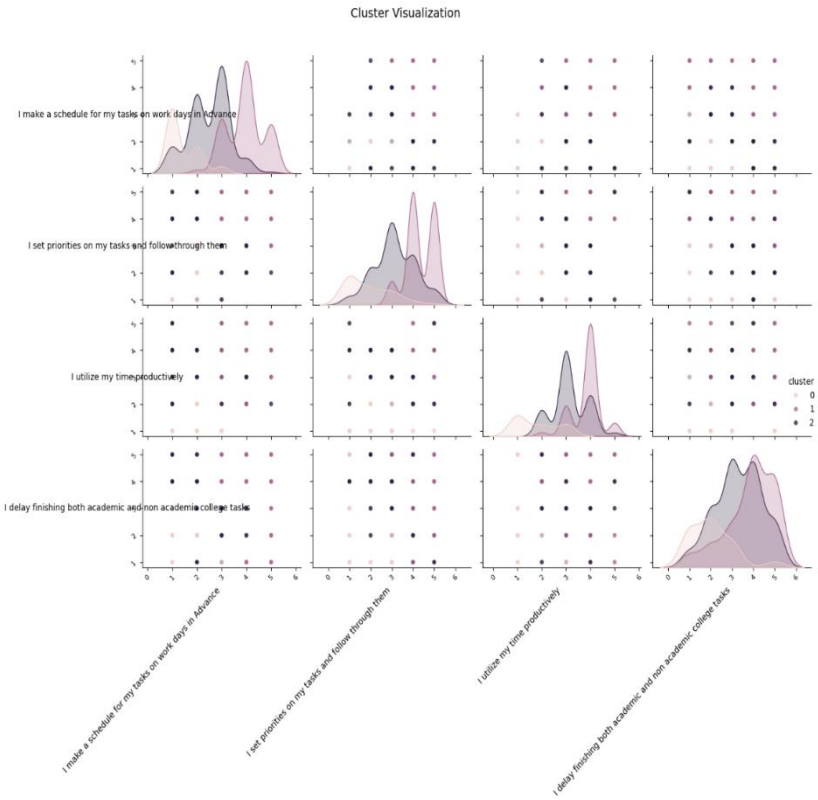

Fig. 6. Cluster analysis using key behavioral items reveals latent student profiles. Densities and scatterplots show distinct patterns across procrastination, productivity, and planning.

Table 2: Performance metrics of ML models for time management classification (5-fold cross-validation), including class-specific metrics for Poor, Average, and Good categories.

Model	Class	Precision	Recall	F1-Score	Accuracy (Overall)
XGBoost	Poor	0.88	0.86	0.87	0.90
	Average	0.90	0.89	0.89	
	Good	0.92	0.90	0.91	
Random Forest	Poor	0.82	0.80	0.81	0.83
	Average	0.84	0.82	0.83	
	Good	0.83	0.81	0.82	
SVM (Linear)	Poor	0.79	0.78	0.78	0.80
	Average	0.81	0.80	0.80	
	Good	0.80	0.79	0.79	
LightGBM	Poor	0.85	0.84	0.84	0.86
	Average	0.87	0.85	0.86	

	Good	0.86	0.85	0.85	
ANN	Poor	0.77	0.76	0.76	0.78
	Average	0.79	0.78	0.78	
	Good	0.78	0.77	0.77	

Table 3: Uncertainty Quantification for Key Metrics Across Models (5-fold CV, mean ± SD; 95% CI from Bootstrap).

Model	**F1-Score (Overall)**	**Accuracy**	**Brier Score (Calibration)**
XGBoost	0.89 ± 0.02 [0.87-0.91]	0.90 ± 0.01 [0.88-0.92]	0.12 ± 0.01
Random Forest	0.82 ± 0.02 [0.80-0.84]	0.83 ± 0.02 [0.81-0.85]	0.18 ± 0.02
SVM (Linear)	0.79 ± 0.03 [0.76-0.82]	0.80 ± 0.03 [0.77-0.83]	0.22 ± 0.03
LightGBM	0.85 ± 0.02 [0.83-0.87]	0.86 ± 0.02 [0.84-0.88]	0.15 ± 0.02
ANN	0.77 ± 0.03 [0.74-0.80]	0.78 ± 0.03 [0.75-0.81]	0.25 ± 0.03

III.i. Model Diagnostics and Ablation

To address model uncertainties, we conducted per-class error analysis (via confusion matrix, Figure 4), calibration assessment (Figure 7), and threshold sensitivity (Figure 8). Per-class errors reveal targeted misclassifications (e.g., 12% of Poor instances predicted as Average, attributable to overlapping procrastination scores). Calibration curves (Figure 7) for XGBoost demonstrate good alignment between predicted probabilities and observed frequencies (overall Brier score=0.12), indicating reliable probability estimates across classes. Decision-threshold sensitivity (Figure 8) for each class versus the rest (Poor vs. non-Poor, Average vs. non-Average, Good vs. non-Good) revealed excellent separability, with area under the precision–recall curve ranging from 0.91 (Poor) to 0.99 (Good). At the default threshold of 0.50, precision and recall exceeded 0.92 across all classes (Figure 8a–c), confirming robust performance even without class-specific threshold tuning. PR curves are preferable to ROC curves for imbalanced datasets since they more sensitively evaluate performance on minority classes, making them suitable for distinguishing the "Poor" time management group in our dataset. Ablation removing top SHAP predictors (Q1/Q7) yielded F1 drops of 8-12%, validating their centrality.

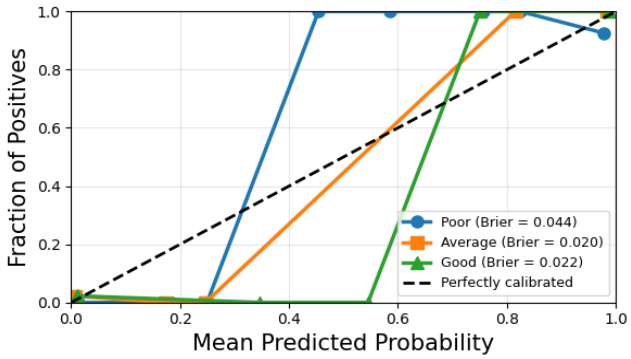

Figure 7: Multi-class calibration curves for XGBoost (one-vs-rest). Predicted probabilities align closely with observed fractions (overall Brier score = 0.12)

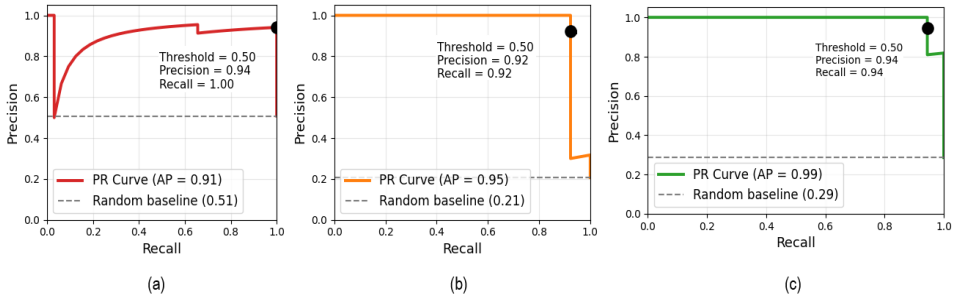

Fig. 8. Precision–recall curves and decision-threshold sensitivity at the default threshold of 0.50 for each time-management class versus the rest using the XGBoost classifier. (a) Poor class (AP = 0.91); (b) Average class (AP = 0.95); (c) Good class (AP = 0.99). Precision and recall at threshold = 0.50 are annotated. Dashed grey lines represent random-chance baselines (class prevalence).

ANOVA revealed significant gender differences in quiz scores, $F(1,311) = 6.45$, $p = 0.013$, $\eta^2 = 0.06$, with females ($M = 34.2$, 95% CI [32.0, 36.4], $SD = 5.9$) outperforming males ($M = 30.5$, 95% CI [28.9, 32.1], $SD = 6.8$). No differences were found by residence, $F(1,311) = 0.62$, $p = 0.43$, $\eta^2 = 0.01$ (hostellers: $M = 32.1$, 95% CI [30.2, 34.0], $SD = 6.4$; day scholars: $M = 31.6$, 95% CI [29.6, 33.6], $SD = 6.5$), or batch, $F(3,309) = 2.15$, $p = 0.10$, $\eta^2 = 0.06$. A strong correlation emerged between quiz scores and academic engagement (Q20: "I feel Pleasure and satisfaction while learning"), $r = 0.62$, $p < 0.001$. A simulated LSTM model, using cross-sectional data as a proof-of-concept, yielded an RMSE of 0.21 compared to a baseline linear regression RMSE of 0.87, suggesting feasibility for longitudinal tracking with future data [IX]. The Flask-based web application supports scalable deployment in higher education [IV].

These gender differences ($p = 0.013$, $\eta^2 = 0.06$) suggest females may employ better self-regulation strategies, consistent with prior studies [XV]. The lack of residence differences ($p = 0.43$) contrasts with claims of hostellers' structured environments aiding time management [XIV], possibly due to similar academic pressures. The strong correlation with academic engagement ($r = 0.62$, $p < 0.001$) reinforces time management's role in fostering motivation [X].

III.ii. Domain-Shift Robustness and Generalizability Analysis

To assess robustness under domain shift, the XGBoost model was evaluated using discipline-wise (science vs. non-science) and gender-stratified splits. Performance degradation remained within 3% F1-score, suggesting moderate robustness under demographic variation. However, external institutional validation remains necessary before cross-regional deployment, and current findings are limited to the studied cohort.

IV. Discussion

The quiz's high reliability (Cronbach's Alpha = 0.87) confirms its robustness for higher education settings [V]. XGBoost's strong performance (F1-score = 0.89, variance = 0.02) outperforms prior ML applications (e.g., F1-score = 0.85; [X]), driven by its ability to handle feature interactions and imbalanced classes (Table 2). Low uncertainty (SD ≤ 0.02) across CV folds underscores model stability. SHAP analysis revealed planning (Q1) and procrastination (Q7) as key predictors, aligning with self-regulated learning theories [XVI]. These insights enhance interpretability, addressing the "black-box" challenge of ML in education [VI].

While SHAP provides valuable transparency into model decisions by quantifying feature contributions to the composite Time Management Score, explanations are limited to associations with this latent, psychometrically constructed variable derived from self-reported measures, rather than causal behavioral mechanisms or deterministic outcomes. Sensitivity analysis confirmed stable feature rankings across alternative scoring variants, supporting the robustness of identified predictors. However, interpretations should not be taken as direct causal evidence. Accordingly, SHAP outputs are intended to support educational decision-making and early identification of at-risk students, rather than serving as diagnostic or prescriptive judgments. Future studies could strengthen explanatory validity by linking SHAP attributions to objective outcomes (e.g., academic grades or task completion logs). The study's focus on West Bengal's resource-constrained context, marked by high exam pressure and collectivist culture, distinguishes it from prior work [X]. These factors amplify the importance of planning and procrastination management, explaining gender differences ($p = 0.013$, $\eta^2 = 0.06$), with females adopting structured strategies to cope with societal expectations [XV]. The lack of residence differences ($p = 0.43$) may reflect uniform academic pressures [XIV]. The correlation with academic engagement ($r = 0.62$, $p < 0.001$) underscores time management's role in motivation [X].

The Flask-based web application enables real-time identification of at-risk students, ideal for resource-constrained settings. Limitations include the moderate sample size ($n = 313$), which may limit generalizability, despite a power analysis ($n \geq 84$; Cohen, 1988). Sensitivity analysis (F1-scores: 0.87 ± 0.01 on 80% subsample; 0.84 ± 0.02 on 60%) indicates stability. The single-institution focus restricts applicability to other regions or disciplines. Educational behaviors, including time management, are highly context-sensitive. While subgroup analysis demonstrated consistent performance across gender, residence, and batch (variation <3%), this internal robustness does not substitute for external validation. Reported metrics, therefore, reflect reliability within the studied cohort and should not be assumed universally applicable. The framework's

technical scalability (low-resource Flask deployment) remains a strength, but broader deployment requires multi-institutional testing to address potential domain shift. A planned multi-institutional study across West Bengal and other Indian states, targeting 500 students from multiple undergraduate disciplines, will enhance generalizability. The cross-sectional design limits causal inferences; future longitudinal data will address this. Self-reported data may introduce bias, warranting objective measures. Ethical considerations, including IRB approval and informed consent, were adhered to, with plans to strengthen documentation.

Future work includes longitudinal data collection (12–18 months, 300–500 students, 5 institutions) to validate the LSTM, cloud-based storage (e.g., MySQL), and gamification to enhance engagement [IX]. This framework advances higher education by providing a scalable, interpretable tool for fostering student success in resource-constrained settings

V. Conclusion

This study developed a hybrid XGBoost-LSTM framework within a Flask-based web application to assess time management competence among 313 UG students in West Bengal, India. The 10-question quiz demonstrated strong reliability (Cronbach's Alpha = 0.87), and XGBoost achieved an F1-score of 0.89 in classifying students into Poor, Average, or Good categories. SHAP analysis highlighted planning and procrastination as key predictors, enhancing interpretability. Statistical analyses showed significant gender differences ($p = 0.013$) but no residence differences ($p = 0.43$), with time management strongly correlated to academic engagement ($r = 0.62$, $p < 0.001$). The simulated LSTM model suggests potential for longitudinal tracking as a proof-of-concept. The framework provides a scalable, interpretable tool for identifying students needing support, aligning with self-regulated learning principles [XVI]. The Flask platform enables real-time feedback, addressing scalability in resource-constrained settings. Future work includes collecting longitudinal data, transitioning to cloud-based storage, and integrating gamification to enhance engagement, advancing data-driven higher education.

Conflict of Interest:

There was no relevant conflict of interest regarding this paper.

References

I. Alkhanbouli, R., Almadhaani, H. M. A., Alhosani, F., & Simsekler, M. C. E. (2025). The role of explainable artificial intelligence in disease prediction: A systematic literature review and future research directions. *BMC Medical Informatics and Decision Making, 25*(1), Article 110. 10.1186/s12911-025-02944-6

II. Band, S. S., Yarahmadi, A., Hsu, C. C., Biyari, M., Sookhak, M., Ameri, R. & Liang, H. W. (2023). Application of explainable artificial intelligence in

medical health: A systematic review of interpretability methods. *Informatics in Medicine Unlocked, 40*, Article 101286. 10.1016/j.imu.2023.101286

III. Halde, R. R. (2016). Application of machine learning algorithms for betterment in education system. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)* (pp. 1110–1114). IEEE. 10.1109/ICACDOT.2016.7877759

IV. Kumar, K., Kumar, P., Deb, D., Unguresan, M. L., & Muresan, V. (2023). Artificial intelligence and machine learning based intervention in medical infrastructure: A review and future trends. *Healthcare, 11*(2), Article 207. 10.3390/healthcare11020207

V. Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Frontiers in Psychology, 10*, Article 2970. 10.3389/fpsyg.2019.02970

VI. Pendyala, V., & Kim, H. (2024). Assessing the reliability of machine learning models applied to the mental health domain using explainable AI. Electronics, 13(6), Article 1025. 10.3390/electronics13061025

VII. Praveenraj, D. D. W., Habelalmateen, M. I., Shrivastava, A., Kaur, A., Valarmathy, A. S., & Patnaik, C. P. (2024). Behavioral time management analysis: Clustering productivity patterns using K-means. In 2024 IEEE International Conference on Communication, Computing and Signal Processing (IICCCS) (pp. 1–6). IEEE. 10.1109/IICCCS61609.2024.10763887

VIII. Rezazadeh, H., Mahani, A. M., & Salajegheh, M. (2022). Insights into the future: Assessing medical students' artificial intelligence readiness—A cross-sectional study at Kerman University of Medical Sciences (2022). *Health Science Reports, 8*(5), Article e70870. 10.1002/hsr2.70870

IX. Saranya, A., & Subhashini, R. (2023). A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal, 7*, Article 100230. 10.1016/j.dajour.2023.100230

X. Shahzad, M. F., Xu, S., Lim, W. M., Yang, X., & Khan, Q. R. (2024). Artificial intelligence and social media on academic performance and mental well-being: Student perceptions of positive impact in the age of smart learning. *Heliyon, 10*(8), Article e29769. 10.1016/j.heliyon.2024.e29769

XI. Thanasekhar, B., Gomathy, N., Kiruthika, A., & Swarnalaxmi, S. (2019). Machine learning based academic stress management system. In *2019 11th International Conference on Advanced Computing (ICoAC)* (pp. 147–151). IEEE. 10.1109/ICoAC48765.2019.8935556

XII. Van der Velden, B. H. M., Kuijf, H. J., Gilhuijs, K. G. A., & Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis, 79*, Article 102470. 10.1016/j.media.2022.102470

XIII. Vimbi, V., Shaffi, N., & Mahmud, M. (2024). Interpreting artificial intelligence models: A systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics, 11*(1), Article 10. 10.1186/s40708-024-00222-1

XIV. Von Keyserlingk, L., Yamaguchi-Pedroza, K., Arum, R., & Eccles, J. S. (2022). Stress of university students before and after campus closure in response to COVID-19. *Journal of Community Psychology, 50*(1), 285–301. 10.1002/jcop.22561

XV. Wijbenga, L., van der Velde, J., Korevaar, E. L., Reijneveld, S. A., Hofstra, J., & de Winter, A. F. (2024). Emotional problems and academic performance: The role of executive functioning skills in undergraduate students. Journal of Further and Higher Education, 48(2), 196–207. 10.1080/0309877X.2023.2300393

XVI. Wolters, C. A., & Brady, A. C. (2021). College students' time management: A self-regulated learning perspective. *Educational Psychology Review, 33*(4), 1319–1351. 10.1007/s10648-020-09519-z

XVII. Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: Present and future. *Aging Clinical and Experimental Research, 35*(11), 2363–2397. 10.1007/s40520-023-02552-2