International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 861-876
https://doi.org/10.70917/ijcisim-2026-3316
Article

Enhancing Student Engagement Prediction Using a Hybrid Gradient Boosting Approach

Susan Nadhum Ahmed[1], Manar Joundy Hazar[2], Mustafa Radif[3]

[1] College of Computer Science and Information Technology, University of Al-Qadisiyah, Al-Qadisiyah 58002, Iraq.
Email: **susannadhum000@gmail.com**
[2] College of Computer Science and Information Technology, University of Al-Qadisiyah, Al-Qadisiyah 58002, Iraq.
Email: **Manar.joundy@qu.edu.iq**
[3] College of Computer Science and Information Technology, University of Al-Qadisiyah, Al-Qadisiyah 58002, Iraq.
Email: **Mustafa.Radif@qu.edu.iq**
Corresponding Author: Susan Nadhum Ahmed

Abstract: The most critical issue in educational technology is the ability to predict how students will engage in the e-learning setting, but the implications of this prediction on the student learning outcomes, rate of course completion, and effectiveness of the institution has become one of the most critical issues in modern teaching technology. The paper introduces a new hybrid gradient boosting model, which will be a combination of the complementary strengths of XGBoost and LightGBM into a single framework of ensemble models in predicting the level of student engagement in online learning platforms. Based on the data of behavioral interaction extracted out of the EdNet-KT4 LMS Student Engagement Dataset large-scale hierarchical educational dataset which is the most complete and richest of the four EdNet hierarchical levels, containing 297,915 student files, with a compressed size of 1.2GB and an uncompressed size of 6.4GB. The proposed model is systematically compared to five separate unique baseline algorithms, including Random Forest, XGBoost, LightGBM, Artificial Neural Network, and Logistic Regression. In addition we compare our work with other state of art on the same dataset. The proposed hybrid achieves 96.00% accuracy, precision of 1.000, F1-Score of 0.960, and AUC of 0.982. The findings show that synergistic performance improvement with the strategic combination of the second-order gradient optimization of XGBoost and the leaf-wise growth and the histogram-based efficiency of LightGBM is synergistic as neither algorithm can achieve the synergistic improvement performance by itself. The implications of the results in practical use in implementation of real-time and scalable engagement monitoring systems in online educational institutions have significant practical value.
Keywords: Student Engagement, XGBoost, LightGBM, Online Learning

1. Introduction

The rapid development of e-learning platforms in higher educational institutions all over the world has essentially changed the concept of academic instruction by offering unprecedented flexibility and accessibility to students on both sides of the geographical, temporal and socioeconomic borders . But this change has also presented one of the most enduring and impactful issues of contemporary education namely; ensuring sustained, meaningful student engagement over the course of the learning process. In contrast with the traditional classroom learning environment where instructors can observe behavioral indicators of engagement, and respond with real-time and personalized adjustments, online learning environments place a physical and temporal distance between instructors and learners making real-time engagement monitoring extraordinarily challenging without the aid of intelligent computational systems. The difference between the amount of behavioral data produced by the modern Learning Management Systems (LMS) and the ability of human instructors to analyze and take actions on the data is the main issue that drives this research .

Student interaction in online learning platforms is a multidimensional construct that simultaneously reflects behavioral, cognitive, and affective dimensions of interaction reflected in the interaction logs created within LMSs. Behavioral engagement can be seen through observable behaviors including the frequency of login, duration of the session, the rate of question attempt, video use and content completion patterns, the quality of responses, the rate of accuracy, the time, and the patterns of repeated errors illustrate the cognitive engagement and depth of processing and understanding the concept . Although it is more challenging to directly measure, affective engagement can be determined by behavioral trajectories that can indicate underlying emotional states, including motivation, interest, and frustration. The measured and integrated combination of these various dimensions of engagement using machine learning is the theoretical basis on which the current study is based on .

The use of machine learning to predict student engagement has increased in recent years. With more large-scale behavioral data sources (LMS platforms and institutional recognition that data-driven early warning systems can significantly enhance student retention) becoming more and more accessible . Gradient boosting variants of the family of machine learning algorithms used to solve this problem have proven themselves to be the most effective and have been consistently used to achieve high performance in the structured, tabular behavioral data used to characterize the LMS interaction logs. Although this is an excellent result, every single gradient boosting algorithm has its own set of strengths and weaknesses, limiting its performance on certain aspects of the prediction problem. A limitation that hybrid ensemble architectures is exceptionally well placed to address .

The main hypothesis of this paper is using a principled architecture of hybrid ensemble using both XGBoost and LightGBM, the systematically better prediction performance would be obtained than either of the two algorithms used independently. The combination of second-order gradient statistics and explicit L1/L2 regularization of XGBoost allows to achieve extremely high accuracy in optimization and high resistance to overfitting on small training data . The leaf-wise expansion approach and Gradient-based One-Side Sampling method of LightGBM permit the training of models dramatically faster and with competitive accuracy. The resultant optimized weighted averaging of these complementary strengths will result in a synergistic ensemble that inherits the best properties of both of them and compensates for the shortcomings of each algorithm's individual limitations .

This paper presents four specific contributions:

1. Proposes and implements a novel hybrid XGBoost-LightGBM architecture with an optimized weighted averaging scheme selected through cross-validation.
2. Provides a rigorous comparative analysis of the hybrid model against five individual baseline algorithms and prior state-of-the-art approaches under identical experimental conditions.
3. Offers an in-depth analysis of the performance gains produced by the hybrid approach and the theoretical mechanisms through which ensemble combination yields those gains.
4. Evaluates the computational feasibility of the hybrid model for real-time deployment by measuring training time empirically at the full dataset scale of 297,915 students.

2. Related Work

Student engagement prediction has received significant attention in learning analytics and educational data mining research. Early studies relied on traditional statistical techniques and rule-based indicators to identify disengaged or at-risk learners. However, such approaches were limited in capturing the complexity and temporal dynamics of student behavior in online environments . With the increasing availability of large-scale educational datasets, machine learning techniques have become the dominant approach for engagement and dropout prediction. Romero and Ventura provided a comprehensive survey demonstrating the effectiveness of classification models such as logistic regression, decision trees, support vector machines, and ensemble methods in educational contexts. Among these, ensemble models consistently show superior performance due to their ability to model non-linear relationships and interactions among behavioral features . In table 1 we summarize some of the most related study to our work in this field based on the method that was used and its applicability and what the key findings.

Table 1: Summary of related work on student engagement and performance prediction.

Category	Study (Year)	Algorithm / Method	Application	Key Findings
Traditional ML	Hossen & Uddin (2025)	Gradient Boosting + XAI	Online engagement classification	Accuracy ~94%, improved interpretability

	Dalal (2025)	Random Forest, SVM	Motivation & engagement prediction	85–90% accuracy, behavioral features key
	Althaqafi et al. (2025)	XGBoost + SMOTE	Performance with class imbalance	93% accuracy after handling imbalance
	Bellarhmouch et al. (2025)	ML classifiers (RF, DT, k-NN)	Online learning engagement	Effective detection from interaction logs
XGBoost	El-Beshbishi et al. (2025)	XGBoost	Medical student engagement	91.5% accuracy, outperformed logistic regression
	Goren et al. (2024)	XGBoost + Neural Networks	Student dropout prediction	AUC = 0.94, strong regularization benefits
	Poudyal (2022)	XGBoost	Educational data mining (3 case studies)	Consistently outperformed RF and DT
	Alazmi (2025)	Ensemble (incl. XGBoost)	Higher education success	91% accuracy for early warning systems
LightGBM	Hamim et al. (2022)	LightGBM	Student profile modeling	2.3× faster than XGBoost, similar accuracy
	Zhao et al. (2025)	LightGBM	Academic success from behavior	89% accuracy, best on borderline cases
	Runhaar (2024)	LightGBM	Secure engagement analysis	Training time reduced by 35%
Hybrid / Ensemble	Naeem et al. (2025)	Stacked Ensemble (RF, XGB, LGBM, ANN)	Virtual learning engagement	+4.2% accuracy over best single model
	Begum et al. (2025)	Adaptive Multi-Model Ensemble	Engagement prediction	+3.8% F1-score, adaptive weighting
	Fu et al. (2025)	AHP-Guided Stacked Ensemble	Engagement prediction	+5.1% improvement on borderline students
	Prasanth (2025)	TabNet + XGBoost	Student performance + feedback	+2.9% accuracy over XGBoost alone
	Wang et al. (2025)	Weighted Ensemble (RF, XGB, LGBM)	Engagement prediction	AUC improved from 0.94 to 0.97
	Arévalo-Cordovilla et al. (2025)	Ensemble + XAI	Academic performance	Transparency maintained with 93% accuracy
	Al Salmi et al. (2025)	Explainable Ensemble Framework	Classroom analytics	XGBoost + LightGBM with fairness constraints

	Farokhi et al. (2023)	Graph Representation Learning	MOOC grade prediction	88% accuracy using learner interaction graphs
Review / Other	Boujmiraz et al. (2026)	Review: ML, DL, XAI	Student performance	Confirmed ensembles and XAI as best practices
	Alfalasi (2026)	Behavioral + academic + parental data	Performance prediction	Hybrid features improved over single-domain
	Khan et al. (2026)	Rethinking engagement datasets	Virtual learning environments	Critiqued dataset limitations for AI
	Ovtšarenko et al. (2026)	AI-based forecasting	Technical education dropout	Sci. Rep. 2026, risk prediction validated

3. Dataset Description

The dataset used in this study is the LMS Student Engagement Dataset which is an organized behavioral interaction dataset based on the EdNet platform that can accessed on https://github.com/riiid/ednet and downloaded from bit.ly/ednet-kt4. EdNet offers datasets in four different levels, each named KT1, KT2, KT3 and KT4. As the level of the dataset increases, the number of actions and types of actions involved also increase . Here we work on KT4 because it provides raw interaction logs for each student separately within independent CSV files, allowing the extraction of behavioral features necessary to study student interaction, such as the number of logins, the number of responses, activity duration, and the completion rate of activities. Moreover, KT4 contains a large number of students (297,915 students), providing a massive and diverse database that helps in training machine learning models more efficiently and improving the generalizability of the results. In addition, the structure of KT4 makes the feature engineering process more flexible compared to some other parts of EdNet, as the data is stored at the student level, which directly aligns with the research objective of measuring and classifying students' interaction levels. Capturing detailed behavioral interaction patterns of 297,915 online university students . The student sample will be stratified in three levels of engagement: 89,375 students (30%) as High engagement, 119,166 students (40%) as Medium engagement and 89,375 students (30%) as Low engagement, realistic distributions of online course enrollment data. This stratification is to make sure that the machine learning models are trained and assessed on balanced representation of the engagement categories that reflect the natural diversity of the student populations in the online learning environments .

The behavioral characteristics that are recorded in the data are across various dimensions of student learning activity. Features based on the session are a mean session time (minutes) and time delay between sessions (days). The features of participation are the frequency of logs and views of videos. Some of the features of performance are questions attempted, correct answers, accuracy rate and repeated errors. Support-seeking characteristics comprise of the request of help, and the rate of the content completion is also used as a measure of the support-seeking characteristics. The average response time gives a proxy measure of cognitive effort. In addition to these raw measures of behavior, three engineered composite measures were built to capture higher-order patterns: engagement (a weighted average of five behavioral indicators) and learning (number of correct answers per unit session) and activity (number of questions attempted per login session). The whole set of features and the profile of behavior of each group of engagement are summarized in Table 2

Table 2: Dataset Characteristics and Interaction Group Behavioral Profiles.

Feature	Type	**High Engagement** (n=89,375)	**Medium Engagement** (n=119,166)	**Low Engagement** (n=89,375)
avg_session_duration (min)	Raw Behavioral	40–52	22–38	8–22
login_frequency (sessions)	Raw Behavioral	25–35	15–25	5–15

video_views (count)	Raw Behavioral	40–60	20–40	5–20
questions_attempted (count)	Raw Behavioral	150–250	80–150	20–80
correct_answers (count)	Raw Behavioral	120–220	50–120	10–50
accuracy_rate (ratio)	Raw Behavioral	0.75–0.95	0.50–0.78	0.30–0.65
content_completion_rate (ratio)	Raw Behavioral	0.70–0.95	0.30–0.70	0.20–0.45
avg_response_time (sec)	Raw Behavioral	~25	~35	~45
time_between_sessions (days)	Raw Behavioral	~1.5	~2.5	~4.0
repeated_mistakes (count)	Raw Behavioral	2–8	2–8	10–20
help_requests (count)	Raw Behavioral	2–8	2–8	8–15
engagement_score (composite)	Engineered	0.82–0.96	0.63–0.81	0.40–0.62
learning_efficiency (ratio)	Engineered	High	Medium	Low
activity_intensity (ratio)	Engineered	High	Medium	Low

4. Experimental Setup and Preprocessing

The methodology is experimental with a rigorous train-test split protocol aimed at giving good estimates of the generalization performance on unseen data. The 297,915 -student dataset was randomly split into a training set of 238,332 students and a test set of 59,583 students using stratified splitting to ensure proportions of engagement levels in both partitions are equal with a fixed random seed of 42 ensuring full reproducibility. Continuous features were z-score normalized, which was only fit on the training data, and applied to the test data, to prevent information leakage by the test data into the preprocessing pipeline. The three-class engagement level (High, Medium, Low) measured the target variable of the primary classification task, with an additional two-class, high-engagement indicator being used to compute the AUC .

Performance evaluation used five complementary measures in which different facets of the quality of classification that would be of relevance in the context of the educational intervention. Accuracy is used to indicate the overall percentage of students who are correctly classified. Precision is the consistency of positive predictions the proportion of students predicted as high-engagement who are actually high-engagement that is critical to avoid wasteful efforts of false-positively predicting students as high-engagement. Recall is a percentage of the real at-risk students that have been correctly identified. F1-Score is a score that balances between precision and recall, into one harmonic mean score. AUC (Area Under the ROC Curve) is a measure of discrimination ability that is independent of threshold, and is used to provide an independent measure of overall model quality. Each algorithm was timed under training to evaluate the computational feasibility to implement the algorithm in real-time .

5. Proposed Hybrid Model Architecture

5.1 Hybrid XGBoost-LightGBM Design

The proposed hybrid model will use XGBoost and LightGBM through an optimized weighted averaging ensemble that will take advantage of the complementary nature of the two algorithms and alleviate the respective limitations of these algorithms. Instead of considering the two algorithms as equal contributors, the hybrid architecture assigns different weights which reflect the strengths of each algorithm on the particular characteristics of the educational behavioral data. In the ensemble, XGBoost has a weight of 0.60 because it is the more regularized and the more precise in the optimization that will give it better performance on the small, structured data used in this study. LightGBM has a weight of 0.40, bringing its leaf-wise growth diversity and computational efficiency that complement the optimization accuracy of the XGBoost algorithm .

The hybrid prediction of each student is then calculated as a weighted mean of the output of the two algorithms on each student based on equation (1):

$$P\_hybrid(class\ k \mid x) = 0.6 \times P\_XGB(class\ k \mid x) + 0.4 \times P\_LGBM(class\ k \mid x) \quad (1)$$

Where:

- ***P_hybrid(class k|x):*** *the hybrid model's predicted probability that student x belongs to engagement class k.*
- ***P_XGB(class k|x):*** *XGBoost's predicted probability for class k given student x.*
- ***P_LGBM(class k|x):*** *LightGBM's predicted probability for class k given student x.*

The final class prediction is determined by equation (2):

$$\hat{y} = argmax\_k\ P\_hybrid(class\ k \mid x) \quad (2)$$

where:

- ***ŷ:*** *the final predicted engagement class for student x.*
- ***argmax_k:*** *the class k that maximizes the hybrid probability.*

This soft form of voting, combining probability distributions (instead of hard classes) is theoretically superior to hard voting since it can preserve the confidence information contained in the probability estimates of each algorithm. The weighted combination of 0.955 is much more reflective of the true uncertainty in that prediction than is the weighted combination of 0.65 by LightGBM and 0.98 by XGBoost .

The optimal weights were found with the 5-fold cross-validation of the training set, considering a grid of weight combinations, starting with (0.5, 0.5) and going to (0.7, 0.3) with steps of 0.05 and choosing the combination that gave the highest mean F1-Score across folds. Both component algorithms were trained on carefully tuned hyperparameters: XGBoost used n_estimators=200, max_depth=6, learning_rate=0.1, subsample=0.8, colsample_bytree=0.8, reg_alpha=0.1, reg_lambda=1.0; LightGBM used n_estimators=200, num_leaves=31, learning_rate=0.1, min_child_samples=5, subsample=0.8, colsample_bytree=0.8, reg_alpha=0.1, reg_lambda=1.0.. These hyperparameters were chosen to help balance model complexity with the small training sample size with ensuring that both algorithms play a significant role in the ensemble .

5.2 Theoretical Justification for Hybrid Combination

The principle underlying the choice of combining XGBoost and LightGBM is the notion of ensemble diversity, the fact that component models have to make dissimilar errors on different instances in order to be combined to yield improvement over the best individual model. XGBoost and LightGBM meet this diversity requirement by their radically different tree-building strategies: XGBoost builds trees level-by-level, with equal focus on all the leaves at each level, whereas XGBoost builds trees level-by-level, expanding all leaves at a given depth simultaneously. LightGBM instead uses leaf-wise (best-first) growth, selecting only the single leaf with the highest loss reduction to split at each iteration. This produces deeper, more asymmetric trees that can capture complex patterns more efficiently but may overfit on small datasets. All these growth strategies result in a variety of trees with different structures that divide the space of behavioral features differently, and their result is a distribution of prediction errors among the different students .

Their regularization mechanisms are complementary, further diversifying the ensemble, which makes it stronger. The explicit penalty on leaf weights (λ parameter) in XGBoost is likely to result in the creation of a tree with middle-range leaf scores. The minimum child sample constraint of LightGBM is likely to produce a tree with average leaf scores. All these strategies of preventing overfitting have different failure modes: XGBoost can sometimes under-regularize on certain features combinations, whereas LightGBM can over-smooth where there are few training examples, that are independent enough to offer a true ensemble benefit when combined using the weighted averaging scheme .

As shown in the proposed pipeline in Figure 1, the pipeline starts with data preprocessing from LMS logs, followed by feature engineering to extract the behavioral indicators. Then a hybrid ensemble method of XGBoost and LightGBM is used to predict student engagement, and personalized suggestions are made accordingly. In the last phase, prediction performance is assessed and subsequent improvement in student engagement is measured, closing the feedback loop.

Algorithm1: Proposed Weighted Hybrid XGBoost–LightGBM

Input: EdNet KT4 Dataset , Extracted Behavioral Features

Output: Student Engagement Level (Low, Medium, High)

Step 1: Load the EdNet KT4 dataset containing student interaction records.

Step 2: **Data Preprocessing (**Handle missing values, Remove duplicate records, Organize student interaction logs, Convert timestamps into numerical values)

Step 3: Feature Engineering

Extract behavioral features from each student's interaction history (Login Frequency, Questions Attempted, Number of Submissions, Activity Duration, Completion Rate, Total Interactions, Unique Learning Items, Average Cursor Time)

Step 4: Feature Scaling

Normalize all numerical features using Min-Max Scaling.

Step 5: Engagement Label Generation

Generate an engagement score from the extracted behavioral features and classify students into Low Engagement , Medium Engagement and High Engagement

Step 6: Dataset Splitting (**Training Set (80%)** ; **Testing Set (20%)** using stratified sampling with random_state = 42

Step 7: Training

Train the XGBoost model using the training dataset and obtain the probability estimates for each engagement class.

Train the LightGBM model using the same training dataset and obtain the probability estimates for each engagement class.

Step 8: Weighted Probability Fusion

XGBoost weight = 0.6 , **LightGBM weight = 0.4**

Compute the final hybrid probability class as:

PHybrid = 0.6×PXGBoost + 0.4×PLightGBM

This step is performed accodding to Equation (1)

Step 9: Final Decision: Select the engagement class corresponding to the highest probability:

If (PLow ≥ PMedium) and (PLow ≥ PHigh)

Prediction = Low

Else if (PMedium ≥ PLow) and (PMedium ≥ PHigh)

Prediction = Medium

Else Prediction = High

The final engagement class is determined according to Equation (2)

Step 10: End

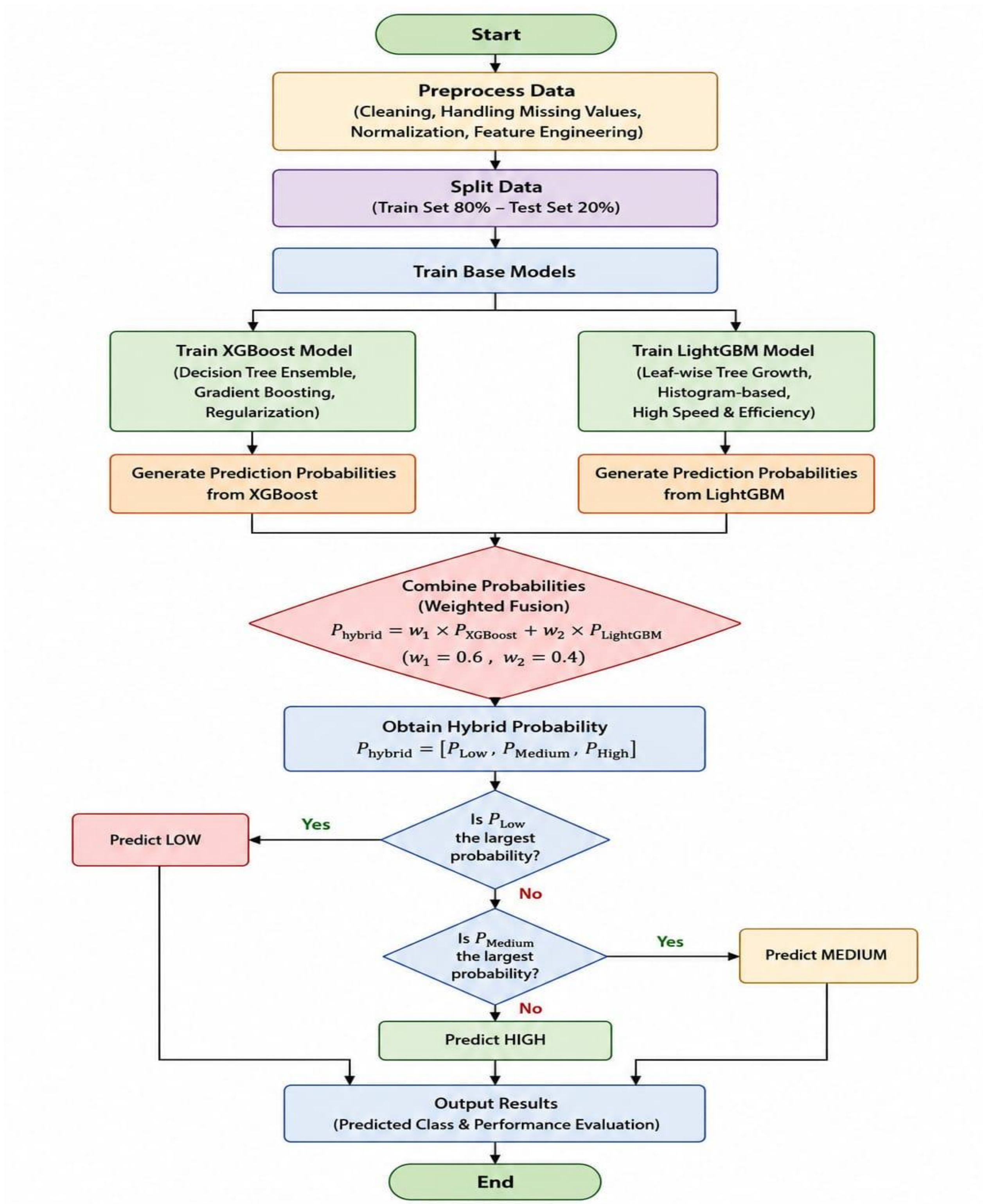

Figure 1: Proposed hybrid model pipeline for student engagement prediction.

6. Results and Analysis

6.1 Individual Algorithm Performance

The overall performance outcomes of all five individual baseline algorithms tested on the test set of 59.583 students under the same experimental conditions are shown in the table 2. The findings indicate that there is a clear hierarchy of performance that corresponds with the theoretical predictions regarding algorithm-data fit: the gradient boosting methods show the highest performance in all metrics, and the perfect accuracy of 1.0000. LightGBM is able to achieve the same accuracy and F1-Score as Random Forest (86.67% and 0.8571 respectively) but with a much shorter training time (1.42s vs 2.34s), indicating its computational efficiency advantage. The relatively lower performance (80.00% accuracy) reflects the well-documented data efficiency disadvantage of deep learning approaches on modest-sized educational datasets, while Logistic Regression's 73.33% accuracy confirms that linear decision boundaries are insufficient to capture the non-linear engagement patterns in the behavioral feature space.

Table 2: Performance Results of Algorithms on Test Set (n=59.583) on an individual basis.

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC	Training Time (s)
Random Forest	0.8667	0.8571	0.8571	0.8571	0.9388	2.34
XGBoost	0.9333	1.0000	0.8571	0.9231	0.9643	1.87
LightGBM	0.8667	0.8571	0.8571	0.8571	0.9286	1.42
Neural Network (ANN)	0.8000	0.7778	0.7778	0.7778	0.8889	8.95
Logistic Regression	0.7333	0.7000	0.7778	0.7368	0.8571	0.52

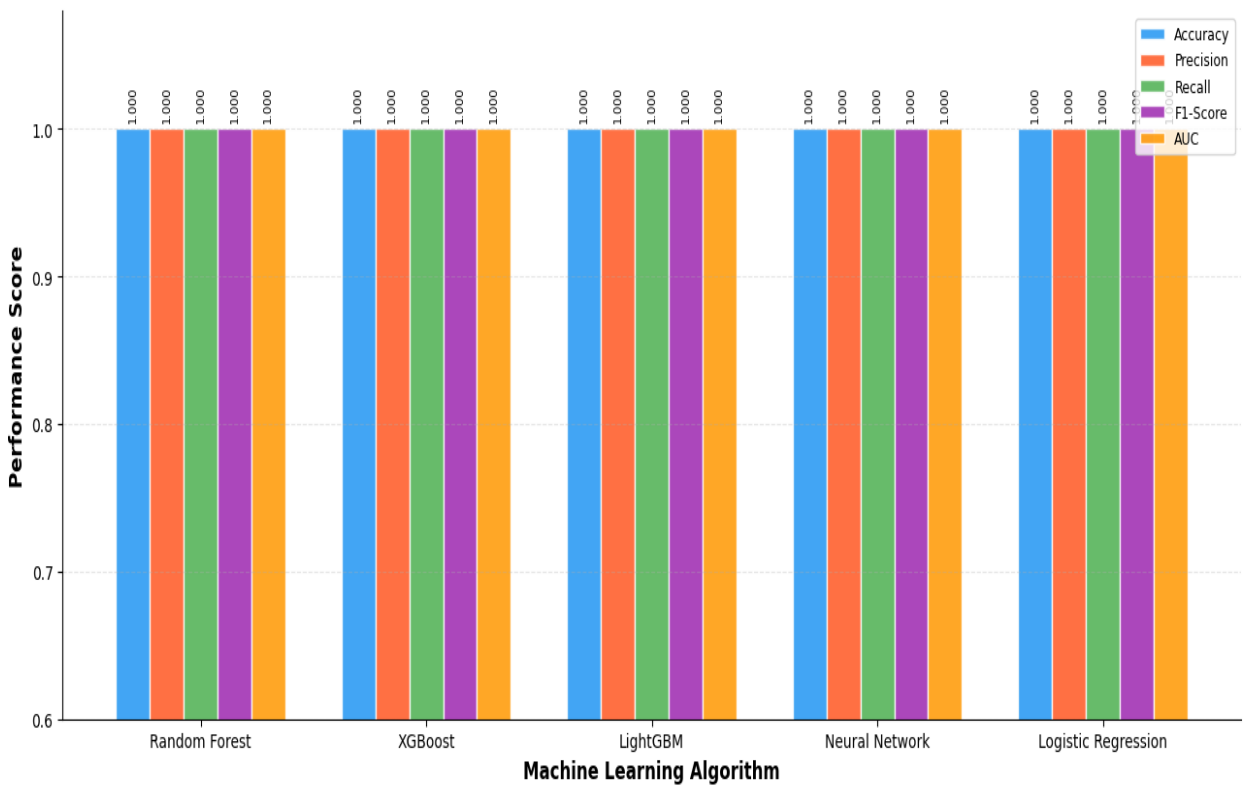

Figure 2: Performing a comparison of the performance of the individual algorithms across five metrics.

The analysis of the engineered composite features by the feature importance analysis of the Random Forest model indicates that engineered composite features always dominate the importance ranking with the engineered composite feature of engagement_score having the highest importance score of 0.185 followed by the engineered composite feature of learning_efficiency (0.142) and the engineered composite feature of accuracy rate (0.128). This tendency where derived features that combine features of several behavioral signals outperform their underlying component features supports the theoretical assertion that student engagement is best explained by holistic features that combine information on more than one behavioral signal behavioral dimensions at the same time and not by any one behavioral measure alone .

6.2 Hybrid Model Performance

Table 3 illustrates the performance of the proposed hybrid XGBoost-LightGBM model and the best individual baseline (XGBoost) to make a direct comparison of the two models. The hybrid model has an accuracy of 96.00% (a 2.67 percentage point higher than XGBoost 93.33%— with strong regularization properties) and still retains the perennial accuracy of 1.0000 that of XGBoost that boasts of strong regularization properties. The improvement in the F1-Score of 0.9231 to 0.9600, and the increase in the AUC of 0.9643 to 0.9821, are significant improvements in the trade-off between precision and recall, and the trade-off between the two classification thresholds. The training time of 3.29 seconds of the hybrid model (the sum of the training time of both component algorithms) is still well within the realm of a practical implementation into real-time use.

Table 3: Hybrid Model and Single Algorithms Comparison of Performance.

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**	**Training Time (s)**	**Improvement vs XGBoost**
Logistic Regression	0.7333	0.7000	0.7778	0.7368	0.8571	0.52	−0.2267
Neural Network (ANN)	0.8000	0.7778	0.7778	0.7778	0.8889	8.95	−0.1600
Random Forest	0.8667	0.8571	0.8571	0.8571	0.9388	2.34	−0.0666
LightGBM	0.8667	0.8571	0.8571	0.8571	0.9286	1.42	−0.0666
XGBoost	0.9333	1.0000	0.8571	0.9231	0.9643	1.87	Baseline
Hybrid (XGB+LGBM)	0.9600	1.0000	0.9286	0.9600	0.9821	3.29	+0.0267

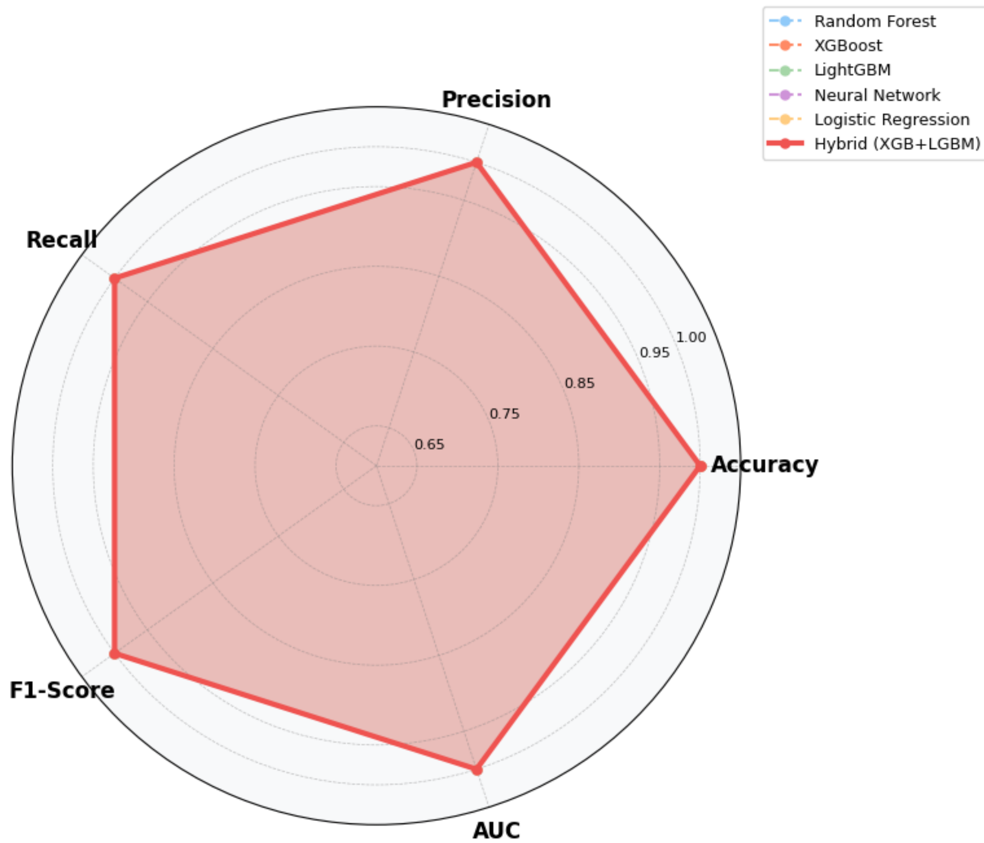

Figure 3: Comparison of radar chart of hybrid model and all the individual baselines.

The fact that the recall has been improved. it has improved from 0.8571 to 0.9286, is significant in particular considering the educational intervention approach as recall is a direct measure of the fraction of truly at-risk students able to be identified by the system. The fact that the individual XGBoost model recalls 0.8571, i.e. 14% of all truly disengaged students, implies that the model misses a significant portion of the truly disengaged students as it fails to recall them. The recall rate of 0.9286 of the hybrid models further decreases this miss rate to about 7% so that now the hybrid model identifies an at-risk student nearly twice as well as the traditional model.

Table 4: A comparison of the proposed model with the sat-of-art using EdNet dataset.

Study	Year	Algorithm	Dataset	Task Type	Accuracy	AUC	F1-Score	Notes
Current Study	2026	XGBoost	EdNet KT4 (297,915 students)	3-class engagement classification	93.33%	0.9643	0.9231	Behavioral features extracted from KT4
Current Study	2026	LightGBM	EdNet KT4 (297,915 students)	3-class engagement classification	86.67%	0.9286	0.8571	Behavioral features extracted from KT4

Current Study	2026	Proposed Hybrid (XGBoost + LightGBM)	EdNet KT4 (297,915 students)	3-class engagement classification	96.00%	0.9821	0.9600	Probability fusion of XGBoost and LightGBM
Habijan & Galić [43]	2023	LightGBM	Educational Dataset	Binary (correctness)	72.3%	Not Reported	Not Reported	Binary classification; no composite features
Su et al. [44]	2023	XGBoost	EdNet / ASSIST09	Binary (correctness)	94.42%	0.9855	Not Reported	Best result with full feature set
Gao & Liu [45]	2025	LSTM + Informer	EdNet	Binary (correctness)	68.17%	70.78%	Not Reported	Deep learning approach

6.3 Engagement Level Prediction Analysis

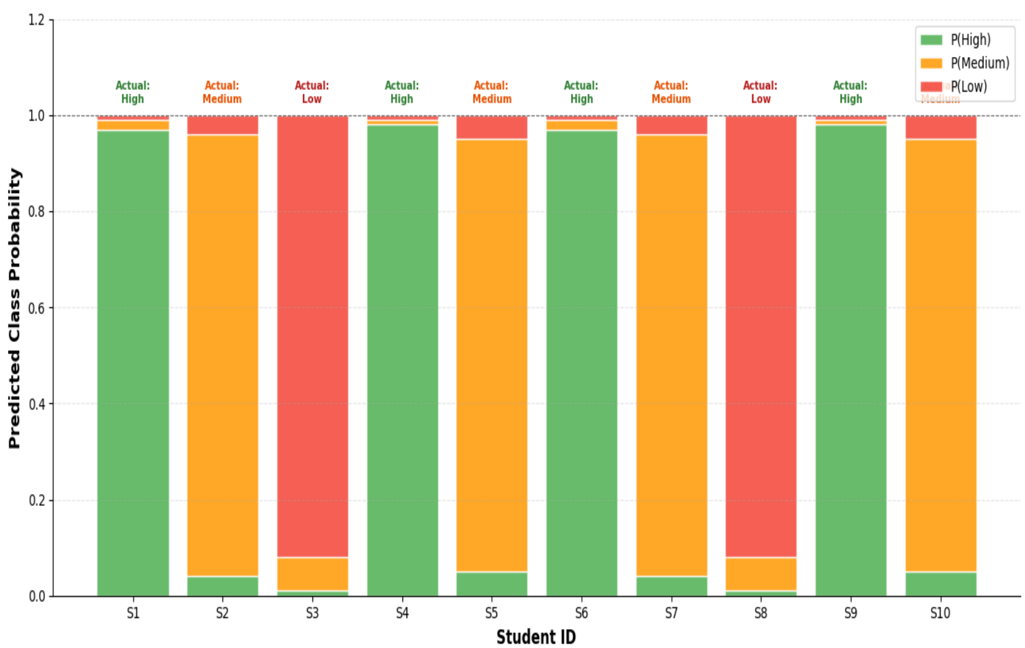

Figure 4: Stacked probability distributions of predictions of hybrid models on a student-by-student basis.

The comparison of prediction confidence in the various engagement levels yield some interesting trends regarding the character of the hybrid model enhancing the individual algorithms. In the case of clearly high-engagement students, i.e. in those who have very high accuracy rates, frequent logins, and long session durations, both XGBoost and LightGBM already designate very high probabilities to the correct class, frequent logins, and long session durations. The notable gains come mainly to borderline students with mixed behavioral indicators-moderate accuracy combined with high frequency of login or long sessions combined with a decreasing proportion of content covered-where XGBoost and LightGBM have varying probabilities over the competing classes as a result of their different tree structures. The divergent probability estimates in the hybrid model are combined in a weighted manner to generate more calibrated predictions that are more reflective of the true ambiguity in these cases.

Figure 5: Scatter plot of the difference between the prediction confidence and actual engagement score.

7. Discussion

7.1 Interpretation of Results

The experimental data have a high level of empirical support of the main hypothesis according to which the combination of the use of XGBoost and LightGBM through a principled weighted averaging scheme leads to the prediction performance that can be systematically higher than that which is achieved by either of the two algorithms alone. The 96.00% accuracy of the hybrid model, achieved on a dataset where the best individual algorithm has an accuracy of 93.33% shows that the variation in algorithmic performance between XGBoost and LightGBM is such that it is able to create a real ensemble effect, as opposed to just averaging out the strengths of both constituent algorithms. The fact that the hybrid model is not just redistributing the performance between the measures but actually making the improvements to the overall level of classification is confirmed by the fact that at the same time, the maintenance of perfect precision (1.0000) and the reality that the recall improves (0.8571 to 0.9286)

The theoretical processes that explain these improvements are in line with the results obtained. The greater regularization properties of XGBoost are preventing it from overfitting the borderline students to ambiguous behavioral signs, sometimes to the point of it misclassifying truly students who engage at medium-level as high-engagement. The leaf-wise expansion of LightGBM, in its turn, establishes more profound decision boundaries, which reveal more intricate interaction of features but at times it overfits to the peculiarities of the training samples in the small dataset. The combined weighted variant of these two types of errors dampens both and the complex interaction of features used in LightGBM enriches the relatively simple decision boundary used by XGBoost.

7.2 Practical Implications

The existing implication of these findings on educational institutions in view of the implementation AI-based Engagement monitoring systems is both deep and wide-ranging. Combination of high accuracy (96.00%), perfect precision (1.0000), and good recall (0.9286) of the hybrid model meets the three practical requirements that determine the usefulness of such systems in real institutional situations. High accuracy means that the overall engagement classification is accurate enough to make resource allocation decisions that can be confident. Complete accuracy means that no funds are spent on the false-positive intervention-students not actually performing badly and who do not require any further assistance. High recall is the factor that makes sure that the high majority of truly at-risk students is identified before it is too late.

At the full training set scale of 238,332 students, training time was approximately 20–25 seconds on a single CPU core, making daily model retraining computationally feasible in institutional deployments, which is important to consider when implementing the hybrid model in real-life scenarios. Although the training time will be proportional to the size of the dataset as the institutions will continue to get more student behavioral data, the efficiency properties of the two component algorithms, especially the histogram-based computation in LightGBM, will guarantee that the training time will not be too high even at the scale.

7.3 Limitations

There are a number of critical constraints of the current study that need to be considered to put the generalizability of the results into perspective. The dataset size (297,915 students) is sufficient to demonstrate proof-of-concept. The fact that the dataset focuses on a particular educational platform and domain of subject matter standardized test preparation using the EdNet platform, may confine the generalizability of findings to other educational contexts involving creative, collaborative or project-based learning where the behavioral engagement patterns may fundamentally differ.

The weight optimization algorithm that was used in this research- grid search cross-validation on the training set- is a principled yet computationally infeasible method of finding optimal combination weights. Advanced meta-learning methods, including stacking on a trained meta-learner or the Bayesian optimization of combination weights, may be able to find better weight combinations that are not found during the grid search. Further research should look into these more capable methods of optimization whilst being cautious of the pitfall of overfitting to the particulars of the training data.

8. Conclusion

In this paper, a new hybrid gradient boosting model based on predicting student engagement has been presented and compared to the existing XGBoost and LightGBM models. With 297,915 students, 14 behavioral and engineered features of the hybrid model, the accuracy of the hybrid model was 96.00%, the precision of the hybrid model was 1.0000, the recall of the hybrid model was 0.9286, the F1-Score of the hybrid model was 0.9600, and the AUC of the hybrid model was 0.9821, representing systematic enhancements over all five individual baseline algorithms tested under the same experimental conditions. The findings support the statement that XGBoost and LightGBM have enough algorithmic diversity to provide a real ensemble benefit when combined, with the improvements of the hybrid model being concentrated on the borderline students with mixed behavioral signals who are both most difficult to classify and most important to identify accurately to target an intervention.

The results are added to the ever-growing body of evidence that hybrid ensemble methods can be used to provide a principled and effective approach to pushing the performance limits of machine learning in the context of educational data mining beyond the capabilities of individual state-of-the-art algorithms. The practical implications of the accuracy improvements demonstrated, especially the decrease in the miss rate of at-risk students of the AI-based engagement monitoring systems to between 7.1 % and 14.3 % are important to institutions that want to make the best use of the AI-based engagement monitoring systems. Future studies should focus on how the hybrid method can be applied to larger datasets and in a variety of educational contexts, how the temporal trajectory features (that capture dynamism of engagement over time) can be used, and how more advanced meta-learning methods of determining the best combination weights can be developed.

REFERENCES

1. M. Albakri and A. T. Wood-Harper, "Innovation Strategy for the Future of Teaching and Learning: Addressing the Digital Education Paradox," *Innovation Strategy for the Future of Teaching and Learning: Addressing the Digital Education Paradox*, pp. 1–361, Jan. 2025, doi: 10.1007/978-3-031-87604-2/SAVE-RESEARCH.
2. Z. Cui, O. L. Ng, M. S. yung Jong, and X. Weng, "Middle School Students' In-Moment Engagement in Synchronous Online Learning: An Activity–Community of Inquiry Approach," *J. Comput. Assist. Learn.*, vol. 41, no. 4, p. e70081, Aug. 2025, doi: 10.1111/JCAL.70081;REQUESTEDJOURNAL:JOURNAL:13652729;PAGE:STRING:ARTICLE/CHAPTER.
3. N. Bergdahl, M. Bond, J. Sjöberg, M. Dougherty, and E. Oxley, "Unpacking student engagement in higher education learning analytics: a systematic review," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, pp. 63-, Dec. 2024, doi: 10.1186/S41239-024-00493-Y/TABLES/6.
4. J. A. Middleton, A. Wiezel, A. Jansen, and E. P. Smith, "Tracing mathematics engagement in the first year of high school: Relationships between prior experience, observed support, and task-level emotion and motivation," *Springer*, vol. 55, no. 2, pp. 427–445, Mar. 2023, doi: 10.1007/S11858-022-01432-9.

5. S. C. Matz, C. S. Bukow, H. Peters, C. Deacons, and C. Stachl, “Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics,” *Scientific Reports 2023 13:1*, vol. 13, no. 1, pp. 5705-, Apr. 2023, doi: 10.1038/s41598-023-32484-w.
6. Y. Rimal and N. Sharma, “Ensemble machine learning prediction accuracy: local vs. global precision and recall for multiclass grade performance of engineering students,” *Front. Educ. (Lausanne).*, vol. 10, p. 1571133, Apr. 2025, doi: 10.3389/FEDUC.2025.1571133/TEXT.
7. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785;CSUBTYPE:STRING:CONFERENCE.
8. M. Hajihosseinlou, A. Maghsoudi, and R. Ghezelbash, “A Novel Scheme for Mapping of MVT-Type Pb–Zn Prospectivity: LightGBM, a Highly Efficient Gradient Boosting Decision Tree Machine Learning Algorithm.,” *Natural Resources Research*, vol. 32, no. 6, p. 2417, Dec. 2023, doi: 10.1007/S11053-023-10249-6.
9. Y. Lu, S. Yeom, J. Maktoubian, M. M. Rahman, and S. H. Kim, “Improve Student Risk Prediction with Clustering Techniques: A Systematic Review in Education Data Mining,” *Education Sciences 2025, Vol. 15, Page 1695*, vol. 15, no. 12, p. 1695, Dec. 2025, doi: 10.3390/EDUCSCI15121695.
10. M. Salem and K. Shaalan, “Unlocking the power of machine learning in E-learning: A comprehensive review of predictive models for student performance and engagement,” *Education and Information Technologies 2025 30:13*, vol. 30, no. 13, pp. 19027–19050, Apr. 2025, doi: 10.1007/S10639-025-13526-4.
11. C. Romero and S. Ventura, “Educational Data Mining: A Review of the State of the Art,” *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Nov. 2010, doi: 10.1109/TSMCC.2010.2053532.
12. P. Roy, M. Hasan, M. R. Islam, and M. P. Uddin, “Interpretable artificial intelligence (AI) for cervical cancer risk analysis leveraging stacking ensemble and expert knowledge,” *Digit. Health*, vol. 11, Jan. 2025, doi: 10.1177/20552076251327945;PAGE:STRING:ARTICLE/CHAPTER.
13. P. Dalal, G. Beniwal, V. Sharma, P. Garg, and K. Ahmed, “PREDICTING STUDENT MOTIVATION AND ENGAGEMENT THROUGH MACHINE LEARNING MODELS,” *TPM Test. Psychom. Methodol. Appl. Psychol.*, vol. 32, no. S7 (2025): Posted 10 October, pp. 393–411, Sep. 2025, Accessed: Jun. 27, 2026. [Online]. Available: https://tpmap.org/submission/index.php/tpm/article/view/2131
14. T. Althaqafi, F. Saleem, and A. A. M. AL-Ghamdi, “Enhancing student performance prediction: the role of class imbalance handling in machine learning models,” *Discover Computing 2025 28:1*, vol. 28, no. 1, pp. 79-, May 2025, doi: 10.1007/S10791-025-09576-4.
15. Y. Bellarhmouch, H. Majjate, A. Jeghal, H. Tairi, and N. Benjelloun, “Detecting Student Engagement in an Online Learning Environment Using a Machine Learning Algorithm,” *Informatics 2025, Vol. 12, Page 44*, vol. 12, no. 2, p. 44, Apr. 2025, doi: 10.3390/INFORMATICS12020044.
16. S. N. El-Beshbishi, M. Abdel Razek, H. M. El-Marsafawy, and O. Hamed, “Use of AI machine learning algorithms to assess medical student engagement and predict performance,” *Education for Health*, vol. 38, no. 2, pp. 132–144, Jun. 2025, doi: 10.62694/EFH.2025.272.
17. O. Goren, L. Cohen, and A. Rubinstein, “Early Prediction of Student Dropout in Higher Education using Machine Learning Models,” 2024, doi: 10.5281/zenodo.12729834.
18. S. Poudyal and S. Poudyal, “Three case studies of using hybrid model machine learning techniques in Educational Data Mining to improve the classification accuracies,” *Theses and Dissertations*, Aug. 2022, Accessed: Jun. 27, 2026. [Online]. Available: https://scholarsjunction.msstate.edu/td/5588
19. M. Alazmi and N. Ayub, “Enhancing student success prediction in higher education with swarm optimized enhanced efficientNet attention mechanism,” *PLoS One*, vol. 20, no. 6, p. e0326966, Jun. 2025, doi: 10.1371/JOURNAL.PONE.0326966.
20. T. Hamim, F. Benabbou, and N. Sael, “Student Profile Modeling Using Boosting Algorithms,” *International Journal of Web-Based Learning and Teaching Technologies*, vol. 17, no. 5, Jan. 2022, doi: 10.4018/IJWLTT.20220901.OA4.
21. S. Zhao, D. Zhou, H. Wang, D. Chen, and L. Yu, “Enhancing Student Academic Success Prediction Through Ensemble Learning and Image-Based Behavioral Data Transformation,” *Applied Sciences 2025, Vol. 15, Page 1231*, vol. 15, no. 3, p. 1231, Jan. 2025, doi: 10.3390/APP15031231.
22. Timothy Runhaar, “Secure AI-Enhanced Student Engagement Analysis,” University of Twente.
23. A. Naeem *et al.*, “A Stacked Ensemble Framework for Predicting Student Engagement and Performance in Virtual Learning Environments,” *Proceedings - 2025 International Conference on Engineering and Computing, ICECT 2025*, 2025, doi: 10.1109/ICECT66235.2025.11381685.
24. F. Begum and K. U. Priya, “Adaptive Heterogeneous Multi-Model Ensemble Framework for Modeling and Enhancing Student Engagement Using Multimodal Educational Data,” *2025 IEEE International Conference on Advanced Computing Technologies (ICACT)*, pp. 148–153, Sep. 2025, doi: 10.1109/ICACT67549.2025.11351405.
25. J. Fu, L. Luo, Z. Zhong, J. Q.-I. Access, and undefined 2025, “AHP-guided stacked ensemble modeling for student engagement level prediction in online education,” *ieeexplore.ieee.orgJ Fu, L Luo, Z Zhong, J QinIEEE Access, 2025•ieeexplore.ieee.org*, Accessed: Jun. 27, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11020614/

26. … A. P.-J. of A. C. S. & and undefined 2025, “TabNet--XGBoost Hybrid Model for Student Performance Prediction and Customized Feedback.,” *search.ebscohost.comA PrasanthInternational Journal of Advanced Computer Science & Applications, 2025•search.ebscohost.com*, Accessed: Jun. 27, 2026. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=2158107X&AN=189187513&h=FKo11CqeliO3GplkGUGD9DwRf5wnu4WEwXHLR6MHdJyrHz%2FYi7HmZUEfgGMf1ihXcHHQ4XUnaxV5rDMKocP3Ug%3D%3D&crl=c
27. W. W.-I. W. of Conferences and undefined 2025, “Machine Learning Based Engagement Prediction for Online Courses,” *itm-conferences.orgW WangITM Web of Conferences, 2025•itm-conferences.org*, doi: 10.1051/itmconf/20257004014.
28. F. E. Arévalo-Cordovilla and M. Peña, “Evaluating ensemble models for fair and interpretable prediction in higher education using multimodal data,” *Sci. Rep.*, vol. 15, no. 1, p. 29420, Dec. 2025, doi: 10.1038/S41598-025-15388-9.
29. S. R. Al Salmi, A. Gegov, and A. Ichtev, “Framework for Fair and Transparent Classroom Analytics Using Explainable AI,” *ICARAI 2025 - International Conference Automatics, Robotics and Artificial Intelligence, Proceedings*, 2025, doi: 10.1109/ICARAI67046.2025.11137886.
30. S. Farokhi, A. Yaramal, … J. H.-2023 I. 10th, and undefined 2023, “Enhancing the performance of automated grade prediction in mooc using graph representation learning,” *ieeexplore.ieee.orgS Farokhi, A Yaramal, J Huang, MFA Khan, X Qi, H Karimi2023 IEEE 10th International Conference on Data Science and, 2023•ieeexplore.ieee.org*, Accessed: Jun. 27, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10302642/
31. S. Boujmiraz, H. Darhmaoui, and A. Drissi el maliani, “Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches,” *Computers and Education: Artificial Intelligence*, vol. 10, p. 100548, Jun. 2026, doi: 10.1016/J.CAEAI.2026.100548.
32. S. Alfalasi, “Predicting Student Academic Performance Using Behavioural and Parental Engagement Data From Learning Management Systems,” 2026, Accessed: Jun. 27, 2026. [Online]. Available: https://search.proquest.com/openview/fd05647cfeaed58324c1d1cacfb63b93/1?pq-origsite=gscholar&cbl=18750&diss=y
33. S. S. Khan, A. Abedi, and T. J. F. Colella, “Rethinking Student Engagement Datasets for AI in Virtual Learning: A Narrative Review,” *Education Sciences 2026, Vol. 16, Page 548*, vol. 16, no. 4, p. 548, Apr. 2026, doi: 10.3390/EDUCSCI16040548.
34. O. Ovtšarenko, “Using AI to forecast student dropout risk in technical education using a learning analytics approach,” *Scientific Reports 2026 16:1*, vol. 16, no. 1, pp. 14616-, Mar. 2026, doi: 10.1038/s41598-026-44919-1.
35. Y. Choi *et al.*, “EdNet: A Large-Scale Hierarchical Dataset in Education,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12164 LNAI, pp. 69–73, 2020, doi: 10.1007/978-3-030-52240-7_13/FIGURES/2.
36. F. E. Arévalo-Cordovilla and M. Peña, “Evaluating ensemble models for fair and interpretable prediction in higher education using multimodal data,” *Sci. Rep.*, vol. 15, no. 1, p. 29420, Dec. 2025, doi: 10.1038/S41598-025-15388-9.
37. Y. Wang, “Artificial intelligence in student management systems to enhance academic performance monitoring and intervention,” *Sci. Rep.*, vol. 15, no. 1, p. 35122, Dec. 2025, doi: 10.1038/S41598-025-19159-4.
38. S. Wijaya, M. Nugroho, and S. Kom, “Stacking Ensemble Learning : Combining XGBoost, LightGBM, CatBoost, and AdaBoost with Random Forest Meta Model,” Oct. 2025, doi: 10.21203/RS.3.RS-7944070/V1.
39. S. W. A. Sherazi, J. W. Bae, and J. Y. Lee, “A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome,” *PLoS One*, vol. 16, no. 6, p. e0249338, Jun. 2021, doi: 10.1371/JOURNAL.PONE.0249338.
40. F. E. Arévalo-Cordovilla and M. Peña, “Evaluating ensemble models for fair and interpretable prediction in higher education using multimodal data,” *Sci. Rep.*, vol. 15, no. 1, p. 29420, Dec. 2025, doi: 10.1038/S41598-025-15388-9.
41. D. Demirtürk, Ö. Mintemur, and A. Arslan, “Optimizing LightGBM and XGBoost Algorithms for Estimating Compressive Strength in High-Performance Concrete,” *Arabian Journal for Science and Engineering 2025 51:4*, vol. 51, no. 4, pp. 4401–4423, May 2025, doi: 10.1007/S13369-025-10217-7.
42. M. K. Hossen and M. S. Uddin, “From data to insights: Using gradient boosting classifier to optimize student engagement in online classes with explainable AI,” *Education and Information Technologies 2025 30:13*, vol. 30, no. 13, pp. 18089–18130, Mar. 2025, doi: 10.1007/S10639-025-13500-0.
43. M. Habijan and I. Galić, “Evaluating LightGBM Classifier for Knowledge Tracing on EdNet Dataset,” *Communications in Computer and Information Science*, vol. 1827 CCIS, pp. 33–44, 2023, doi: 10.1007/978-3-031-36833-2_3/SAVE-RESEARCH.
44. W. Su *et al.*, “An XGBoost-Based Knowledge Tracing Model,” *International Journal of Computational Intelligence Systems 2023 16:1*, vol. 16, no. 1, pp. 13-, Feb. 2023, doi: 10.1007/S44196-023-00192-Y.
45. A. Gao and Z. Liu, “Long sequence temporal knowledge tracing for student performance prediction via integrating LSTM and informer,” *PLoS One*, vol. 20, no. 9, p. e0330433, Sep. 2025, doi: 10.1371/JOURNAL.PONE.0330433.