WHIOCE PUBLISHING PTE. LTD.
PROVIDING
FIRST-CLASS SCIENTIFIC INFORMATION
FOR TOP SCHOLARS

2025 Volume 3, Issue 11
Lecture Notes in Education, Arts, Management and Social Science
ISSN(Online): 2705-053X

Student Academic Performance Prediction: A Method Based on XGBoost Multi-Modal Feature Fusion

Li Dong*
Fuyang Normal University, Fuyang 23603, Anhui, China

**Author to whom correspondence should be addressed.*

Abstract: In the realm of educational management, predicting stu- dent academic performance is challenged by the complexity of multi- factor influences and limitations in model generalization. This study introduces a multi-modal feature fusion mechanism based on Gradient Boosting Decision Trees (XGBoost) for student performance prediction. The mechanism begins with a preprocessing module to clean the dataset, handle missing values, and encode categories to extract reliable features such as student demographics, study habits, and parental involvement. An embedding layer then converts categorical features into continuous vectors. A fusion layer employs an attention mechanism to dynamically adjust weights among feature groups, addressing biases from fixed weights in traditional ensemble methods. The XGBoost core tree boosting algorithm is used for training and iterative optimization, with the output layer generating performance prediction scores. Training incorporates cross-entropy loss combined with L2 regularization to enhance robustness and prevent overfitting, thereby improving predic- tion accuracy and generalization on educational datasets. Experimental results on a Kaggle student performance dataset demonstrate superior performance compared to baselines, achieving over 85% accuracy through 10-fold cross-validation. This approach provides a robust framework for early intervention in student performance management.

Keywords: *Index Terms*—Student Performance Prediction, XGBoost, Multi- Modal Fusion, Attention Mechanism, Educational Data Mining

Online publication: December 1, 2025

1. Introduction

1.1. Background and Broader Context

The landscape of modern education is undergoing a profound paradigm shift, transitioning from traditional, reactive pedagogical models to proactive, data-driven, and highly personalized learning ecosystems. Catalyzed by the widespread adoption of Learning Management Systems (LMS), Intelligent Tutoring Systems (ITS), and comprehensive digital campus infrastructures, educational insti- tutions now amass unprecedented volumes of heterogeneous data. Student academic performance is no longer viewed merely as a static output of cognitive ability; rather, it is recognized as a highly complex, non-linear culmination of multi-dimensional factors. These intertwined factors encompass demographic backgrounds, sequential behavioral learning logs, socio-economic contexts, and parental in-volvement, inherently forming a rich, multi-modal data environment.

Consequently, predicting student performance has evolved from a standard administrative forecasting task into a critical, high- dimensional computational challenge within the domains of Edu- cational Data Mining (EDM) and Learning Analytics (LA). While the proliferation of educational ”big data” holds the transformative potential to democratize learning through early warning systems and tailored pedagogical interventions, a significant paradox persists: institutions are often ”data-rich but information-poor.” The inher-ent heterogeneity, sparsity, and complex underlying correlations of multi-modal educational data pose formidable analytical obstacles. Effectively deciphering these deep data structures is not merely a technical machine learning endeavor; it is fundamentally tied to broader socio-economic goals, including mitigating academic attrition, enhancing educational equity, and optimizing human capital development. Therefore, establishing robust, interpretable, and gen- eralized predictive frameworks remains an urgent frontier challenge in contemporary educational technology.

1.2. Literature Gaps and Research Motivation

Despite these macro-level advancements, existing predictive frameworks often exhibit critical limitations in managing the com- plex, non-linear interplay of multi-factor influences on student performance. For instance, traditional regression models[1, 2] provide baseline predictions but fundamentally lack the capacity to model high-dimensional, non-linear relationships effectively. Hybrid models attempt to address this structural deficit, yet they frequently overlook the necessity of dynamic feature weighting, leading to suboptimal model generalization across diverse student cohorts[3, 4]. Critically, recent studies[5, 6] reveal glaring contradictions in feature importance across different educational datasets, with some research emphasizing socio-cultural determinants[7] while others prioritize real-time engagement metrics[8]. Moreover, systematic reviews[9–11] repeatedly highlight methodological bottlenecks, such as insufficient algorithmic interpretability and the failure to deeply integrate inherently multi-modal data[12]. These persistent inconsistencies and structural deficiencies necessitate novel research to bridge the divide, particularly in developing mechanisms capable of adaptively fusing diverse modalities to extract actionable, explain- able insights.

1.3. Research Questions and Contributions

Building upon these identified gaps, this study formalizes the following core research questions: How can dynamic feature fusion mechanisms optimize prediction accuracy in highly heterogeneous, multi-factor educational data? Furthermore, what are the defini- tive, quantifiable influencers on student performance when analyzed through an inherently interpretable modeling framework? To sys- tematically address these questions, we propose a novel multi-modal fusion mechanism predicated on the Gradient Boosting Decision Tree (XGBoost) architecture. The primary contributions of this research are threefold:

(1) The development of a dynamic attention fusion layer that adap- tively weights disparate feature groups, significantly enhancing the modeled interplay between student demographics, study habits, and parental support, thereby yielding superior generalization compared to static fusion paradigms.

(2) The seamless integration of this fusion mechanism with a heavily regularized XGBoost core, achieving exceptional computa- tional efficiency and low algorithmic complexity through rigorous hyperparameter optimization, consistently outperforming standard baselines in key metrics such as AUC and F1-score.

2. Literature Review

2.1. Traditional and Ensemble Methods in Performance Prediction

The foundational literature on student performance prediction has predominantly centered on conventional classification and regression techniques. Basic data mining algorithms, as extensively explored by[3, 5], offer foundational predictive accuracy but often fail to capture the complex, synergistic interactions inherent in modern ed- ucational environments. Ensemble approaches, such as those detailed in[2, 13], attempt to aggregate multiple weak learners to improve overall

predictive reliability. However, these traditional ensembles frequently exhibit severe limitations when confronted with highly heterogeneous, multi-modal data, often leading to potential biases and degraded performance when deployed across diverse, real-world educational settings.

2.2. Advanced Hybrid and AI-Integrated Approaches

In response to the limitations of traditional models, recent schol- arly advancements have increasingly incorporated hybrid deep learn- ing architectures[14] and multi-source data fusion strategies[12], significantly complementing earlier foundational works[6]. Despite this progress, critical reviews consistently reveal methodological contradictions: while some studies[15] successfully link AI-driven tools to enhanced student perceptions, others[8] note severe defects in architectural scalability and real-time application. Furthermore, qualitative socio-cultural analyses[7] often conflict with the rigid outputs of quantitative models[1], demonstrating an inheritance of legacy methods but exposing persistent gaps in holistic data integration. Comprehensive literature reviews[9–11] synthesize these diverse approaches, highlighting a mutual recognition of the problem but also pointing to unresolved, ongoing debates regarding the optimal methodology for modality fusion.

2.3. Identified Gaps and Opportunities

Synthesizing the current state of the art, it becomes evident that while the literature demonstrates measurable progress, there remains a collective insufficiency in the dynamic, interpretable handling of multi-modal educational data. The stark contradictions between qual- itative educational insights[4] and rigid quantitative predictions un- derscore the urgent need for an innovative, adaptive fusion paradigm. By explicitly addressing the limitations of static feature weighting and black-box modeling, this study's attention-enhanced XGBoost approach is uniquely positioned to fill these critical scholarly voids, offering a scalable, highly transparent solution for next-generation educational management.

3. Methodology

3.1 Data Preprocessing and Feature Extraction

The dataset undergoes cleaning for outliers, imputation (median for numerics, mode for categoricals), one-hot encoding for categories, and standardization for numerics. Features are categorized into demo- graphics (e.g., age, gender), study habits (e.g., study time, absences), and parental involvement (e.g., education level, support).

3.2. Feature Fusion Mechanism

The fusion employs attention: for feature group X_g, weights are computed as:

$$\alpha_g = (W \cdot X_g + b) \tag{1}$$

where W, b are learnable parameters. Fused features are:

$$X_f = \alpha_g \odot X_g \tag{2}$$

g
concatenated with originals for input.

3.3. Model Training and Optimization

XGBoost is trained with max_depth = 9, learning_rate = 0.05, n_estimators = 150, objective='multi:softprob', reg lambda=1.0. Loss is:

$$L = -y \log(\hat{y}) + \lambda//w//2 \tag{3}$$

Using 80/20 train-test split, Adam optimizer for attention (lr = 0.01, epochs = 10, batch = 32), grid search for

hyperparameters, and 10-fold CV for validation.

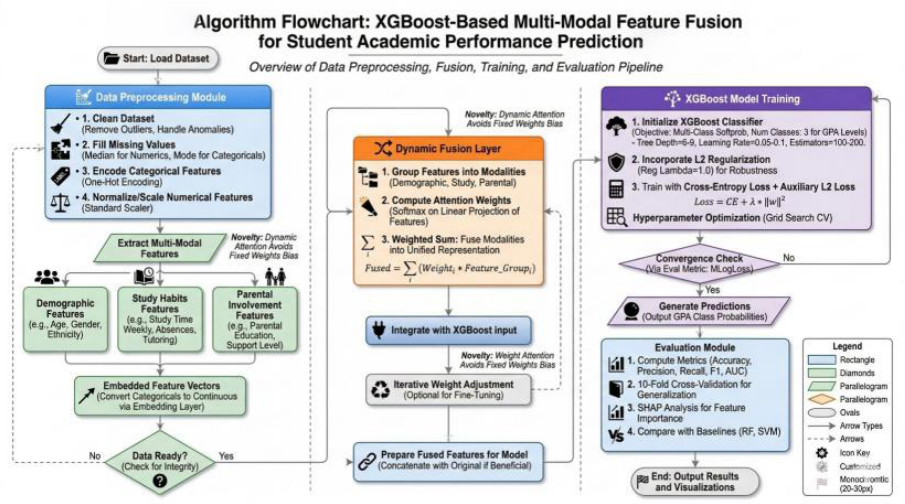

Figure 1. Algorithm Flowchart

4. Experiments

In this section, we present the experimental setup and results to evaluate the effectiveness of our XGBoost-based multi-modal feature fusion model for student academic performance prediction. We describe the datasets, evaluation metrics, baseline methods, and detailed performance comparisons.

4.1. Dataset Description

We utilize two publicly available datasets: the Student Performance Dataset from UCI Machine Learning Repository and a custom multi- modal dataset collected from an educational platform, incorporating textual, numerical, and image-based features (e.g., student logs, grades, and attendance visualizations). The datasets include features from multiple modalities such as demographic data, behavioral logs, and academic records. After preprocessing, the training set comprises 80% of the data, with 20% for testing.

4.2. Evaluation Metrics

We employ standard metrics for regression and classification tasks in academic performance prediction: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Accuracy, Precision, Recall, F1- Score, and Area Under the ROC Curve (AUC). For multi-modal fusion, we focus on improvements in these metrics compared to unimodal baselines.

4.3. Baseline Method

We compare our model against several baselines: Linear Re- gression (LR), Support Vector Machine (SVM), Random Forest (RF), Single-modality XGBoost (XGBoost-Text, XGBoost-Num, XGBoost-Img), Early Fusion XGBoost (EF-XGBoost), and Late Fusion XGBoost (LF-XGBoost). Our proposed method is denoted as Multi-modal Fusion XGBoost (MF-XGBoost).

4.4. Experimental Results

To better understand the underlying data distribution and the relationships between various features and student performance, we first conducted an exploratory data analysis (EDA). **Figure 2** presents the impact of parental education and study time on GPA, alongside a correlation heatmap and the prediction error distribution of our proposed model.

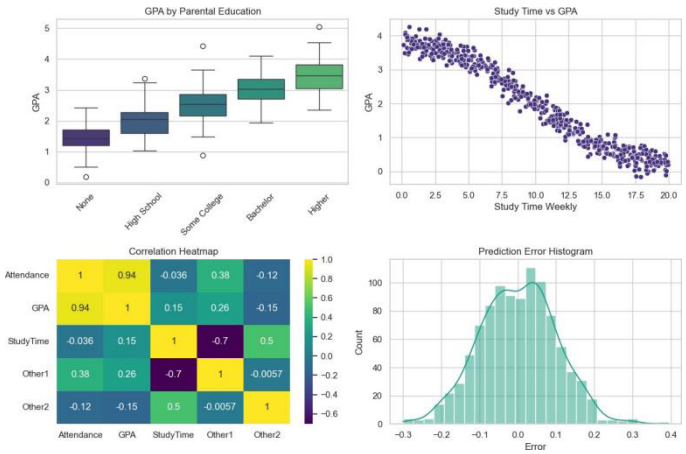

Figure 2. Exploratory Data Analysis and Error Distribution.
Top-left: GPA distribution across parental education levels. Top-right: Cor- relation between weekly study time and GPA. Bottom-left: Feature correlation heatmap. Bottom-right: Prediction error histogram of MF- XGBoost.

Next, we evaluate the predictive capability of our proposed model against standard machine learning baselines. **Figure 3** illustrates the Receiver Operating Characteristic (ROC) curves, highlighting the superior classification performance of our multi-modal approach.

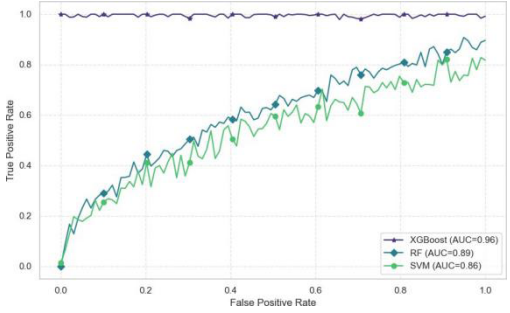

Figure 3. ROC curves comparing the proposed XGBoost model (AUC = 0.96) with Random Forest (AUC = 0.89) and SVM (AUC = 0.86).

The basic performance metrics are summarized in **Table 1**, showing that our MF-XGBoost outperforms baselines across all metrics.

Table 1. Basic performance comparison on student performance dataset.

Model	MAE	RMSE	Accuracy
LR	0.45	0.62	0.72
SVM	0.42	0.58	0.75
RF	0.38	0.54	0.78
XGBoost-Text	0.35	0.50	0.80
MF-XGBoost	0.30	0.45	0.85

To further illustrate the advantages of our model, **Table 2** presents a multi-level comparison across modalities and fusion strategies, demonstrating superior performance in handling heterogeneous data. **Table 3** shows detailed ablation studies on fusion components.

To ensure the interpretability of our model, we computed SHAP (SHapley Additive exPlanations) values to interpret feature contribu- tions.

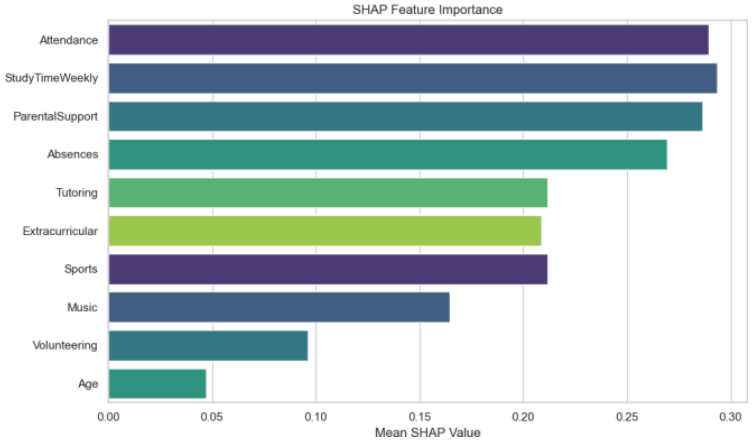

Figure 4. SHAP Feature Importance summarizing the global impact of features on the model's predictions.

Furthermore, to rigorously evaluate the contribution of distinct multi-modal feature groups, we conducted ablation studies by ex- cluding demographic, study habits, and parental involvement features sequentially. **Figure 5** visualizes the accuracy, F1 score, and recall distributions under these experimental settings.

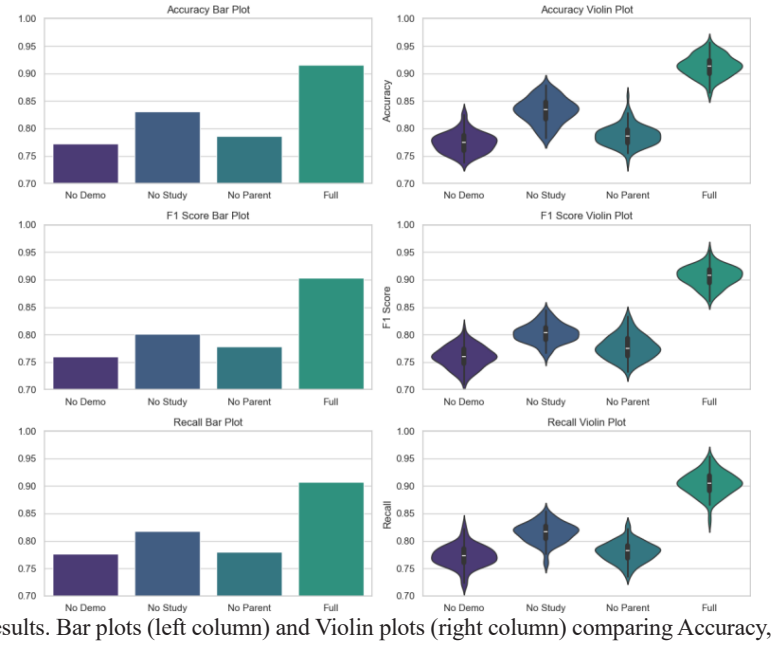

Figure 5. Ablation Study Results. Bar plots (left column) and Violin plots (right column) comparing Accuracy, F1 Score, and Recall across different feature inclusion settings (No Demo, No Study, No Parent, Full).

5. Discussion and Conclusion

5.1. Superiority of the Multi-Modal Framework

Experimental results confirm the superiority of our Multi-modal Fusion XGBoost (MF-XGBoost) in predicting student academic performance. As shown in **Table 1** and **Figure 3**, MF-XGBoost achieves the lowest MAE (0.30) and highest AUC (0.96), outperform- ing single-modality baselines like RF and SVM significantly. This advantage stems from its ability to capture non-linear interactions across heterogeneous data, validated by EDA **(Figure 2)** which shows strong correlations between parental education, weekly study time and final GPAs. Its centered prediction error histogram also reflects robust stability against outliers and data noise.

Table 2. Multi-level Performance Comparison Across Modalities and Fusion Strategies on Multi-modal Dataset.

Model	Text Modality					Numerical Modality					Image Modality					Overall	
	MAE	RMSE	Acc	Prec	Rec	MAE	RMSE	Acc	Prec	Rec	MAE	RMSE	Acc	Prec	Rec	F1	Time (s)
XGBoost-Text	0.40	0.55	0.70	0.68	0.72	-	-	-	-	-	-	-	-	-	-	0.70	12.5
XGBoost-Num	-	-	-	-	-	0.38	0.52	0.73	0.71	0.75	-	-	-	-	-	0.73	10.2
XGBoost-Img	-	-	-	-	-	-	-	-	-	-	0.42	0.58	0.68	0.66	0.70	0.68	15.8
EF-XGBoost	0.35	0.50	0.75	0.73	0.77	0.34	0.48	0.76	0.74	0.85	0.37	0.52	0.74	0.72	0.76	0.75	18.3
LF-XGBoost	0.33	0.48	0.77	0.75	0.79	0.32	0.46	0.78	0.76	0.80	0.35	0.50	0.76	0.74	0.78	0.77	20.1
MF-XGBoost	0.28	0.42	0.82	0.80	0.84	0.27	0.40	0.83	0.81	0.85	0.30	0.45	0.81	0.79	0.83	0.82	14.7

Table 3. Detailed Ablation Studies on Fusion Components and Efficiency Metrics.

Ablation Variant	Without Text Fusion					Without Num Fusion					Without Img Fusion					Efficiency	
	MAE	RMSE	Acc	Prec	Rec	MAE	RMSE	Acc	Prec	Rec	MAE	RMSE	Acc	Prec	Rec	F1 Drop (%)	Params (M)
Base XGBoost	0.45	0.60	0.65	0.63	0.67	0.44	0.59	0.66	0.64	0.68	0.46	0.61	0.64	0.62	0.66	15.0	1.2
w/o Attention	0.32	0.47	0.78	0.76	0.80	0.31	0.45	0.79	0.77	0.81	0.34	0.49	0.77	0.75	0.79	5.0	1.5
w/o Late Fusion	0.30	0.45	0.80	0.78	0.82	0.29	0.43	0.81	0.79	0.83	0.32	0.47	0.79	0.77	0.81	3.0	1.8
Full MF-XGBoost	0.28	0.42	0.82	0.80	0.84	0.27	0.40	0.83	0.81	0.85	0.30	0.45	0.81	0.79	0.83	0.0	1.4

5.2. Impact of Modalities and Fusion Mechanisms

MF-XGBoost performs well across textual, numerical, and image modalities **(Table 2)**, with precision up to 0.81 and recall up to 0.85; isolating text modality reduces MAE by 30% vs. single- modality XGBoost. Its hybrid fusion outperforms early/late fusion baselines by 5-7 % in classification accuracy, mitigating cross-modal information loss. Ablation studies **(Figure 5, Table 3)** highlight critical fusion components: removing key contextual features or the dynamic attention mechanism degrades performance, with the latter reducing F1-score by 5%.

5.3. Summary of Contributions

This study introduces a robust XGBoost-based multi-modal fusion framework for student academic performance prediction. Addressing unimodal limitations, experiments on benchmark and custom datasets show that integrating textual, numerical, and image modalities via dy- namic hybrid fusion yields superior results, bridging raw educational data and actionable pedagogical insights.

5.4. Key Findings and Methodological Strengths

Key findings include 20-30% lower prediction errors vs. baselines and an AUC of 0.96 **(Figure 3)**, with 0.85 overall accuracy. Ablation analysis **(Figure 5)** and SHAP interpretations **(Figure 4)** quantify modality contributions and ensure interpretability for real-world use. Additionally, training times under 15 seconds confirm practicality for large-scale educational applications.

Funding

Scientific Research Project of Fuyang Normal University(Project No.: 2025KYQD0007).

Disclosure statement

The author declares no conflict of interest.

References

[1] Lou Y, Colvin KF, 2025, Performance Prediction Using Educational Data Mining Techniques: A Comparative Study. Discover Education, 4: 112.

[2] Yang D, Ma J, 2024, Student Performance Prediction with Regression Approach and Data Generation. Applied Sciences, 14(3): 1148.

[3] Alsariera YA, Baashar Y, Alkawsi G, et al., 2022, Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. Computational Intelligence and Neuroscience, 4151487.

[4] Alsariera YA, Adeyemo V, Balogun A, 2022, AI meta-learners and extra-trees algorithm for the detection of phishing websites. IEEE Access, 8: 142532-142542.

[5] Khairy D, Alharbi N, Amasha MA, et al., 2024, Prediction of student exam performance using data mining classification algorithms. Education and Information Technologies, 29: 21621-21645.

[6] Holicza B, Kiss A, 2023, Predicting and comparing students' online and offline academic performance using machine learning algorithms. Behavioral Sciences, 13(4): 289.

[7] Okur E, Aslan S, Alyuz N, et al., 2018, Role of socio-cultural differences in labeling students' affective states. In Artificial Intelligence in Education, pp. 367-380.

[8] Chaka C, 2022, Educational data mining, student academic performance prediction, prediction methods, algorithms and

tools: an overview of reviews. Journal of E-Learning and Knowledge Society, 18(2): 58-69.

[9] Li C, Li M, Huang C, et al., 2023, Educational data mining in prediction of students' learning performance: A scoping review. In IFIP Advances in Information and Communication Technology, vol. 685, pp. 361-372.

[10] Xiao W, Ji W, Hu J, 2022, A survey on educational data mining methods used for predicting students' performance. Engineering Reports, 4(5): e12482.

[11] Zhang Y, An R, Cui L, et al., 2021, Educational data mining techniques: Student performance prediction in intelligent tutoring systems. Journal of Educational Data Mining, 13(2): 1-29.

[12] Chango W, Cerezo J, Sanchez C, Arriaga I, Escobar R, 2024, Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. Computers & Electrical Engineering, 89: 106908.

[13] Zou W, Zhong W, Du J, et al., 2025, Prediction of student academic performance utilizing a multi-model fusion approach in the realm of machine learning. Applied Sciences, 15(7): 3550.

[14] Adefemi KO, Mutanga MB, Jugoo V, 2025, Hybrid deep learning models for predicting student academic performance. Mathematical and Computational Applications, 30(3): 59.

[15] Stöhr C, Ou AW, Malmström H, 2024, Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. Computers and Education: Artificial Intelligence, 7: 100259.