

Enhancing Educational Quality with Explainable AI: Interpretable Prediction of Student Success

Ahmet ADALIER¹, Kian JAZAYERI¹, Moein JAZAYERI², Damla KARAGOZLU¹

¹ Department of Management Information Systems, Cyprus International University, Nicosia, North Cyprus.

² Department of Electrical and Electronics Engineering, Cyprus International University, Nicosia, North Cyprus.
aadalier@ciu.edu.tr; kjazayeri@ciu.edu.tr; mjazayeri@ciu.edu.tr; dkaragozlu@ciu.edu.tr

Abstract

Ensuring academic success and educational equity has become a critical priority for higher education institutions worldwide. This study investigates the determinants of academic success through interpretable machine learning and explainable artificial intelligence techniques. A dataset comprising 80,000 students was utilized to develop multiple regression models for GPA prediction, followed by SHAP dependence plots and Individual Conditional Expectation (ICE) plots to examine both global and instance-level contributions of behavioral, psychological, and environmental factors. Local interpretability was further explored through SHAP waterfall plots for selected students. Results indicate that study hours, motivation level, stress, sleep duration, study environment, and access to tutoring exert particularly strong influences on academic outcomes. The findings underscore that educational quality hinges on holistic student well-being beyond curriculum design alone. By delivering transparent, evidence-based explanations at both aggregate and individual levels, this research advances United Nations Sustainable Development Goal 4 (Quality Education) and offers higher education institutions actionable, data-driven guidance for equitable, student-centered enhancement strategies.

Keywords - Explainable Artificial Intelligence, Student Success, Learning Analytics, Higher Education Quality, Quality Education

1. Introduction

In recent years, the pursuit of academic excellence in higher education has been increasingly supported by data-driven approaches (Rodríguez-Ortiz et al., 2025). The availability of rich student data, combined with advances in machine learning (ML) and explainable artificial intelligence (XAI), has enabled educators and researchers to better understand the multifaceted factors influencing student performance (López-Meneses et al., 2025). However, the application of these technologies has often been limited to predictive accuracy, with less attention given to the interpretability and educational value of the models. To ensure meaningful and equitable educational improvement, models that not only predict but also explain learning outcomes (at both global and local levels) are essential.

The current study was designed to address this need by developing an interpretable ML-based framework capable of identifying the most influential behavioral, psychological, and environmental factors affecting students' academic performance. Using a large-scale dataset of 80,000 students that includes demographic, lifestyle, and learning habit indicators, the study employs SHapley Additive exPlanations (SHAP) dependence plots, Individual Conditional Expectation (ICE) plots, and local SHAP waterfall explanations to uncover how each feature contributes to variations in GPA and to illustrate individual-level prediction mechanisms. By prioritizing explainability over mere prediction, this approach aligns with the goals of evidence-based decision-making and transparency in educational quality enhancement.

This research contributes directly to United Nations Sustainable Development Goal 4 (Quality Education) by providing insights that promote data-informed educational interventions and institutional improvement. Through explainable modeling, the findings enable educators and policy-makers to design targeted strategies that nurture academic success, psychological well-being, and learning equity. Moreover, the study exemplifies how AI can be responsibly and ethically integrated into higher education systems to support continuous quality assurance and student-centered development.

2. Literature Review

Learning analytics (LA) and machine learning (ML) have emerged as central components of the data-driven transformation of higher education. Over the past decade, large-scale digitalization of learning environments and institutional information systems has generated vast quantities of student data, opening opportunities for predictive modeling, personalized learning, and quality enhancement (Sciarrone, 2018). The convergence of LA, educational data mining (EDM), and ML has enabled researchers to go beyond basic descriptive statistics towards modeling

complex behavioral, psychological, and contextual factors that shape learning outcomes. Within this paradigm, LA is viewed not merely as an evaluative tool but as an evidence-based mechanism for enhancing teaching, learning, and institutional decision-making.

Early work in LA primarily concentrated on integrating computational techniques from statistics, human–computer interaction, and data science to monitor and interpret learning processes (Gasevic et al., 2014). These studies emphasised two persistent challenges: the lack of analytics methods native to the educational domain and the limited reflection of social or pedagogical dimensions within computational models. The subsequent development of LA research addressed these issues by embedding ML algorithms into educational contexts to enhance prediction and intervention capabilities. Such efforts have yielded robust models for student classification, dropout prediction, and performance forecasting, which are core applications that remain central to modern educational analytics (Vinoth Kumar et al., 2025).

Parallel to methodological advances, the philosophical and epistemological foundations of learning analytics have also been re-examined. Doroudi (2024) argued that the field must adopt a constructivist epistemology that recognizes learning as a socially situated and interpretive process rather than a purely statistical phenomenon. Doroudi positioned different LA approaches along the bias–variance trade-off, proposing that the next generation of analytics should balance predictive precision with conceptual interpretability. This perspective aligns with the current movement towards transparent, explainable, and pedagogically grounded AI models.

Systematic reviews provide strong empirical backing for this methodological advancement. Rodríguez-Ortiz et al. (2025) conducted a comprehensive synthesis of 101 studies published between 2018 and 2025, demonstrating that ML predominantly drives current LA applications, especially in engagement prediction, dropout modelling, and academic performance forecasting. However, the review also indicated that generative AI (GenAI) techniques, such as transformer-based architectures including GPT-4 and BERT, are emerging in adaptive feedback and sentiment analysis but remain largely experimental due with transparency and implementation issues. Similarly, Ersozlu et al. (2024) examined 77 studies and found that 88 percent utilised supervised ML models (e.g., decision trees, support vector machines, random forests, and logistic regression) to forecast student performance. Their findings highlight the dominance of predictive modelling over interpretative analysis, while calling for greater methodological diversity and more substantial links to pedagogical theory.

Recent reviews on predictive learning analytics (PLA) confirm the increasing importance of ML in predicting academic success. Sghir et al. (2022) summarized a decade of PLA research, identifying a wide range of predicted outcomes (grades, retention, engagement) and data types (behavioural logs, assessment scores, demographic attributes). Despite significant progress, they pointed out a key gap: most models favour predictive accuracy over explainability, limiting their usefulness for educational decision-making. Alalawi et al. (2024) addressed this challenge by proposing a broader framework that combines pedagogical approaches with ML-based performance prediction and targeted intervention. Their Student Performance Prediction and Action (SPPA) model showed that integrating predictive analytics with instructor-led interventions can considerably enhance retention and pass rates, providing a replicable infrastructure for educators without needing institution-wide systems.

A parallel line of inquiry has examined the relationship between artificial intelligence and multimodal learning analytics. Mohammadi et al. (2025) reviewed 43 studies on AI-enhanced multimodal LA, concluding that while AI facilitates real-time feedback and customised learning, integration with learning theory and ethical governance remains limited. They identified gaps in deep AI implementation, data integration, and scalability, emphasising the need for interpretable and ethically aligned AI systems capable of managing complex, multimodal student data. These insights align with the broader educational movement towards transparency and fairness in AI applications.

Beyond higher education, machine learning has proven transformative for organizational and policy-level decision-making in development and quality assurance. Garbero et al. (2024) demonstrated how ML can be used to extract insights from large sets of project documentation, combining text mining and predictive analytics to strengthen data-driven evaluation frameworks. Their work parallels educational quality assurance systems, where diverse qualitative and quantitative indicators must be synthesized for informed decision-making. In the higher-education context, explainable AI (XAI) can similarly support continuous improvement cycles by linking institutional actions to measurable learning outcomes.

Complementary evidence links LA-driven analytics to student behavior, motivation, and well-being, which are variables that form the psychosocial foundation of learning. Walck-Shannon et al. (2021) found that the quality of learning strategies outweighs total study time, with active methods such as self-testing and elaboration predicting

higher exam performance. Aljaffer et al. (2024) reported that motivation, visual learning preferences, and memory recall significantly correlate with academic achievement among medical students. Lifestyle studies further reveal that balanced sleep and nutrition positively affect GPA (Mahfouz et al., 2024; Shafie et al., 2022), while irregular habits increase anxiety and stress (Rahimi et al., 2024; AlHamlan et al., 2025). Physical activity has been consistently associated with reduced stress and improved concentration, though standardized intervention protocols remain scarce (Guerrero et al., 2025).

Psychological health is a key factor in student success. Pascoe et al. (2019) established that academic strain directly harms well-being and learning outcomes. More recent research confirms that high stress and poor sleep quality increase the risks of anxiety and depression (Rahimi et al., 2024), while regular exercise helps reduce these effects. These behavioural insights highlight the importance of including psychological and physiological factors in predictive and explanatory models of academic performance. De Filippis and Foyals (2024) showed that anxiety and sleep quality are among the most important predictors of student stress, emphasising the need for comprehensive analytical frameworks that consider cognitive, emotional, and environmental factors.

The COVID-19 pandemic further highlighted how these factors are interconnected. Studies by Salazar-Granizo et al. (2024a, 2024b) showed that extended online learning disrupted healthy routines and increased stress levels, while face-to-face settings supported well-being. Pérez-Jorge et al. (2025) found that overload and imbalance of time were major stressors, but structured planning and emotional support helped lessen negative effects. These findings have led to international calls for universities to develop resilience-focused quality strategies that include both technological innovation and student well-being monitoring.

Taken together, the literature shows a clear shift from data collection and performance prediction towards interpretability, ethics, and holistic educational quality. Contemporary frameworks such as Rodríguez-Ortiz et al. (2025) and Mohammadi et al. (2025) agree on the importance of transparent and explainable AI to ensure that analytics improve, rather than hinder, human understanding of learning. This reflects a broader change in educational research from algorithmic optimisation to making meaning and gaining actionable insights. In this context, explainable ML techniques like SHapley Additive exPlanations (SHAP) are an important methodological step, enabling educators to trace how individual behaviours and environmental factors influence learning outcomes.

Despite these advances, most earlier studies treat prediction and explanation as separate aims. Few have systematically modelled behavioural, psychological, and environmental factors together within understandable frameworks. This study addresses this gap by using SHAP-based analysis of large-scale student data to show how daily habits, motivation, stress, sleep, and study environment together influence academic success. By combining interpretable ML methods with institutional quality goals, this research builds on previous work in learning analytics towards a clearer, more human-centred approach that directly supports the aims of United Nations Sustainable Development Goal 4 (Quality Education).

3. Methodology

3.1 Dataset and Preprocessing

The dataset used in this study is the *Student Habits and Academic Performance* dataset (Kumar, 2025). It contains data on 80,000 students' demographic characteristics, lifestyle habits, study environments, and academic performance. The dataset includes both numerical and categorical features, with "GPA" serving as the target variable representing students' academic success.

The features include demographic features (*age*, *gender*, *family_income_range*, *parental_education_level*), lifestyle and behavioral features (*study_hours_per_day*, *social_media_hours*, *netflix_hours*, *sleep_hours*, *diet_quality*, *exercise_frequency*, *screen_time*), academic features (*semester*, *major*, *attendance_percentage*, *access_to_tutoring*, *study_environment*, *time_management_score*, *learning_style*), psychological and emotional features (*stress_level*, *mental_health_rating*, *exam_anxiety_score*, *motivation_level*), and social and environmental features (*part_time_job*, *extracurricular_participation*, *social_activity*, *parental_support_level*, *internet_quality*). This categorization ensures that both internal and external factors influencing students' academic outcomes are comprehensively represented, covering demographic background, daily habits, cognitive and emotional states, academic behaviors, and social conditions.

Prior to analysis, the dataset was carefully cleaned and preprocessed. All rows containing missing values were removed to ensure data quality and model reliability. Non-numerical variables were converted into numerical form through logical enumerations. For instance, the *diet_quality* variable was encoded as Poor (0), Fair (1), and Good (2) to represent increasing levels of nutritional quality, while *study_environment* was ordered from less quiet to more quiet, Café (0), Dorm (1), Co-Learning Group (2), Library (3), and Quiet Room (4). This ordinal encoding

approach allows the model and subsequent SHAP analysis to capture meaningful progressions in the data rather than arbitrary categorical values.

All categorical enumerations applied during preprocessing are summarized in Table 1. Finally, all features were reviewed for consistency, ensuring that the resulting dataset was clean, standardized, and suitable for supervised regression modeling and SHAP-based interpretability analysis.

Table 1. Enumeration map of non-numerical features

Column Name	Category	Encoded As
gender	Male	0
	Female	1
	Other	2
major	Arts	0
	Business	1
	Psychology	2
	Biology	3
	Computer Science	4
	Engineering	5
part_time_job	No	0
	Yes	1
diet_quality	Poor	0
	Fair	1
	Good	2
parental_education_level	High School	0
	Some College	1
	Bachelor	2
	Master	3
	PhD	4
internet_quality	Low	0
	Medium	1
	High	2
extracurricular_participation	No	0
	Yes	1
study_environment	Cafe	0
	Dorm	1
	Co-Learning Group	2
	Library	3
	Quiet Room	4
access_to_tutoring	No	0
	Yes	1
family_income_range	Low	0
	Medium	1
	High	2
learning_style	Auditory	0
	Reading	1
	Visual	2
	Kinesthetic	3

3.2. Machine Learning and Explainability

To predict students' academic success, as measured by GPA, several supervised regression models were employed. Specifically, Random Forest, Logistic Regression, and Naïve Bayes algorithms were implemented to capture both linear and non-linear relationships between student characteristics and academic outcomes. Each model was trained using the preprocessed dataset described earlier, and performance was evaluated through mean absolute error (MAE) and mean absolute percentage error (MAPE).

After model comparison, the highest-performing model was selected for explainability analysis. To ensure transparency and interpretability of the predictions, SHAP was applied to this model. SHAP values quantify each feature's contribution to the predicted GPA, enabling identification of the factors that most strongly influence higher or lower academic performance.

The detailed results of the regression models and SHAP-based feature explanations are presented in the following section.

4. Results

4.1. Machine Learning Results

The performance of the three regression models, Random Forest, Logistic Regression, and Naïve Bayes, was evaluated and compared based on two key metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The results of this comparison are summarized in Table 2.

Table 2. Comparison of machine learning models' performances

	MAE	MAPE
Random Forest	0.3451	10.28%
Logistic Regression	0.4211	13.21%
Naïve Bayes	0.6452	20.27%

Among the tested models, the Random Forest regressor achieved the highest predictive accuracy, demonstrating superior ability to capture complex, non-linear relationships in the data. Specifically, it attained an MAE of 0.3451 and a MAPE of 10.28%, outperforming the other models by a clear margin.

Given its strong predictive performance and robustness, the Random Forest model was selected as the final model for the subsequent SHAP- and ICE-based explainability analysis, which is presented in the next subsection.

4.2. SHAP and ICE Explainability Results

To interpret the contribution of each feature to the predicted student GPA, SHAP analysis was applied to the Random Forest model. Two complementary visualizations were generated: the SHAP beeswarm plot (Figure 1, left) and the SHAP bar plot (Figure 1, right). Together, these plots reveal which features most strongly influence academic success and how their values affect GPA predictions.

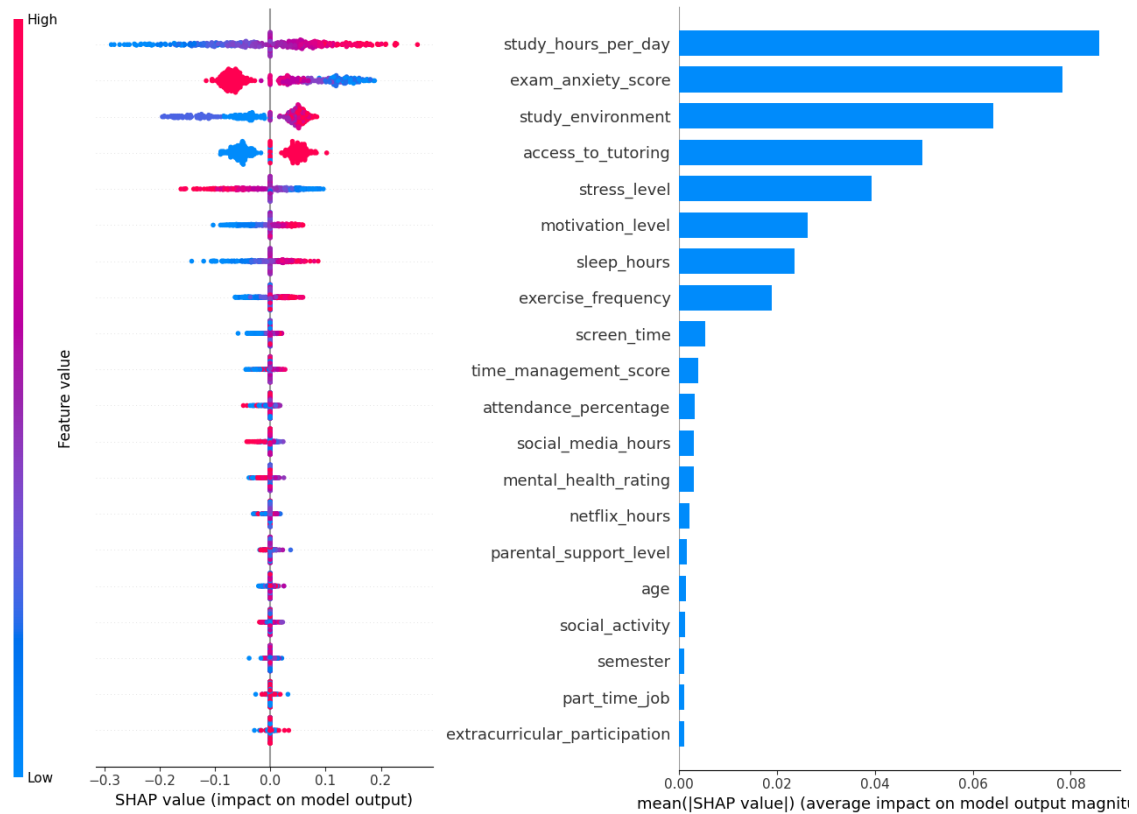


Figure 1. The SHAP Beeswarm (left) and Bar (right) plots

The beeswarm plot (Figure 1, left) provides a detailed, instance-level view of feature importance and direction of influence. The y-axis lists the features ranked by their overall importance, while the x-axis shows the SHAP value, representing each feature’s impact on the model’s output (i.e., predicted GPA). Dots to the right of the vertical line (positive SHAP values) indicate contributions that increase the predicted GPA, whereas dots to the left (negative SHAP values) indicate contributions that decrease it.

Each dot corresponds to a single student, and the color represents the feature value, pink for high values and blue for low values. For example, in the `study_hours_per_day` feature, pink dots appear predominantly on the right side, meaning that higher study hours are associated with higher GPAs. Conversely, for `exam_anxiety_score`, pink dots are concentrated on the left, showing that higher anxiety levels tend to lower GPA predictions.

The spread of dots along the x-axis indicates the magnitude of variation in that feature’s effect on GPA. Features such as `study_hours_per_day`, `exam_anxiety_score`, and `study_environment` show a wide spread, indicating strong and varied influences on academic performance across students. In contrast, features near the bottom, such as `part_time_job` or `extracurricular_participation`, have narrow distributions, implying limited impact on GPA.

The encoded values used for categorical variables (e.g., study environment encoded from 0 = Café to 4 = Quiet Room) enable this interpretation. Because the categories were ordered logically, higher encodings correspond to conditions generally more conducive to learning. This ordering makes the SHAP patterns meaningful. For instance, higher `study_environment` values (quieter settings) are associated with higher GPA predictions.

The SHAP bar plot (figure 1-right) aggregates the mean absolute SHAP values across all students to provide a global ranking of feature importance. The features are sorted from most to least influential based on their average impact on the model output.

According to the results, `study_hours_per_day` is the most important predictor of GPA, followed closely by `exam_anxiety_score`, `study_environment`, and `access_to_tutoring`. Students who dedicate more time to studying and have access to tutoring services tend to achieve higher academic performance, while those with higher anxiety levels generally perform worse. `Stress_level`, `motivation_level`, and `sleep_hours` also emerge as key determinants, highlighting the role of psychological and lifestyle factors in shaping academic success.

In contrast, features like `part_time_job`, `semester`, and `extracurricular_participation` exhibit relatively minor influence, suggesting that these external or temporal factors have limited direct effect on GPA compared to study habits and mental well-being.

To better understand how specific features affect student success, SHAP dependence plots and Individual Conditional Expectation (ICE) plots were generated for the top eight influential variables. Although the marginal contribution of any single feature, as reflected in the SHAP values, may appear modest (typically on the order of a 0.1 increase or decrease in predicted GPA) it should be emphasized that academic performance is shaped by the combined influence of multiple factors. When favorable (or unfavorable) conditions across several features align simultaneously, their cumulative effects can produce substantial shifts in predicted GPA, underscoring the multifaceted and interdependent nature of the determinants of student success.

The SHAP dependence plots illustrate how the actual value of each feature (x-axis) relates to its SHAP value (y-axis), which represents its positive or negative contribution to the predicted GPA. Each dot in these plots corresponds to an individual student. When SHAP values increase with the feature value, the relationship is positive, meaning that higher values of that feature contribute to a higher GPA. The ICE plots complement this analysis by showing the predicted GPA (y-axis) as a function of the feature value (x-axis) for individual cases, represented by thin lines, with a bold blue line indicating the average ICE across all instances. These plots reveal both the average effect and the heterogeneity in how changes in the feature influence predictions across students. In the dependence plot for `study_hours_per_day` (Figure 2-a), a clear positive trend is visible. As daily study hours increase, SHAP values also rise, indicating that longer study durations contribute positively to GPA predictions. Students who study fewer than approximately 4 hours per day generally have negative SHAP values, suggesting a negative effect on predicted GPA, while those who study more than 4 hours per day show a consistent positive contribution. This pattern highlights the strong influence of regular study habits on academic success. The corresponding ICE plot (Figure 2-b) reinforces this trend, with the average ICE line increasing steadily from approximately 3.3 to 3.7 as study hours rise from 0 to 7, demonstrating that higher study time generally leads to higher predicted GPAs. Individual lines exhibit some variation, but the overall upward trend indicates a consistent positive impact across most students.

In Figure 2-c, the SHAP dependence plot for `exam_anxiety_score` shows a clear negative relationship between exam anxiety and predicted GPA. As the anxiety score increases along the x-axis, the SHAP values decrease, indicating that higher anxiety levels contribute negatively to academic performance. Students with lower anxiety scores (around 5-6) have distinctly positive SHAP values, meaning their exam calmness positively influences GPA predictions. In contrast, when anxiety levels rise above 8, SHAP values drop below zero, reflecting a detrimental effect on predicted GPA. This pattern suggests that excessive exam anxiety undermines students' academic outcomes, while maintaining moderate or low anxiety supports better performance. The ICE plot in Figure 2-d further illustrates this, with the average ICE line declining from about 3.7 to 3.3 as anxiety scores increase from 5 to 10. The spread of individual lines highlights heterogeneity, where some students experience steeper declines in predicted GPA with rising anxiety, emphasizing the variable sensitivity to this factor.

In Figure 2-e, the SHAP dependence plot for `study_environment` reveals a non-linear trend, indicating that quieter and more structured study settings are generally linked to higher predicted GPAs. The x-axis represents the ordered study environments-Café (0), Dorm (1), Co-Learning Group (2), Library (3), and Quiet Room (4), while the y-axis indicates their contribution to GPA predictions. Students studying in cafés and especially dorms tend to have negative SHAP values, suggesting these environments negatively impact academic performance. The negative influence is stronger for dorms than cafés, likely due to the higher levels of distraction and lack of dedicated study space typically found in dorm rooms. In contrast, environments such as co-learning groups, libraries, and quiet rooms show less negative or positive SHAP values, indicating that structured and distraction-free spaces enhance learning effectiveness and contribute to higher GPAs. The ICE plot in Figure 2-f complements this by depicting the average ICE line starting around 3.5 in café environments, dipping to approximately 3.4 in dorms, and then rising to about 3.6 for quieter settings. Individual lines exhibit variation, with some showing pronounced dips in predicted GPA for less optimal environments, underscoring the context-dependent effects on student outcomes.

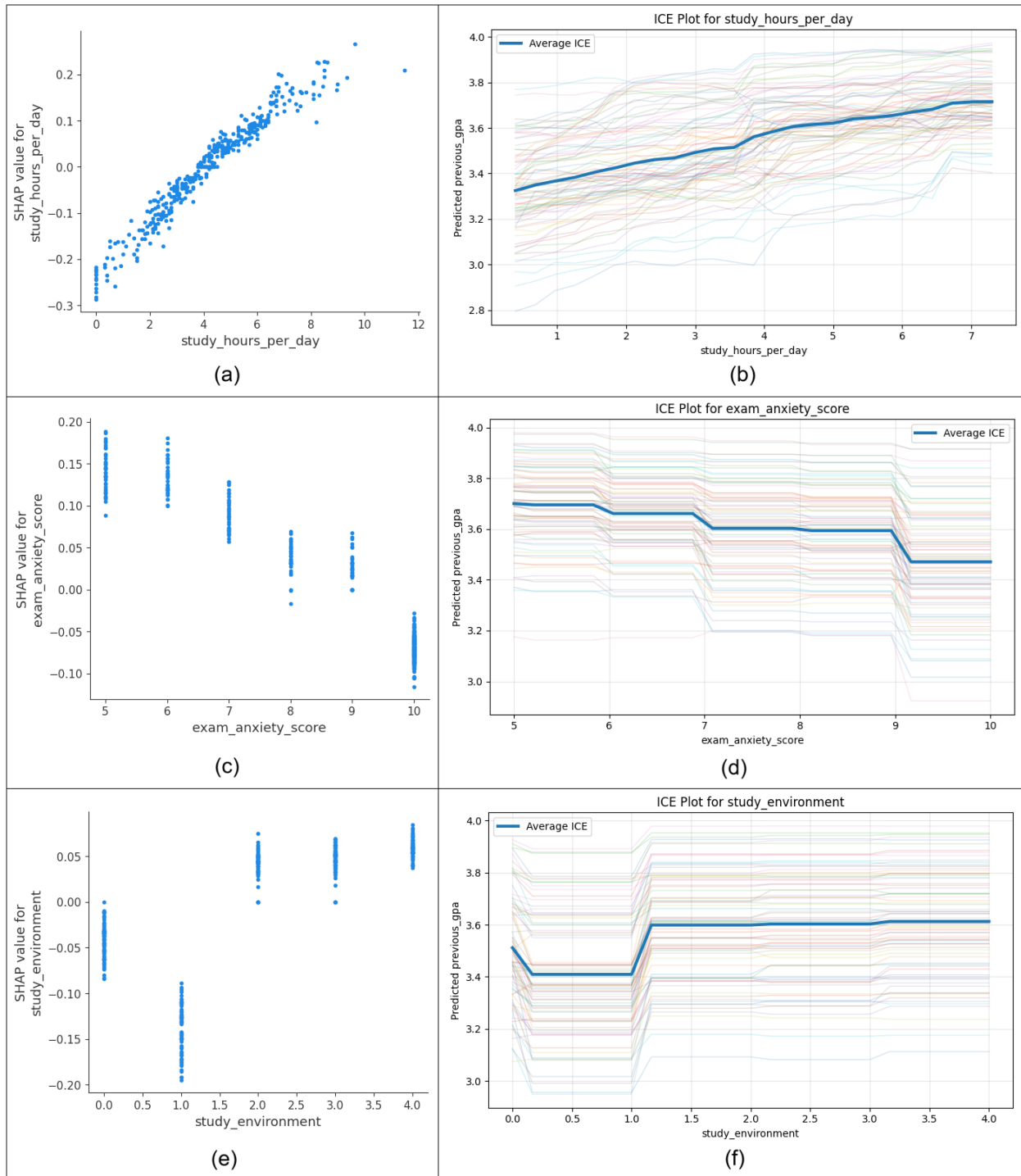


Figure 2. SHAP dependence and ICE plots for study hours per day (a, b), exam anxiety score (c, d), and study environment (e, f)

In Figure 3-a, the SHAP dependence plot for `access_to_tutoring` highlights a clear distinction between students with and without tutoring support. The feature is binary, where 0 represents no access to tutoring, and 1 indicates access to tutoring services. The plot shows that students without access to tutoring generally have negative SHAP values, meaning this condition contributes to lower predicted GPAs. Conversely, those with access to tutoring exhibit positive SHAP values, reflecting a significant positive effect on academic performance. This pattern underscores the importance of academic support mechanisms. Students who engage in tutoring or supplemental instruction are more likely to overcome learning challenges and achieve higher GPAs. The ICE plot in Figure 3-b supports this binary effect, with the average ICE line jumping from approximately 3.5 at no access to 3.6 with access. Individual lines cluster around these levels, indicating a relatively uniform positive shift in predicted GPA when tutoring is available.

In Figure 3-c, the SHAP dependence plot for `stress_level` shows a strong negative relationship between stress and predicted GPA. As stress levels increase along the x-axis, SHAP values steadily decline, indicating that higher stress contributes to lower academic performance. Students with stress levels below 4 generally have positive SHAP values, suggesting a beneficial or at least neutral influence on GPA. However, beyond this threshold, particularly at stress levels above 6, SHAP values drop sharply into the negative range. This pattern demonstrates that while moderate stress may have minimal impact, excessive stress severely hinders students' ability to perform well academically. The ICE plot in Figure 3-d confirms the negative trend, with the average ICE line decreasing from about 3.6 to 3.5 as stress rises from 0 to 10. The individual lines vary in slope, revealing that some students are more adversely affected by increasing stress than others.

In Figure 3-e, the SHAP dependence plot for `motivation_level` displays a clear positive association with student success. As motivation increases along the x-axis, SHAP values rise steadily, indicating that higher motivation consistently contributes to better academic outcomes. Students with motivation levels below 4 tend to have negative SHAP values, indicating a detrimental effect on GPA, whereas those scoring above 6 exhibit increasingly positive SHAP values. This upward trend suggests that maintaining high motivation is one of the strongest psychological factors supporting academic success, as motivated students are more likely to engage effectively in learning and sustain productive study behaviors. The ICE plot in Figure 3-f illustrates this positivity, with the average ICE line ascending from approximately 3.5 to 3.6 as motivation increases from 0 to 10. Individual lines show parallel increases, though with some dispersion, highlighting consistent but varying degrees of benefit across students.

In Figure 4-a, the SHAP dependence plot for `sleep_hours` reveals a positive, non-linear relationship between sleep duration and predicted GPA. As sleep hours increase from 4 to around 7-8 hours, SHAP values rise sharply, indicating that adequate sleep contributes significantly to better academic performance. Students sleeping fewer than 6 hours tend to have negative SHAP values, suggesting that sleep deprivation negatively affects GPA predictions. Beyond approximately 8 hours, the SHAP values plateau, implying that while sufficient rest is beneficial, excessive sleep does not provide additional academic advantage. This trend highlights that maintaining a balanced sleep schedule (around 7 to 8 hours per night) optimally supports learning and cognitive performance. The ICE plot in Figure 4-b depicts this non-linearity, with the average ICE line rising from about 3.5 at low sleep hours to a plateau around 3.6 at 7-9 hours. Individual lines follow similar patterns but with variations in the point of plateau, indicating personalized optimal sleep ranges.

In Figure 4-c, the SHAP dependence plot for `exercise_frequency` shows a clear positive relationship between physical activity and predicted GPA. As the number of exercise days per week increases along the x-axis, SHAP values gradually rise, indicating that more frequent exercise contributes positively to academic performance. Students who exercise fewer than two times per week generally have negative SHAP values, reflecting lower predicted GPAs, while those engaging in regular exercise (five or more times per week) consistently show positive SHAP contributions. This pattern suggests that maintaining an active lifestyle not only supports physical health but also enhances cognitive functioning, concentration, and overall academic success. The ICE plot in Figure 4-d reinforces this, with the average ICE line increasing from approximately 3.5 to 3.6 as exercise frequency rises from 0 to 7. The spread of individual lines suggests heterogeneity, where the positive effect is more pronounced for some students than others.

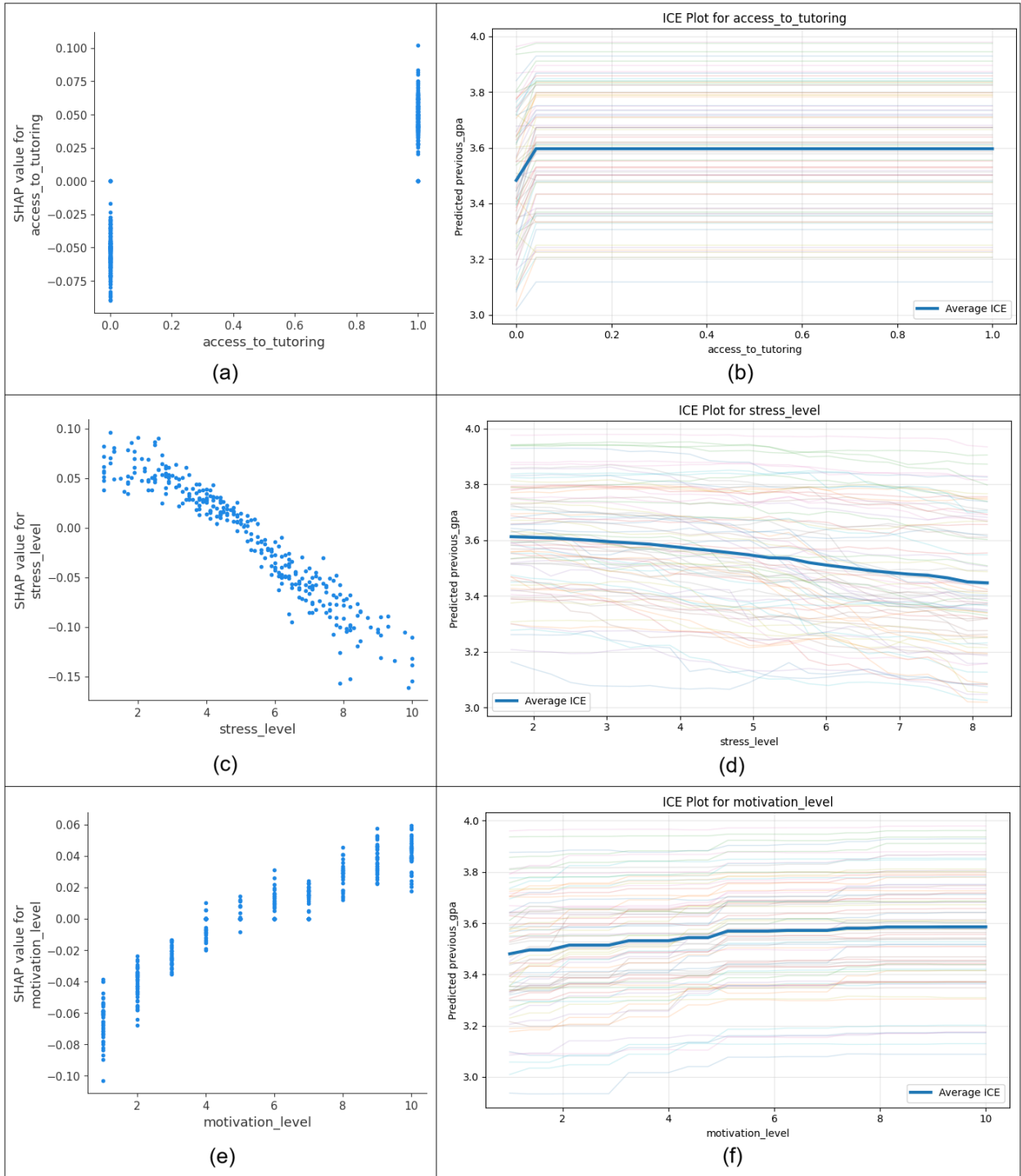


Figure 3. SHAP dependence and ICE plots for access to tutoring (a, b), stress level (c, d), and motivation level (e, f)

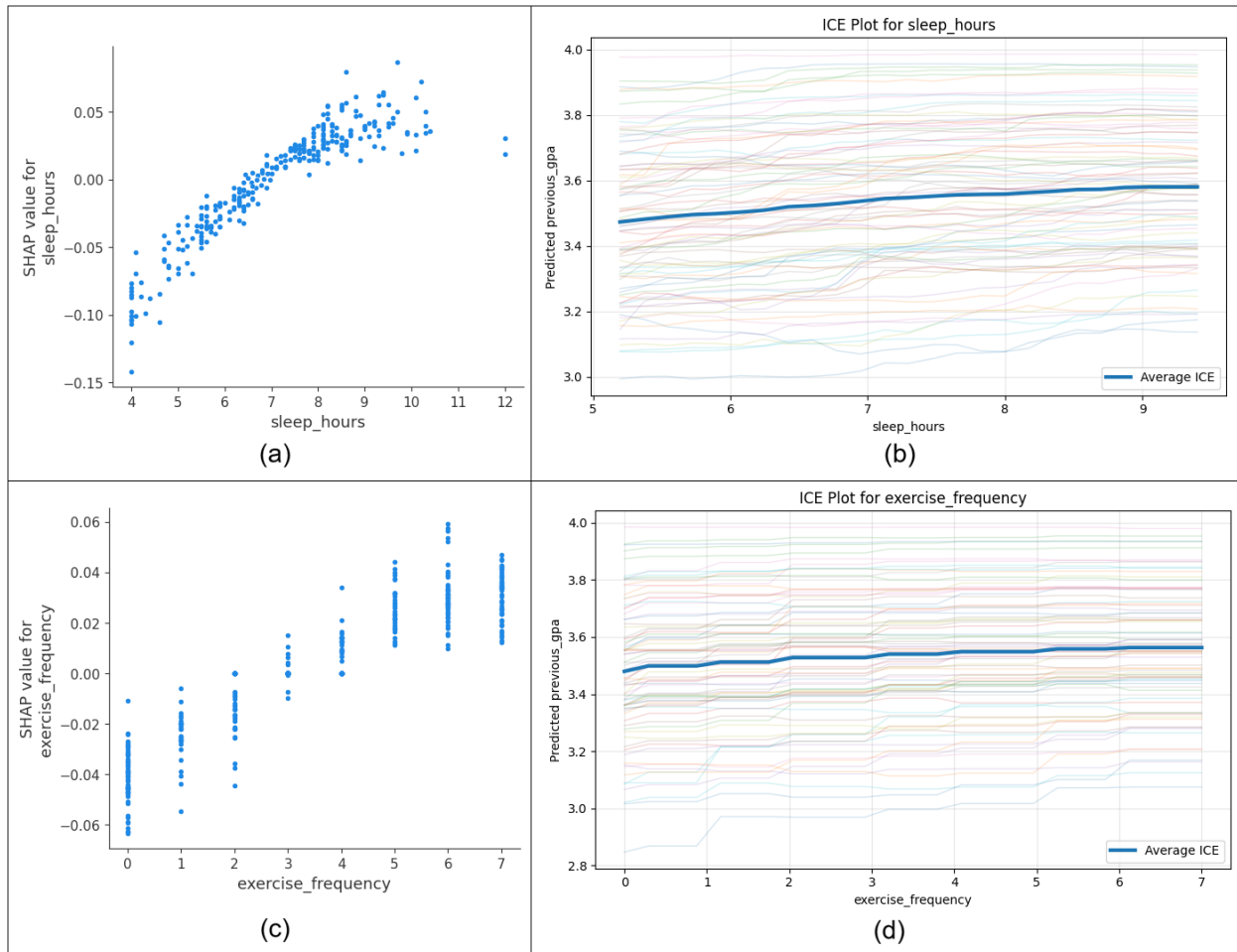


Figure 4. SHAP dependence and ICE plots for sleep hours (a, b) and exercise frequency (c, d)

To further elucidate the local interpretability of the model, SHAP waterfall plots were employed to examine the feature contributions for three randomly selected students. These plots demonstrate how each feature's SHAP value modifies the base expected GPA of 3.604, resulting in the final predicted GPA for the individual student. Positive contributions are shown in blue and increase the prediction, whereas negative contributions are shown in red and decrease it. The analysis begins with a student whose features collectively elevate the predicted GPA above the baseline (Figure 5), proceeds to a student in whom positive and negative contributions largely offset one another, yielding a predicted GPA near the expected value (Figure 6), and concludes with a student whose features predominantly reduce the predicted GPA below the baseline (Figure 7).

In Figure 5, the waterfall plot illustrates a case where the predicted GPA is elevated to 3.936. The most influential positive drivers are a moderate exam anxiety score and a relatively high number of daily study hours, both of which contribute substantially to raising the prediction above the baseline. These effects are further supported by adequate sleep duration, very high motivation, a reasonably favorable study environment (Co-Learning Group), frequent exercise, good mental health, and limited social media use. The remaining features have only negligible or neutral influence. This configuration demonstrates that a balanced psychological state, characterized by manageable anxiety levels, combined with strong study discipline, sufficient rest, intrinsic motivation, and healthy lifestyle habit can collectively produce a meaningful improvement in predicted academic performance.

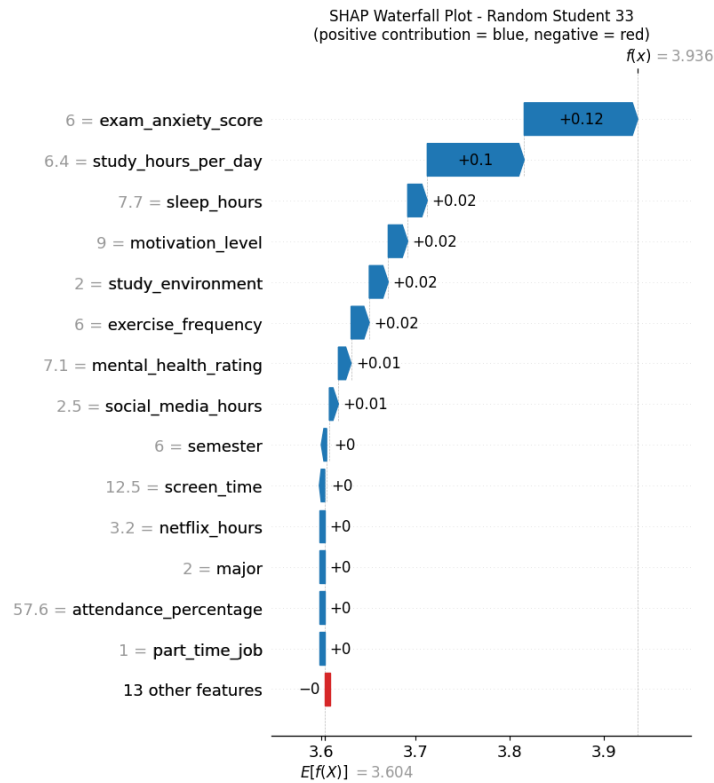


Figure 5. SHAP Waterfall Plot for a Random Student with Features Pushing Predicted GPA Higher (Student 33).

In Figure 6, the waterfall plot depicts a student whose predicted GPA remains close to the expected value at 3.642, reflecting a near-neutral net effect. The prediction is bolstered by a solid amount of daily study time, an optimal study environment (Quiet Room), moderate stress (which in this range appears to exert a mildly positive influence), and sufficient sleep. Smaller positive contributions arise from age, major, time management ability, attendance, diet quality, and limited recreational screen time. However, these gains are largely counteracted by very high exam anxiety, extremely low motivation, and lack of access to tutoring. This example highlights how strengths in behavioral and contextual factors can be substantially undermined by intense psychological distress and insufficient motivational drive or academic support, resulting in an outcome that deviates only minimally from the average prediction.

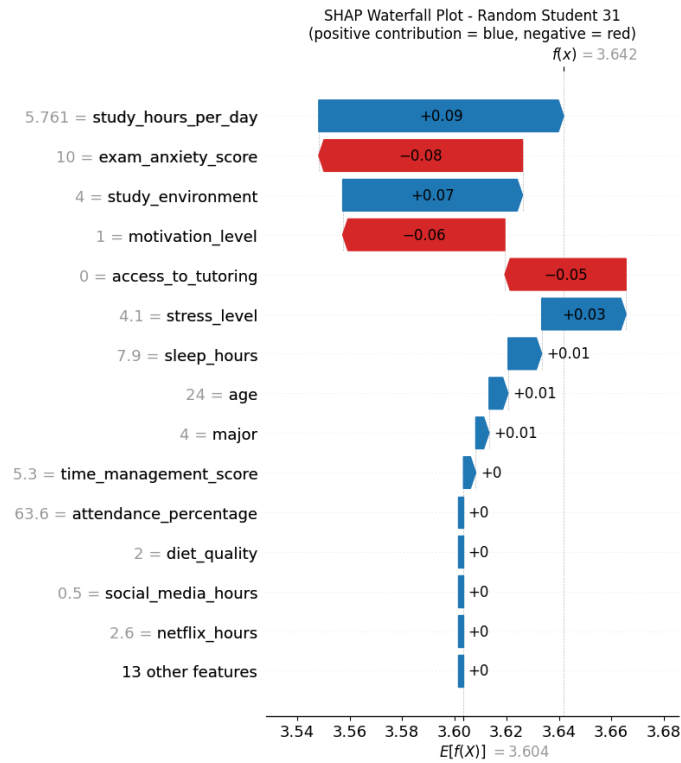


Figure 6. SHAP Waterfall Plot for a Random Student with Features Largely Canceling Each Other (Student 31).

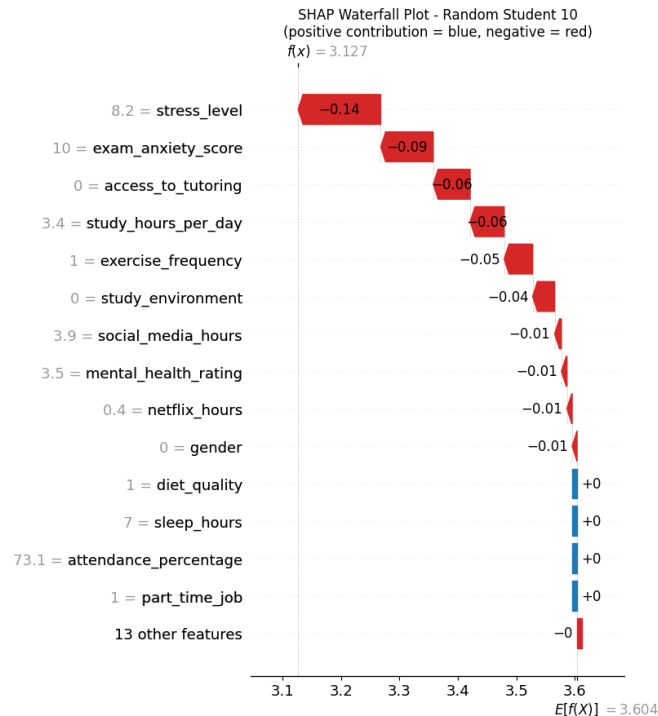


Figure 7. SHAP Waterfall Plot for a Random Student with Features Pushing Predicted GPA Lower (Student 10).

In Figure 7, the waterfall plot reveals a student whose predicted GPA is reduced to 3.127. The strongest negative influences stem from very high stress levels and maximum exam anxiety, both of which exert pronounced downward pressure on the prediction. Additional detrimental effects arise from insufficient daily study time, infrequent exercise, a highly distracting study environment (Café), excessive social media use, poorer mental health, and suboptimal diet quality. The absence of tutoring access further compounds the negative impact. Features such as sleep duration, attendance, part-time employment status, and most remaining variables contribute negligibly or neutrally and fail to offset the cumulative adverse effects. This profile underscores the severe

consequences of elevated psychological strain combined with inadequate study habits, poor environmental conditions, limited physical activity, and lack of academic support mechanisms.

5. Discussion

The findings of this study offer actionable insights for advancing Quality Education (UN SDG 4) through evidence-based decision-making in higher education. By integrating interpretable machine learning with both global (SHAP beeswarm/bar/dependence and ICE plots) and local (SHAP waterfall) explainability techniques, the research illustrates how large-scale student data can be transformed into transparent and meaningful indicators of academic well-being and performance. The results consistently highlight the strong influence of behavioral, psychological, and environmental factors (most notably study hours, motivation level, stress, sleep duration, study environment, and access to tutoring) on GPA outcomes. These patterns should be carefully considered by higher education institutions and policy-makers seeking to strengthen learning quality, student support systems, and academic equity.

From an institutional perspective, the interpretable patterns revealed suggest the value of developing data-informed student success frameworks. Universities can implement personalized advising systems capable of identifying students at risk due to insufficient study engagement, elevated stress or anxiety, suboptimal sleep, or limited access to support services, and then guide them toward targeted interventions such as structured study programs, counseling, tutoring, or improved study spaces. For instance, the observed thresholds, such as markedly reduced predicted performance below four hours of daily study or in highly distracting environments, underscore the potential benefits of proactive measures to enhance study conditions and well-being resources.

At the policy level, the insights call for integrating learning analytics and well-being monitoring into institutional and national quality assurance frameworks. Explainable AI tools, including both aggregate-level dependence/ICE visualizations and instance-specific explanations, can support continuous evaluation of student learning conditions, mental health indicators, and resource accessibility. Funding priorities could accordingly be directed toward expanding tutoring availability, mental health services, and initiatives that promote healthy lifestyle habits and motivation, all of which have been shown to exert measurable effects on academic outcomes.

Ultimately, the study reinforces that quality in higher education extends far beyond curriculum design or institutional metrics and encompasses the holistic support of the learner. The adoption of transparent, explainable AI approaches enables universities to design interventions that are equitable, evidence-based, and aligned with the human-centered objectives of SDG 4, thereby nurturing inclusive and high-quality education capable of empowering all students to achieve their potential.

6. Conclusion

The results of this study demonstrate that interpretable machine learning, augmented by SHAP dependence plots, ICE plots, and local SHAP waterfall explanations, can yield deep and actionable understanding of the determinants of student success. It has been shown that academic performance is significantly shaped by a combination of behavioral factors (study hours, sleep duration, exercise frequency), psychological dimensions (motivation level, exam anxiety, stress, mental health), and environmental conditions (study environment, access to tutoring), rather than by cognitive ability alone. These influences can be monitored and targeted through an institutional intervention.

From a quality standpoint, the presented framework offers a practical and transparent pathway toward realizing UN SDG 4, Quality Education. By employing explainable AI techniques at both global and individual levels, universities can monitor learning conditions, identify at-risk students early, optimize support services, and cultivate healthier academic environments in a manner that is fair, interpretable, and student-centered. Such approaches ensure that data-driven policies remain accountable and aligned with educational equity goals.

In conclusion, integrating explainable artificial intelligence into learning analytics represents a significant advancement in educational quality management. When applied ethically and responsibly, these methods empower decision-makers to convert complex student data into targeted, evidence-informed actions that enhance learning outcomes. Accordingly, this study contributes both methodological rigor and strategic guidance to the continuing efforts of higher education institutions and policy-makers to deliver inclusive, equitable, and high-quality educational experiences in accordance with the principles of SDG 4.

Note: This article is an extended and revised version of the paper titled "Quality-Centered Learning Analytics: Interpretable Modeling of Students' Success with Explainable AI," presented at the International Conference on

Quality in Higher Education in November 2025. The present work incorporates substantial new content, additional analyses, updated sections, and minor changes to the title and structure.

References

- Alalawi, K., Athauda, R., & Chiong, R. (2024). An Extended Learning Analytics Framework Integrating Machine Learning and Pedagogical Approaches for Student Performance Prediction and Intervention. *International Journal of Artificial Intelligence in Education*, 35(3), 1239–1287. <https://doi.org/10.1007/s40593-024-00429-7>
- AlHamlan, A. W., AlDhafiri, D. A., AlHajri, N. A., AlHouli, A. H., AlHubaidah, S. M., AlJassar, A. M., AlMutairi, N. N., AlRabiah, A. S., Al-Sultan, A. T., & Qasem, W. (2025). The impact of academic stress on the lifestyle of university students in Kuwait. *BMC Public Health*, 25(1). <https://doi.org/10.1186/s12889-025-23569-4>
- Aljaffer, M. A., Almadani, A. H., AlDughaiter, A. S., Basfar, A. A., AlGhadir, S. M., AlGhamdi, Y. A., AlHubaysh, B. N., AlMayouf, O. A., AlGhamdi, S. A., Ahmad, T., & Abdulghani, H. M. (2024). The impact of study habits and personal factors on the academic achievement performances of medical students. *BMC Medical Education*, 24(1). <https://doi.org/10.1186/s12909-024-05889-y>
- de Filippis, R., & Foysal, A. A. (2024). Comprehensive analysis of stress factors affecting students: a machine learning approach. *Discover Artificial Intelligence*, 4(1). <https://doi.org/10.1007/s44163-024-00169-6>
- Doroudi, S. (2024). On the paradigms of learning analytics: Machine learning meets epistemology. *Computers and Education: Artificial Intelligence*, 6, 100192. <https://doi.org/10.1016/j.caeai.2023.100192>
- Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, 29(16), 22125–22145. <https://doi.org/10.1007/s10639-024-12704-0>
- Garbero, A., Carneiro, B., Cerulli, G., & Resce, G. (2024). Leveraging artificial intelligence and machine learning analytics for development effectiveness: a comprehensive approach to extract insights from project data. *Journal of Development Effectiveness*, 17(1), 71–100. <https://doi.org/10.1080/19439342.2024.2361008>
- Gasevic, D., Rose, C., Siemens, G., Wolff, A., & Zdrahal, Z. (2014). Learning analytics and machine learning. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 287–288). LAK '14: Learning Analytics and Knowledge Conference 2014. ACM. <https://doi.org/10.1145/2567574.2567633>
- Guerriero, M. A., Dipace, A., Monda, A., De Maria, A., Polito, R., Messina, G., Monda, M., di Padova, M., Basta, A., Ruberto, M., Capasso, E., Moscatelli, F., & Limone, P. (2025). Relationship Between Sedentary Lifestyle, Physical Activity and Stress in University Students and Their Life Habits: A Scoping Review with PRISMA Checklist (PRISMA-ScR). *Brain Sciences*, 15(1), 78. <https://doi.org/10.3390/brainsci15010078>
- Kumar, A. (2025). *Student Habits and Academic Performance Dataset* [Dataset]. Kaggle. <https://www.kaggle.com/datasets/aryan208/student-habits-and-academic-performance-dataset>
- López-Meneses, E., López-Catalán, L., Pelicano-Piris, N., & Mellado-Moreno, P. C. (2025). Artificial Intelligence in Educational Data Mining and Human-in-the-Loop Machine Learning and Machine Teaching: Analysis of Scientific Knowledge. *Applied Sciences*, 15(2), 772. <https://doi.org/10.3390/app15020772>
- Mahfouz, H. A., Alhazmi, N. F., Almatrafi, M. K., Almeahmadi, S. S., Alharbi, J. K., Qadi, L. R., & Tawakul, A. (2024). The Influence of Lifestyle on Academic Performance Among Health Profession Students at Umm Al-Qura University. *Cureus*. <https://doi.org/10.7759/cureus.56759>
- Mohammadi, M., Tajik, E., Martinez-Maldonado, R., Sadiq, S., Tomaszewski, W., & Khosravi, H. (2025). Artificial intelligence in multimodal learning analytics: A systematic literature review. *Computers and Education: Artificial Intelligence*, 8, 100426. <https://doi.org/10.1016/j.caeai.2025.100426>
- Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2019). The impact of stress on students in secondary school and higher education. *International Journal of Adolescence and Youth*, 25(1), 104–112. <https://doi.org/10.1080/02673843.2019.1596823>
- Pérez-Jorge, D., Boutaba-Alehyan, M., González-Contreras, A. I., & Pérez-Pérez, I. (2025). Examining the effects of academic stress on student well-being in higher education. *Humanities and Social Sciences Communications*, 12(1). <https://doi.org/10.1057/s41599-025-04698-y>
- Rahimi, A., Wardak, M. F., & Shayan, N. A. (2024). Assessing the relationship between lifestyle factors and mental health outcomes among Afghan university students. *Journal of Affective Disorders Reports*, 17, 100827. <https://doi.org/10.1016/j.jadr.2024.100827>
- Rodríguez-Ortiz, M. Á., Santana-Mancilla, P. C., & Anido-Rifón, L. E. (2025). Machine Learning and Generative AI in Learning Analytics for Higher Education: A Systematic Review of Models, Trends, and Challenges. *Applied Sciences*, 15(15), 8679. <https://doi.org/10.3390/app15158679>

- Salazar-Granizo, Y. E., Hueso-Montoro, C., & Caparros-Gonzalez, R. A. (2024-a). Lifestyles and Academic Stress in University Students of Health Sciences: A Mixed-Methodology Study. *Healthcare*, 12(14), 1384. <https://doi.org/10.3390/healthcare12141384>
- Salazar-Granizo, Y. E., Hueso-Montoro, C., & Caparros-Gonzalez, R. A. (2024-b). Lifestyles and academic stress among health sciences students at the National University of Chimborazo, Ecuador: a longitudinal study. *Frontiers in Public Health*, 12. <https://doi.org/10.3389/fpubh.2024.1447649>
- Sciarrone, F. (2018). Machine Learning and Learning Analytics: Integrating Data with Learning. In 2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET) (pp. 1–5). 2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET). IEEE. <https://doi.org/10.1109/ithet.2018.8424780>
- Sghir, N., Adadi, A., & Lahmer, M. (2022). 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>
- Shafie, S. N. M., Shahri, M. S., Izuddi, N. N. N. I., Shukri, N. M., Aziz, N. A., Amran, A., & Nafi, M. N. A. (2022). Association between Lifestyle-Related Behaviors and Academic Performance among Students. In International Academic Symposium of Social Science 2022 (p. 105). International Academic Symposium of Social Science. MDPI. <https://doi.org/10.3390/proceedings2022082105>
- Vinoth Kumar, E. S., Augustian Isaac, R., Sundaravadivel, P., & Janaki Ramal, P. (2025). Machine learning-driven development of a behaviour-based student classification system (SCS-B) for enhanced educational analytics. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-21332-8>
- Walck-Shannon, E. M., Rowell, S. F., & Frey, R. F. (2021). To What Extent Do Study Habits Relate to Performance? *CBE—Life Sciences Education*, 20(1), ar6. <https://doi.org/10.1187/cbe.20-05-0091>