


Research

Enhancing high-school dropout identification: a collaborative approach integrating human and machine insights

Okan Bulut¹  · Tarid Wongvorachan²  · Surina He²  · Soo Lee³ 

Received: 17 January 2024 / Accepted: 17 July 2024

Published online: 23 July 2024

© The Author(s) 2024 

Abstract

Despite its proven success in various fields such as engineering, business, and healthcare, human–machine collaboration in education remains relatively unexplored. This study aims to highlight the advantages of human–machine collaboration for improving the efficiency and accuracy of decision-making processes in educational settings. High school dropout prediction serves as a case study for examining human–machine collaboration’s efficacy. Unlike previous research prioritizing high accuracy with immutable predictors, this study seeks to bridge gaps by identifying actionable factors for dropout prediction through a framework of human–machine collaboration. Utilizing a large dataset from the High School Longitudinal Study of 2009 (HSLS:09), two machine learning models were developed to predict 9th-grade students’ high school dropout history. Results indicated that the Random Forest algorithm outperformed the deep learning algorithm. Model explainability revealed the significance of actionable variables such as students’ GPA in the 9th grade, sense of school belonging, self-efficacy in mathematics and science, and immutable variables like socioeconomic status in predicting high school dropout history. The study concludes with discussions on the practical implications of human–machine partnerships for enhancing student success.

Keywords High school dropout · Machine learning · Explainable AI · Human–machine collaboration

1 Introduction

In the wake of substantial technological advancements over the last two decades, such as increased computing power, data storage, and network capabilities, machines have undergone a profound transformation from basic mechanization and automation to a state of intelligentization in recent decades [1]. The contemporary landscape witnesses intelligent machines not only keeping pace with human capabilities but also outperforming them in various scenarios. This transformative trajectory has propelled the dynamic between humans and machines beyond the conventional paradigm of basic human–machine interaction, where machines primarily present information and humans make decisions [2]. The current era embraces an advanced stage of human–machine collaboration characterized by the convergence of cognitive abilities. In this evolved paradigm, both humans and machines exhibit prowess in thinking, decision-making, and synergistically working together to elevate the overall quality of decisions. This symbiotic relationship heralds a new era where the synergy between human intellect and machine capabilities transcends traditional boundaries.

✉ Okan Bulut, bulut@ualberta.ca; Tarid Wongvorachan, wongvora@ualberta.ca; Surina He, surina1@ualberta.ca; Soo Lee, slee@air.org | ¹Centre for Research in Applied Measurement and Evaluation, University of Alberta, 6-110 Education Centre North, 11210 87 Ave NW, Edmonton, AB T6G 2G5, Canada. ²Measurement, Evaluation, and Data Science, University of Alberta, Edmonton, AB, Canada. ³American Institutes for Research, Arlington, VA, USA.



The effectiveness of human–machine collaboration can also be significantly augmented by integrating an explainability layer into machines. This innovative approach, commonly known as Explainable Artificial Intelligence (XAI), serves as a crucial bridge between the complex, opaque decision-making processes of machine learning (ML) models and the human stakeholders interacting with them [3, 4]. In essence, XAI acts as a cornerstone for building trust in the symbiotic relationship between humans and intelligent systems. By mitigating the “black box” problem associated with complex ML algorithms, the layer of explainability empowers machines to elucidate their reasoning processes in a manner that is comprehensible and inherently trustworthy to humans [5]. This transparency not only enhances accountability but also allows human stakeholders to validate and contextualize the decisions made by intelligent systems.

Recent research indicates a growing inclination among humans to engage in collaborative efforts with machines, as highlighted by Haesevoets et al. [6]. The synergistic decision-making process involving both humans and machines has consistently outperformed decisions made in isolation by either party, as evidenced by the findings of Xiong et al. [7]. Although human–machine collaboration has become prevalent in domains like engineering, healthcare, business, and organizational settings, its adoption in education remains significantly understated. This study aims to address this gap by exploring the integration of human–machine collaboration within the educational landscape. Specifically, our focus is on demonstrating the practical implementation of human–machine collaboration in education. To achieve this, we examine the critical issue of predicting high school dropouts. Our goal is to showcase the effectiveness of this collaborative approach within this context and to identify actionable factors contributing to dropout prediction by utilizing various XAI techniques.

2 Theoretical framework

Given the multidisciplinary nature of this study, the following sections provide a detailed literature review covering: (1) a broad definition of human–machine collaboration; (2) the application of human–machine collaboration across various domains; (3) an overview of XAI and commonly used XAI techniques; (4) a review of previous studies utilizing machine learning techniques to predict high school dropouts; and (5) the motivation behind this study.

2.1 Human–machine collaboration

Contemporary human–machine collaboration refers to the synergistic partnership between humans and intelligent machines. These machines can take various forms, including automated systems, autonomous agents, robots, algorithms, or artificial intelligence (AI) entities [8, 9]. This collaborative approach results in enhanced performance by leveraging the strengths of both intelligent machines and human intelligence, addressing their respective limitations [10–12]. Intelligent machines also excel in processing vast amounts of information and generating rational outcomes without succumbing to cognitive biases (e.g., availability bias, representativeness bias, and anchoring effect) or being swayed by internal and external factors (e.g., ability, cognitive style, emotions, workload, fatigue, and time pressure) [13]. Conversely, humans possess unique advantages in employing intuition and experience to discern critical factors, adapt to novel conditions, and rapidly learn and apply reasoning to navigate high uncertainty or tackle new, complex, and rare challenges.

In light of these benefits, human–machine collaboration has experienced a growing application across diverse domains, including engineering [13], healthcare [11, 14], and business [15, 16]. For example, a study by Wilson and Daugherty [17] analyzed 1500 companies spanning 12 industries and found that the most substantial performance enhancements occurred when humans collaborated with machines. Moreover, research indicates that human–machine collaboration surpasses the efficacy of operations involving only humans or machines separately. For instance, a study by Xiong et al. [7] explored the performance of human-only, machine-only, and human–machine joint teams in a sequential risky decision-making task. Specifically, in the Balloon Analogue Risk Task (BART), human and machine participants pump up balloons of different colors in a virtual interface to earn money, facing a choice between securing the current amount in a permanent bank or risking another pump for greater rewards. Each additional pump increases the potential earnings stored in a temporary bank but also raises the risk of the balloon exploding, which would result in losing all the money accumulated in that trial. In the human–machine joint team, the machine can either serve as a subordinate, offering recommendations on whether to pump, or as an equal partner, where both human and machine decisions hold equal weight in a ‘one vote, one rule’ system. The findings of this study revealed that two types of human–machine joint teams both outperformed human-only and machine-only teams. In the human–machine joint team, the machine as

a partner entailed human decision-makers to cede power and coordinate, and their pumping decisions became more conservative and fluctuating.

The trajectory of human–machine collaboration also extends to medical disciplines. A noteworthy example of successful human–machine collaboration in healthcare is evident in cancer detection through the analysis of lymph node cell images [11]. The study demonstrated that combining predictions from a deep learning system with diagnoses from a human pathologist achieved an area under the receiver operating curve (AUC) of 0.995, surpassing the AUC of the deep learning system alone (0.925) and that of the pathologist alone (0.966). This integration resulted in a remarkable reduction in error rates, amounting to at least 85%. Beyond cancer detection, collaborative frameworks have been instrumental in areas such as personalized medicine, where the integration of machine-generated insights with clinical expertise allows for tailored treatment plans based on individual patient characteristics (e.g., [18]).

2.2 Explainable AI

The human–machine collaboration can be enhanced by enabling machines to explain their reasoning in a way that is understandable and trustable to humans [5]. This can be achieved through the integration of XAI [19]. XAI, a sub-field of AI, provides human-interpretable explanations regarding the rationale, strengths, weaknesses, and anticipated behavior of AI systems [4, 20]. In recent years, the significance of XAI has increased due to the widespread applications of advanced AI techniques such as deep learning models. Despite their remarkable accuracy in predictions and classifications, these models are often characterized as “black box” models [20, 21]. This label stems from the reliance of machine learning models on mathematical constructs, featuring an extensive array of abstract, numerical parameters, often numbering in the millions or even billions. These parameters are learned from training data, presenting a challenge in offering profound insights into the intricate dependencies, causal relationships, and internal structures of the models [3, 22]. The opaqueness inherent in these black box models introduces the potential for misleading users [23], raising substantial concerns, particularly in sensitive domains such as healthcare and other applications that involve human life, rights, finances, and privacy [3].

To enhance the interpretability of AI outputs, researchers have proposed various XAI methods. According to the latest comprehensive review conducted by Minh et al. [24], XAI methods fall into three main categories: pre-modeling explainability, interpretable models, and post-modeling explainability. The *pre-modeling explainability* method involves a set of data processing approaches applied to gain insights into datasets used for training ML models. This includes data analysis, summarization, and transformation. On the other hand, *interpretable models* refer to those that can be understood by humans through examination of the model summary or parameters, such as linear models, decision trees, k-nearest neighbors, and rule-based models. Lastly, the *post-modeling explainability* method aims to enhance the interpretability of existing black-box ML models by employing various techniques.

Given its widespread application, Minh et al. [24] categorized post-modeling explainability techniques into four main types. First, *textual justification* generates explanatory text in the form of phrases or sentences. Second, *visualization* provides clarity through visual images, utilizing techniques like layer-wise relevance propagation (LRP) and local interpretable model-agnostic explanation (LIME). Third, *simplification* creates a new and simpler system from complex ML models, employing techniques such as local explanation and example generation. Fourth, *feature relevance* quantifies the importance of input variables, incorporating techniques like SHapley Additive exPlanations (SHAP). According to Minh et al.'s [24] summary, visualization, simplification, and feature relevance emerge as the three commonly used XAI methods, emphasizing their role in rendering AI systems more transparent and understandable.

2.3 Predicting high-school dropout

The issue of high school dropouts has long been a focal point in education. For example, research conducted in Wisconsin revealed that approximately 3000 students discontinue their education before reaching the 12th grade, with around 1500 of these dropouts occurring during the 9th and 10th grades [25]. In response to this concerning trend, researchers, policymakers, and school administrators have focused on developing early warning systems powered by ML models. These systems aim to identify students at risk of dropping out of high school and uncover actionable predictors to inform future interventions and policy adjustments [26, 27].

To date, a plethora of studies have harnessed ML models to forecast high school dropout, though they have not explicitly focused on the perspective of human–machine collaboration. Additionally, most of these studies base their predictions on various background and demographic characteristics exhibited by students, including low grades, aggressive

behavior, student poverty, and high absenteeism. For instance, Sara et al. [28] employed the Random Forest algorithm to predict the dropout status of Danish high school students, utilizing demographic and school-related variables such as gender, school and class size, and teacher–pupil ratio. Chung and Lee [29] similarly utilized RF to anticipate the dropout status of Korean high school students. In contrast, Sansone [30] delved into the dropout phenomena among American students, employing Support Vector Machine, Boosted Regression, and Post-LASSO algorithms. Interestingly, this study discovered that GPA, rather than demographic variables, emerged as the most accurate predictor.

While ML models have been extensively employed in dropout prediction, only a limited number of studies have integrated XAI to comprehend high school or college dropout [16, 31, 32]. For example, Krüger et al. [31] investigated dropout factors within the Brazilian technical school system using XAI methods, specifically SHAP and LIME. The findings highlighted the significance of the year of elementary school completion, the family's minimum wages, and the mother's education and work characteristics as important predictors of dropout. Additionally, Nagy and Molontay [22] also employed XAI techniques, SHAP and LIME, revealing that a higher GPA in high school or higher marks in the mathematics section of the matura exam could significantly reduce the likelihood of college dropout.

A significant drawback in prior research exploring high school dropouts through ML models or XAI lies in the substantial reliance on immutable predictors rather than actionable predictors. Immutable predictors encompass variables over which students, teachers, administrators, and family or community members possess limited or no control—examples include gender, ethnicity, and socioeconomic status. On the other hand, actionable predictors, also known as malleable predictors, denote variables that are recent or real-time, adaptable, and amenable to intervention. These predictors can be utilized to implement tailored interventions or modify the current education system. Examples of actionable predictors include orientation to the future and academic habits of mind, such as self-regulation, self-efficacy, and time management [33].

2.4 Current study

While prior studies have made significant contributions to the domains of human–machine collaboration, explainable AI, and high school dropout prediction, there remain notable gaps that warrant further exploration. Firstly, despite the burgeoning use of human–machine collaboration in fields such as engineering, business, and healthcare, its application within the context of education remains underexplored. However, the potential benefits of incorporating human–machine collaboration in education are substantial. This approach has the capacity to enhance efficiency and accuracy, allowing educators to dedicate more time to personalized teaching methods. Furthermore, the integration of AI can yield results that are more user-friendly, ultimately assisting teachers in improving student engagement and achievement. Therefore, this study seeks to exemplify the implementation of human–machine collaboration in education, using high school dropout prediction as a case study. Secondly, previous research on predicting high school dropouts through ML or XAI techniques has primarily concentrated on achieving higher prediction accuracy based on immutable predictors. However, these predictors offer limited guidance for conducting interventions or modifying the current education system. Consequently, this study aims to address these dual gaps by identifying actionable factors for predicting high school dropouts through the incorporation of the human–machine collaboration paradigm shown in Fig. 1.

3 Method

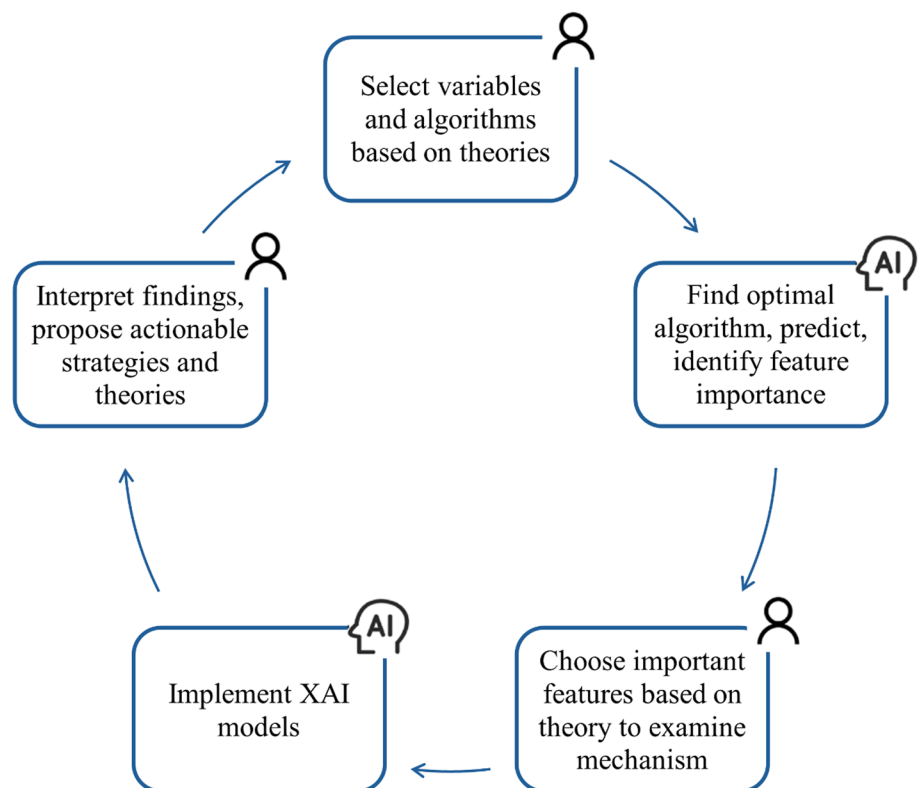
In this section, we detail: (1) the dataset used in this study, including the distribution of demographic information; (2) the preprocessing procedures for selecting variables and handling missing data; (3) data splitting and augmentation procedures; (4) two classification algorithms (i.e., Random Forest and deep learning) employed in this study; and (5) the XAI methods used to interpret the models.

3.1 Dataset

This study used empirical data from the High School Longitudinal Study of 2009 (HSLs:09).¹ HSLs:09 is a nationally representative, longitudinal study that investigated possible factors impacting 9th-grade students' postsecondary

¹ <https://nces.ed.gov/surveys/hsls09/>.

Fig. 1 The proposed framework for human-machine collaboration



education and career trajectories in the United States [32]. In this study, we excluded students for whom the submitted information was not reported by their parents to preserve an accurate representation of variables involving parent-related constructs such as parent education and parental expectations of their child. After this removal, the final sample consisted of 16,137 students.

A review of the students' demographic background revealed that the majority of their parents attained an educational level equivalent to high school or General Educational Development, with $n = 6298$ (39%) for mother/female guardian and $n = 5376$ (33%) for father/male guardian. The gender distribution among the students was balanced, with 8111 males and 8026 females. In terms of ethnicity, the majority of students were white ($n = 9313$; 58%), followed by Hispanic ($n = 2433$; 15%) and Black ($n = 1480$; 9%) students. Students from other races account for $n = 2911$ (18%). Geographically, the majority of the students came from the Southern ($n = 6525$, 40%) and Midwestern ($n = 4332$, 27%) regions of the United States. With respect to the school locale, 5803 students (36%) were from suburban areas, 4686 students (29%) were from city areas, 3784 students (23%) were from rural areas, and 1864 (12%) students were from town areas.

3.2 Data preprocessing

The data preprocessing started with an initial review of the dataset obtained from the HSLS:09 website, encompassing 23,503 students and 67 variables, including the target variable indicating students' high school dropout history. As previously mentioned, students with missing parental responses were excluded, resulting in a final sample of 16,137 students. Within the dataset, 42 variables (63%) were categorical, while 25 variables (37%) were continuous. A thorough examination of the variables in the dataset was conducted, with a focus on identifying and quantifying missing values. The dataset exhibited an overall missing value rate of 14.5%.

To enhance data quality, variables with missing value percentages exceeding 30% were eliminated, thereby reducing the initial set of 67 variables to 51. Subsequently, the remaining missing values underwent replacement using a Random Forest-based multivariate imputation through chained equations, employing the mice package [13] in R (R Core Team, 2022). Following imputation, a correlation analysis was executed on the dataset to identify variables that were not correlated with the target variable (i.e., high school dropout history). Figure 2 shows the correlation matrix

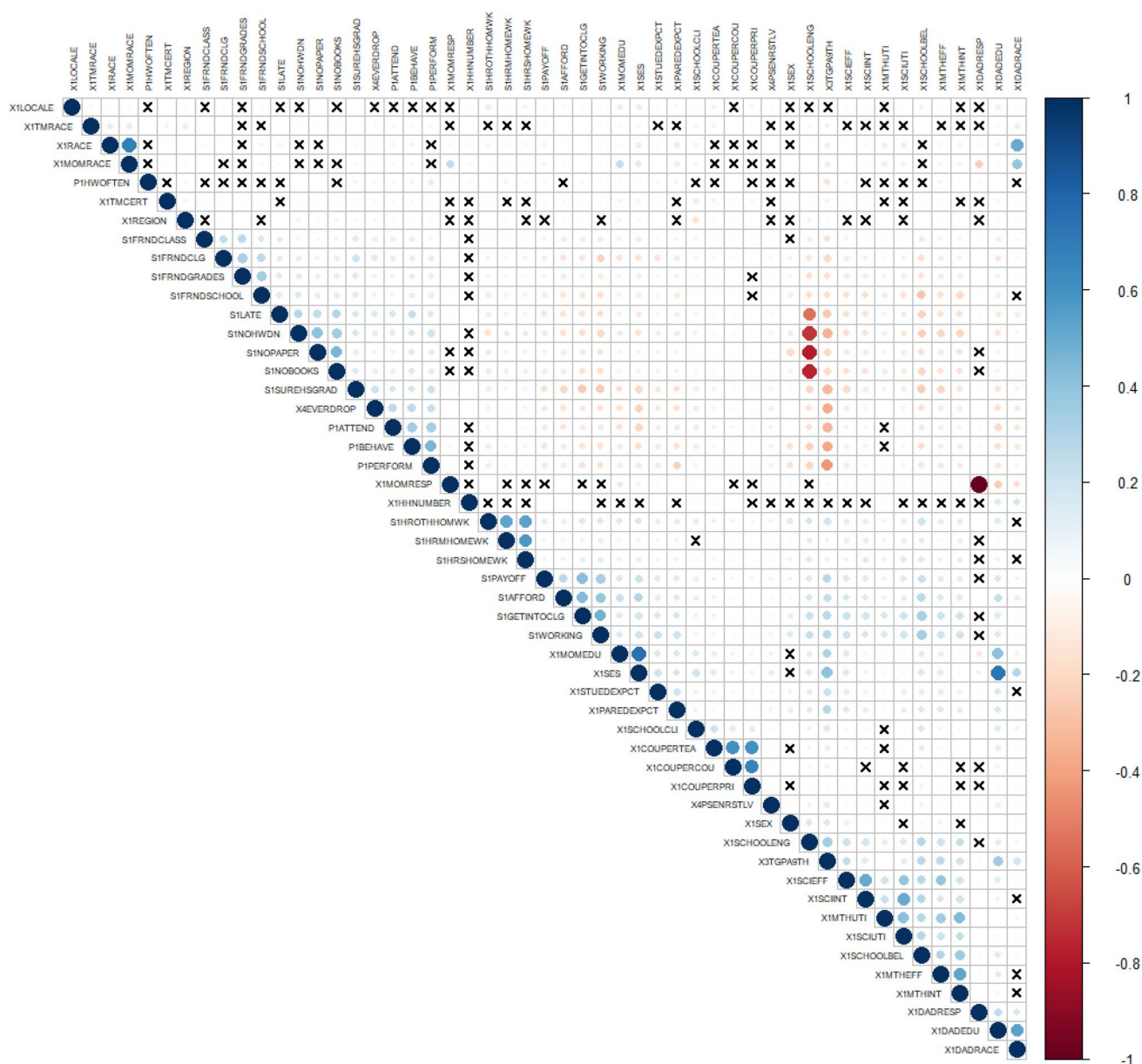


Fig. 2 The correlation matrix plot for the pre-trimmed dataset

of the pre-trimmed dataset. Further refinement involved removing variables based on their theoretical relevance and correlation insignificance. The final dataset, comprising 16,137 students and 37 variables (23 categorical and 14 continuous), underwent a final correlation analysis, as illustrated in Fig. 3.

3.3 Data split and augmentation

The preprocessed dataset was split into two parts: a training dataset (70%) and a testing dataset (30%). Next, the proportion of the target variable (i.e., high school dropout history) was examined prior to the predictive modeling phase. The proportion between the two classes of the target variable in the dataset appeared highly skewed (i.e., a small number of dropout cases relative to the number of students who graduated from high school). In the training dataset, there was a severe class imbalance in the high school dropout history variable, with 1376 students dropping out before high school graduation and 9919 students who did not drop out (the disparity ratio was approximately 7:1). This imbalance was expected due to the occurrence rarity of the dropout phenomenon [34].

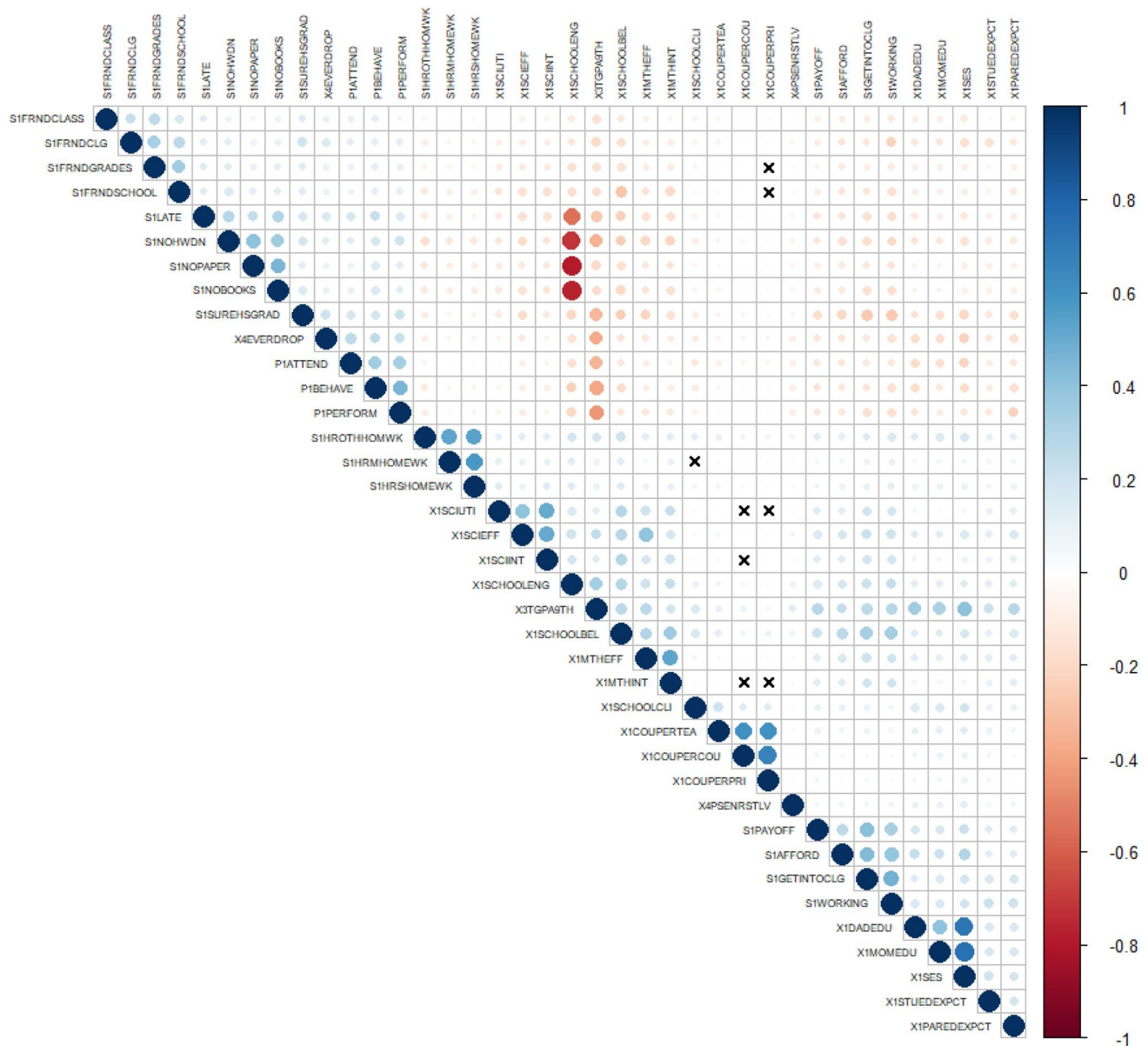


Fig. 3 The correlation matrix plot for the final dataset

To address the imbalance issue, we utilized a hybrid resampling technique involving both the Synthetic Minority Oversampling Technique for Nominal and Continuous (SMOTE-NC) and Random Undersampling (RUS) [12, 15]. This combination of techniques allowed us to synthesize the minority class while also undersampling the majority class. The SMOTE-NC configuration for synthesizing the minority data points was based on their five nearest neighbors and 0.8 resampling ratios. The final sample of the training dataset was $n = 15,870$, with 7935 cases for each class of high school dropout history. We did not balance classes of the target variable in the testing dataset to reflect real-life conditions with class imbalance. Hence, the final sample size of the testing dataset was $n = 4842$, with 4214 for the majority class (i.e., non-dropped-out students) and 628 for the minority class (i.e., dropout students) of high school dropout history.

3.4 Classification algorithms

We utilized two classification algorithms to predict high school dropout, namely Random Forest with a collection of decision tree classifiers [35] and deep learning through the Keras library [27, 32] in Python.

3.4.1 Classification with random forest

For the Random Forest classifier, a randomized grid search was performed to look for the optimal hyperparameter values. The search space comprised 50 sets of hyperparameter values fitted with threefold cross-validation (threefold CV), totaling 150 model fits. Candidate hyperparameter values were selected for the maximum tree depth (`max_depth`), number of trees (`N_estimators`), and number of features factored in determining the best node split (`max_features`). The options for `max_depth` values were generated from an array of 20 evenly spaced values between 100 and 500 (inclusive). The options for `N_estimators` were generated from an array of 20 evenly spaced values between 200 and 2000 (inclusive). For `max_features`, the options were 'auto' in which all features were used, 'sqrt' in which the square root of the total number of features was used to split, and "log2," in which log 2 of the total number of features was used to split. The resulting hyperparameter values for the high school dropout prediction in this study were `max_depth`: 436, `n_estimators`: 1621, and `max_features`: 'sqrt'.

Subsequently, the fine-tuned Random Forest model was used to select optimal features for the prediction with recursive feature elimination with cross-validation (RFECV). RFECV was configured with `step = 1` to sequentially remove one predictor at a time and `CV = 5` to perform fivefold cross-validation to fit and evaluate predictor candidates. As a result, 16 predictors were retained for the classification (see Table 2). The performance of the fine-tuned Random Forest model was evaluated with tenfold cross-validation on the test dataset. The prediction results of the model were consulted with a mean and standard deviation of accuracy, precision, recall, and the AUC score.

3.4.2 Classification with deep learning

For the deep learning classifier with Keras, we utilized a dropout regularization layer with a sequential model to prevent overfitting [36]. To ensure methodological consistency, we applied the deep learning classifier to the same dataset used for the Random Forest classifier. We developed a six-layer neural network model. The model architecture was as follows: (1) an input layer with 17 features to reflect the maximum number of features; (2) a hidden layer with 128 units and a rectified linear unit (ReLU) activation, followed by a dropout layer with a dropout rate of 0.4;² (3) hidden layer with 32 units and ReLU activation, followed by another dropout layer with a dropout rate of 0.3; and (4) an output layer with 1 unit and a sigmoid activation function. The sigmoid activation reflected the nature of the classification task, as the output was a probability ranging between 0 and 1.

In the training phase, the deep learning model was compiled using the binary cross entropy loss function and the Adam optimizer to adjust the learning rate throughout training. The model was fitted with `validation_split = 0.2` to subset 20% of the dataset for testing purposes, `epoch = 500` to make the model iterate through the dataset 500 times, and `batch_size = 50` to make each batch contain 50 cases in updating the model. Evaluation metrics include loss rate, mean squared error (MSE), and binary accuracy. Note that AUC was not included in the evaluation metric of this algorithm because it is a global metric that evaluates the model as a whole. However, Keras operates on batches of data during training, making it potentially misleading to compute AUC directly.

3.5 Model explainability

After performing classification tasks, results from the highest-performing model were examined with XAI to explain the prediction, both at the global and local levels. We utilized the moDel Agnostic Language for Exploration and eXplanation (DALEX) module in Python to perform the XAI analysis [34]. For a global-level explanation, where the impact of variables on the model's prediction as a whole can be assessed, we employed permutation-based variable importance analysis to identify influential predictors of high school dropout [37]. Subsequently, the influential predictors identified through this analysis were further examined using partial dependence profiling to understand how the prediction result changes in relation to the selected explanatory variable [38].

For the local-level explanation that concerns the prediction of each individual case, the breakdown method, SHAP value, and the LIME method were used to explain non-dropout cases and dropout cases [34, 37]. The breakdown method indicates the contribution of variables to the model's prediction of a selected observation [25]. SHAP value, similar to the

² During the model training phase, a fraction of randomly selected neurons (in this case, 40%) in the previous layer was set to zero at each update, which helped prevent overfitting.

Table 1 Classification results from the random forest and deep learning classifiers

Classifier	Performance metrics	<i>M</i>	<i>SD</i>
Random forest	Accuracy	0.88	0.01
	Precision	0.63	0.14
	Recall	0.17	0.05
	ROC-AUC	0.78	0.03
	MSE	0.11	0.01
Deep learning	Accuracy	0.87	0.01
	Precision	0.42	0.06
	Recall	0.21	0.02
	ROC-AUC	0.70	0.02
	MSE	0.10	0.01

breakdown method, explains the contribution of each variable to the final prediction. The difference, however, is that this method calculates the average contribution of each feature over all possible orders to account for possible interactions [25]. Finally, the LIME method explains the classification result by outlining features that contribute to the prediction or serve as evidence against the prediction [39].

4 Results

The results section is organized into three parts. First, we present the classification results of Random Forest and deep learning models. Second, we provide global-level explanations, identifying which features are crucial for predicting high school dropout. Finally, we delve into local-level explanations, exploring how specific features contribute to the predictions.

4.1 Classification outcomes

Table 1 presents the outcomes of the Random Forest classifier, providing the mean and standard deviation (SD) values for accuracy, precision, recall, AUC, and MSE over 10 iterations of cross-validation. For comparison, the results for the deep learning classifier in Table 1 also describe the mean and SD values for accuracy, precision, recall, and MSE across 500 epochs. Comparing the evaluation metrics of both algorithms revealed that the Random Forest classifier exhibited a comparable mean accuracy to the deep learning classifier, with 0.88 for the Random Forest classifier and 0.87 for the deep learning classifier. Both models also exhibited comparable MSE, with 0.11 for the Random Forest classifier and 0.10 for the deep learning classifier.

Although the Random Forest and Deep Learning classifiers yielded comparable outcomes in terms of accuracy and MSE, the Random Forest classifier outperformed the deep learning classifier in two key areas. Specifically, it achieved a higher precision score (0.63 compared to 0.42), and a higher AUC value (0.78 compared to 0.70). In the context of high school dropout prediction, precision is an important metric because resources for dropout prevention programs are often limited, making the allocation of the resources to students who are most likely to drop out highly important. Similarly, AUC is an important metric that measures the classifier's ability to assign higher probabilities to positive instances than to negative instances. A high AUC value suggests that the classifier is effective in distinguishing between students who drop out and those who do not. This discrimination capability is vital for ensuring that intervention and support efforts are directed towards students at higher risk of dropping out, enhancing the overall efficacy of dropout prevention strategies. The combination of higher precision and AUC values positions the Random Forest classifier as a more suitable choice for this high-stakes task of dropout prediction, where the consequences of misallocation of resources can have significant real-world implications.

The observed performance gap between the deep learning and Random Forest classifiers was anticipated, aligning with the well-documented characteristics of deep learning algorithms. Deep learning models, known for their intricate neural network architectures, typically demand a substantial amount of data to generalize effectively. However, the inherent complexity of these algorithms may result in diminished accuracy when faced with relatively smaller datasets, and an increased risk of overfitting can further exacerbate performance issues [40, 41]. The HSL5:09 dataset may not

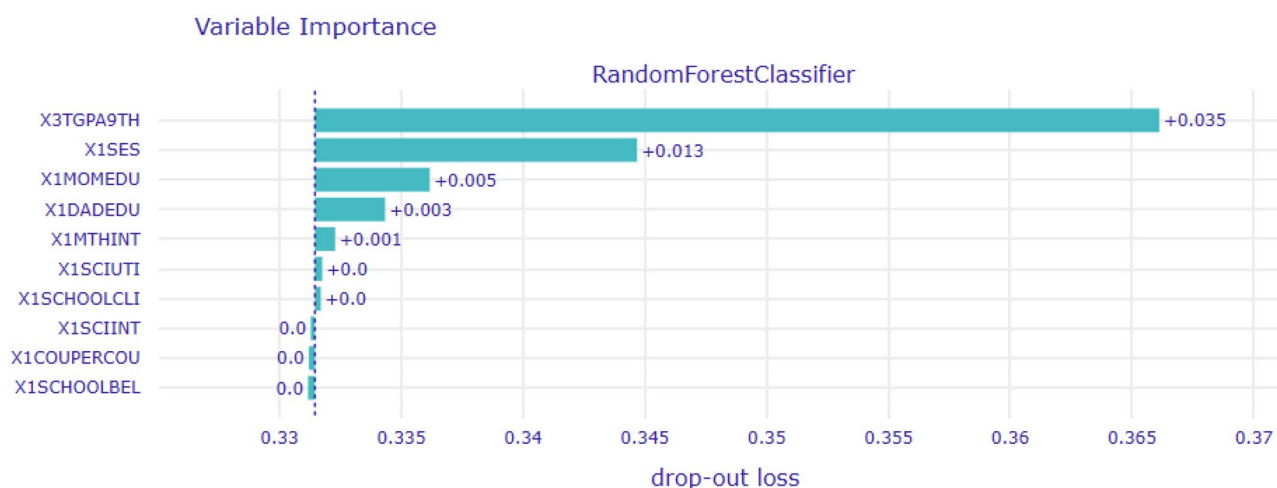


Fig. 4 The results of permutation-based variable importance analysis

fully harness the potential of deep learning models, leading to a significant difference in performance compared to the Random Forest classifier. The Random Forest classifier's superior accuracy and AUC values underscore its suitability for the specific task of dropout prediction in this dataset. Recognizing the importance of interpretability in decision-making processes involving human stakeholders, including teachers, principals, and other school-based professionals, we acknowledge the need for an XAI analysis. Given the Random Forest classifier's enhanced performance, it will be the focus of the XAI analysis. The interpretability afforded by the XAI analysis is critical for empowering stakeholders with the ability to comprehend the model's rationale, facilitating more informed decisions regarding intervention strategies for at-risk students.

4.2 XAI: global-level explanation

For a global-level explanation, Fig. 4 presents the results of the permutation-based variable importance analysis. The most influential variable in students' high school dropout was students' 9th-grade GPA (X3TGPA9TH), followed by students' socioeconomic status (X1SES), father's/male guardian's highest level of education (X1DADEDU), mother's/female guardian's highest level of education (X1MOMEDU), and students' interest in math (X1MTHINT). The examination of the remaining variables indicated that the contribution of the variables to the dropout prediction was relatively minor. This is evidenced by their dropout loss (see Appendix 2), which was less than 0.001 on the prediction outcome.

Figure 5 presents the results of the partial dependence profiling. Students' 9th-grade GPA (X3TGPA9TH) exhibited the strongest impact on the prediction of the model, as seen from the broad range of changes in the prediction, spanning from approximately 0.1 to 0.6 on the y-axis. As students attained higher GPAs, their likelihood of being predicted as high school dropouts gradually decreased. A substantial dip in the prediction was observed at a GPA value of 2.5, suggesting that students whose GPA was around 2.5 or higher had a much smaller probability of being predicted as high school dropouts compared to those within the GPA range of 1 to 2.5.

The second most influential predictor was students' socioeconomic status (X1SES). Intriguingly, the non-linear impact of socioeconomic status on high school dropout rates introduces a nuanced perspective on the conventional understanding of the relationship between socioeconomic status and educational outcomes. Contrary to the assumption that higher levels of socioeconomic status would invariably correlate with an increased likelihood of graduation, the U-shaped pattern in Fig. 4 suggests a more complex dynamic. While students with lower socioeconomic status are at a heightened risk of dropout, the surprising downturn in the likelihood of dropout within the range of -1 to 1 implies that a high level of socioeconomic status may not guarantee graduation either. This unexpected finding underscores the multifaceted nature of the factors influencing educational attainment, emphasizing the need for a more comprehensive understanding of the interplay between socioeconomic status and academic success.

The two variables of the father's level of education (X1DADEDU) and the mother's highest level of education (X1MOMEDU) similarly indicated a negative influence on the prediction of dropout status. Specifically, students whose

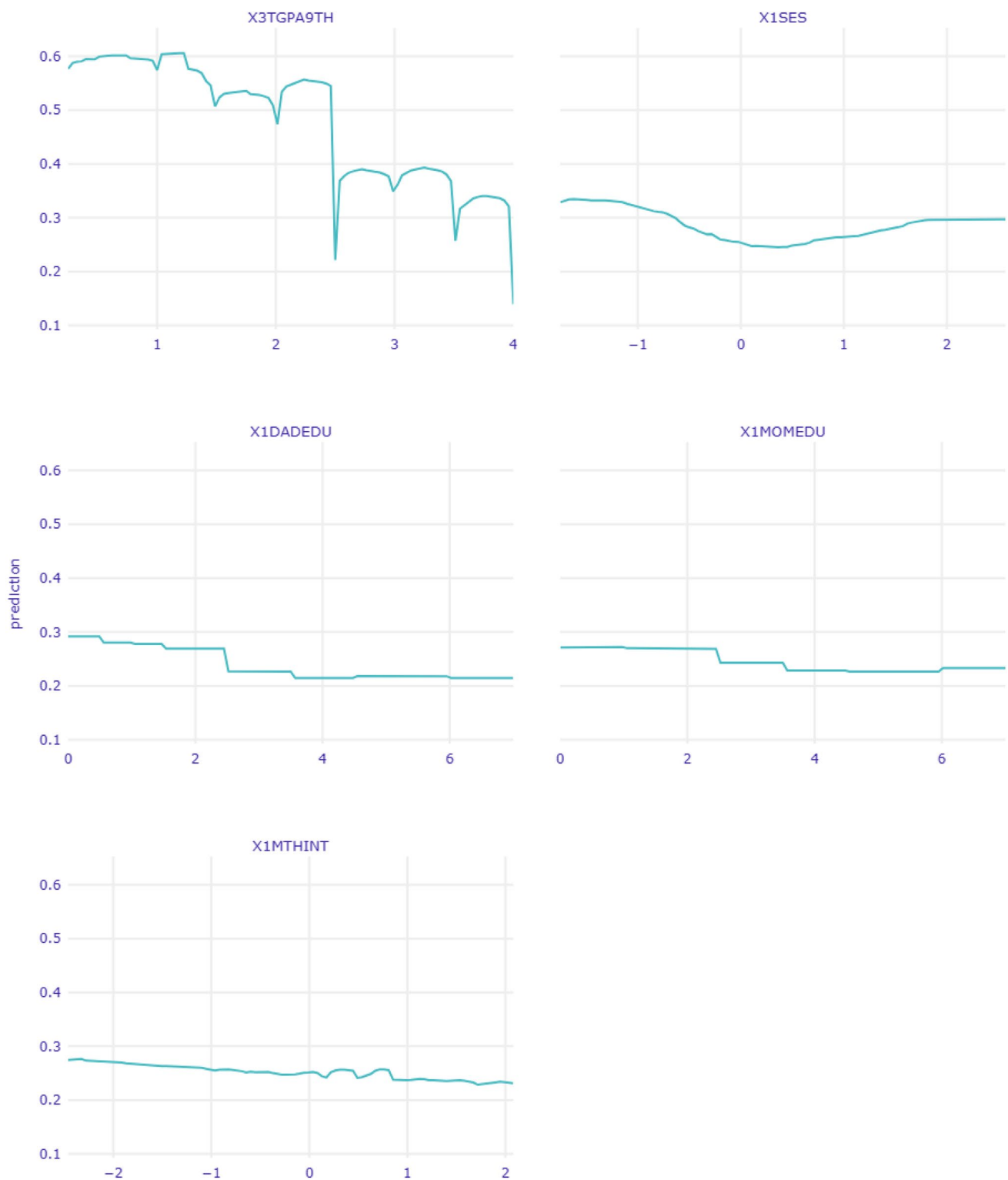


Fig. 5 Partial dependence profiles of the five most influential variables

parents had lower levels of education exhibited the highest likelihood of being predicted as high school dropouts. This likelihood gradually diminished as the educational level of the parents increased.

A parallel pattern emerged in the analysis of students' interest in mathematics (X1MTHINT), revealing a significant correlation with the predicted outcomes. Notably, a negative impact on the prediction results was discerned, as students

Table 2 Predictor values for three students in the HSL5:09 dataset

Predictors	No dropout	Dropout with a low GPA	Dropout with a moderate GPA
X1MOMEDU	1	2	1
X1DADEDU	1	0	1
X1SES	−1.6338	−0.6333	−0.8049
X1MTHEFF	0.9	−2.04	0.1
X1MTHINT	0.55	−1.68	−0.18
X1SCIUTI	−0.33	−1.32	0.1
X1SCIEFF	0.67	0.38	−1.02
X1SCIINT	0.16	−1.38	−0.17
X1SCHOOLBEL	−0.39	−0.39	0.49
X1SCHOOLENG	−0.32	1.39	−0.72
X1SCHOOLCLI	−1.5	−0.22	0.12
X1COUPERTEA	0.31	0.78	0.87
X1COUPERCOU	0.33	1.15	0.61
X1COUPERPRI	0.8	1.1	0.54
X3TGPA9TH	4	1	3
S1HROTHHOMWK	2	1	1

with lower X1MTHINT values, indicative of diminished interest in mathematics, exhibited a heightened probability of being predicted as high school dropouts. Conversely, those with higher values demonstrated a lower likelihood of dropout prediction. This observation implies a compelling link between students' level of interest in mathematics and their motivation to actively participate in their studies, suggesting that a diminished interest in the subject may contribute to a decline in overall academic performance [1]. This underscores the importance of recognizing and addressing motivational factors, such as interest in specific subjects, in the formulation of strategies aimed at reducing high school dropout rates.

4.3 XAI: local-level explanation

The local-level explanation of this study presents three cases of prediction: (1) a non-dropout case, (2) a dropout case with a low-grade 9th-grade GPA, and (3) a dropout case with a moderate 9th-grade GPA but low in other predictors. Table 2 displays the values of the predictors of these three cases.

4.3.1 Non-dropout case

Figure 6 presents the LIME explanation, the breakdown plot, and the SHAP plot for the non-dropout case. The LIME explanation indicates that the student holds an 83% likelihood of being predicted as a non-dropout student and a 17% chance of being predicted as a dropout student. Key predictors opposing the dropout prediction include their 9th-grade GPA (X3TGPA9TH: 4.0) and their positive self-efficacy levels in mathematics (X1MTHEFF: 0.90) and science (X1SCIEFF: 0.67). Among these, the 9th-grade GPA stands out as the most influential factor against the dropout prediction.

Conversely, predictors contributing to the dropout prediction are the low level of education for their father/male guardian (X1DADEDU: 1) and mother/female guardian (X1MOMEDU: 1), the low socioeconomic status of their family (X1SES: −1.63), the negative feeling of belonging at school (X1SCHOOLBEL: −0.39), lower levels of expectations from the counselor (X1COUPERCOU: 0.33) and the principal (X1COUPERPRI: 0.80) in their school, and the low perception of science utility (X1SCIUTI: −0.33). For this particular student, family socioeconomic status (X1SES) was the highest contributing factor to the dropout prediction. The predictive variables for this student revealed that, despite the presence of certain unfavorable factors, a high GPA played a pivotal role in contributing to the individual's success in continuing their education without dropping out.

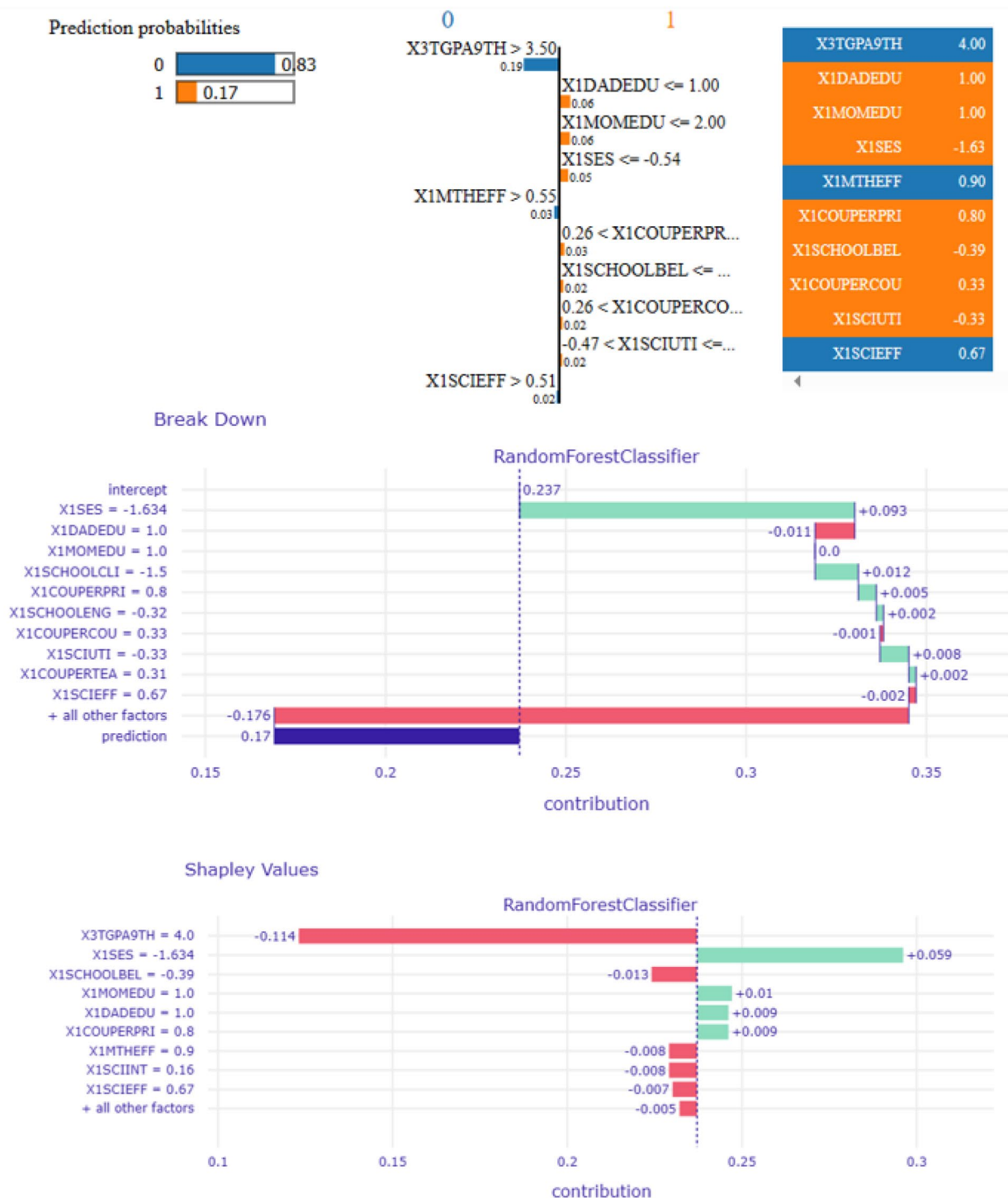


Fig. 6 The LIME explanation, the breakdown plot, and the SHAP plot of a non-dropout case

4.3.2 Dropout case with low 9th-grade GPA

Figure 7 presents the LIME explanation, the breakdown plot, and the SHAP plot of the dropout case with a low 9th-grade GPA. The LIME explanation indicates that this student holds a 10% likelihood of being predicted as a non-dropout student and a 90% chance of being predicted as a dropout student. The key predictors opposing the dropout prediction include positive expectations of counselors at their school (X1COUPERCOU: 1.15) and time spent doing homework on a typical

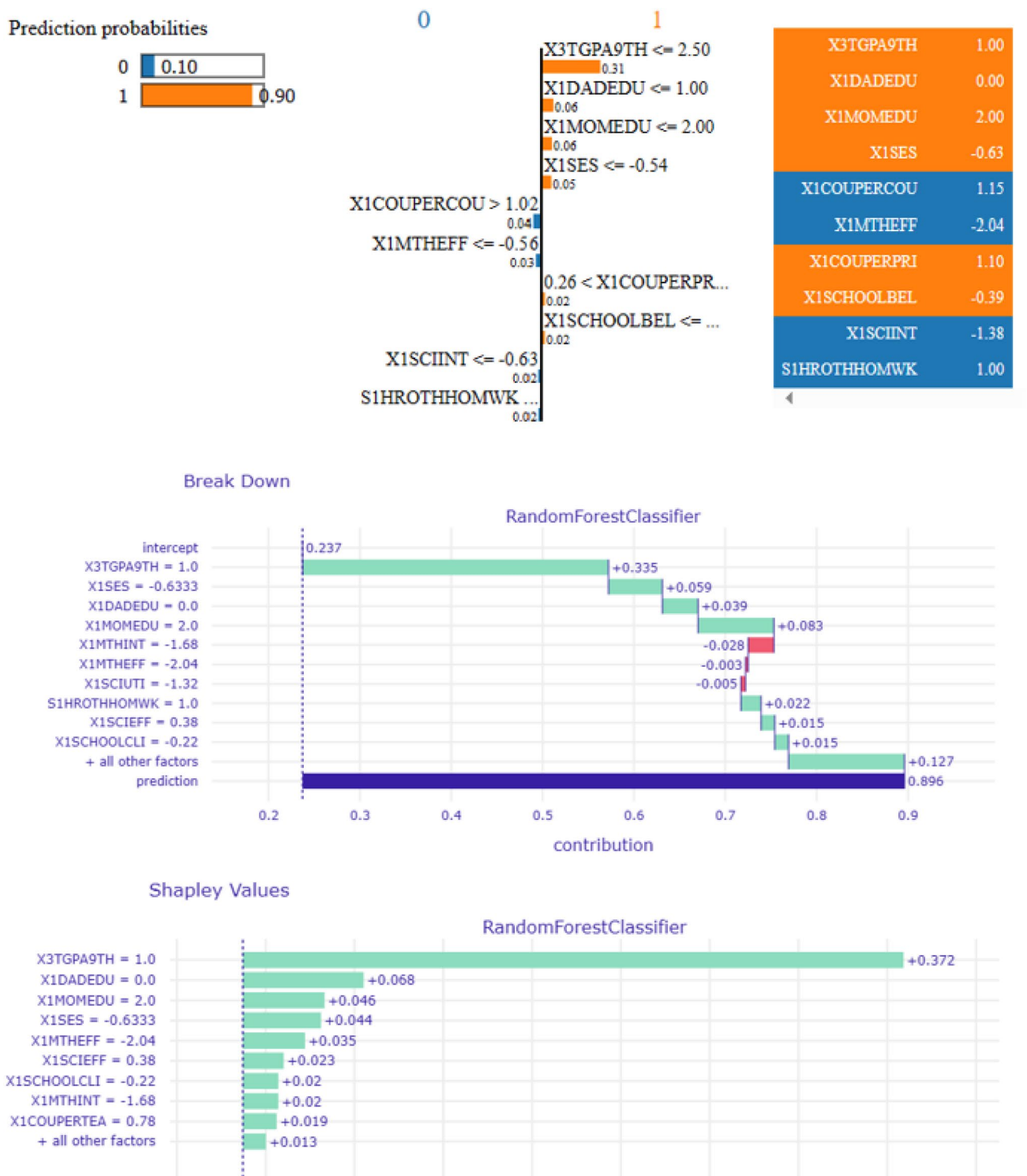


Fig. 7 The LIME explanation, the breakdown plot, and the SHAP plot of a dropout case with a low 9th-grade GPA

school day (S1HROTHHOMWK: 1). Among these predictors, the expectation of counselors at their school stands out as the most influential factor against the dropout prediction.

Conversely, predictors contributing to the dropout prediction are the low 9th-grade GPA (X3TGPA9TH: 1.0), low socioeconomic status of their family (X1SES: -0.63), low level of education for their father/male guardian (X1DADEDU: 0) and mother/female guardian (X1MOMEDU: 2), negative sense of school belonging (X1SCHOOLBEL: -0.39), negative interest

in math (X1MTHINT: -1.68), adverse school climate (X1SCHOOLCLI: -0.22), low teacher expectations (X1COUPERTEA: 0.78), low levels of self-efficacy in science (X1SCIEFF: 0.38) and mathematics (X1MTHEFF: -2.04). Among these variables, the most influential factor driving the dropout prediction was the student's 9th-grade GPA. The explanatory variables for this case revealed a different trend compared to the non-dropout case discussed above. In this case, a low GPA emerged as the primary factor increasing the student's risk of dropout. The explanation may emphasize that a low GPA reflects poor academic performance, serving as a crucial indicator of the student's struggles in their classes that contribute to the increased likelihood of dropout.

4.3.3 Dropout case with moderate 9th-grade GPA but low in other predictors

Figure 8 presents the LIME explanation, the breakdown plot, and the SHAP plot of the dropout case with a moderate ninth-grade GPA, respectively. The LIME explanation indicates that this case holds a 27% likelihood of being predicted as a non-dropout student and a 73% chance of being predicted as a dropout student. Key predictors opposing the dropout prediction include their 9th-grade GPA (X3TGPA9TH: 3.0) and their positive feeling of school belonging (X1SCHOOLBEL: 0.49). Among these variables, the 9th-grade GPA stands out as the most influential factor against the dropout prediction.

Conversely, predictors contributing to the dropout prediction were the low socioeconomic status of their family (X1SES: -0.80), low levels of self-efficacy in mathematics (X1MTHEFF: 0.10) and science (X1SCIEFF: -1.02), low levels of education for their father/male guardian (X1DAEDU: 1) and mother/female guardian (X1MOMEDU: 1), low albeit positive principal expectation (X1COUPERPRI: 0.54), negative school engagement (X1SCHOOLENG: -0.72), negative interest in math (X1MTHINT: -0.18), low counselor expectation (X1COUPERCOU: 0.61), low teacher expectation (X1COUPERTEA: 0.87), adverse school climate (X1SCHOOLCLI: 0.12), and limited hours spent on homework/studying on typical school day (S1HROTHHOMWK: 1.0). Among these, the most influential factor driving the dropout prediction is the student's low self-efficacy in science (X1SCIEFF). The explanation may emphasize that despite having a moderate GPA, this student might have many challenges both at the individual level (i.e., low self-efficacy in math and science, interest in math, low school engagement) and the socio-environmental level (i.e., low socioeconomic status, low educational level in parents, and adverse school climate). These challenges may outweigh the mitigating effect of their GPA, heightening the risk of dropout.

5 Discussion

The primary objective of this study was to highlight the efficacy of human-machine collaboration in education and exemplify human-machine collaboration through a case study of high school dropout prediction. The growing importance of human-machine collaboration in education is driven by the need to enhance learning experiences, personalize education, and improve educational outcomes [42]. In this study, we argue that a human-in-the-loop approach can combine the strengths of both educators and AI technologies, leading to more effective teaching and learning processes. While educators participate in the model development stage (e.g., identifying important and actionable predictors of high school dropout), machines can help find the best predictive model more efficiently. Furthermore, we posit that utilizing XAI techniques can improve educators' understanding of the insights and recommendations provided by AI systems, thereby fostering trust in the adoption and use of AI-powered tools in education.

Beyond mere advancements in predictive accuracy, this study harnesses the synergistic potential of various XAI techniques to unveil crucial factors influencing dropout predictions. The integration of human expertise with machine learning algorithms not only enhanced the performance of predictive models but also yielded actionable insights critical for targeted interventions and systemic improvements. This approach exemplifies how combining human insights with machine intelligence can lead to superior outcomes compared to relying on either independently. The efficiency gained through human-machine collaboration was particularly evident in the feature selection process. Initially, important predictors were identified based on their theoretical relevance, leveraging human expertise. Subsequently, the RFECV algorithm was employed to further refine the selection, illustrating the valuable assistance provided by machine learning algorithms.

The findings of this study also underscore the importance of XAI in understanding and utilizing predictors of high school dropout. By applying various XAI techniques, our study unveiled critical factors influencing dropout predictions. These factors included high school GPA as an indicator of student achievement, students' sense of school belonging, perception of science utility, interest in mathematics, and self-efficacy in both mathematics and science. XAI not only

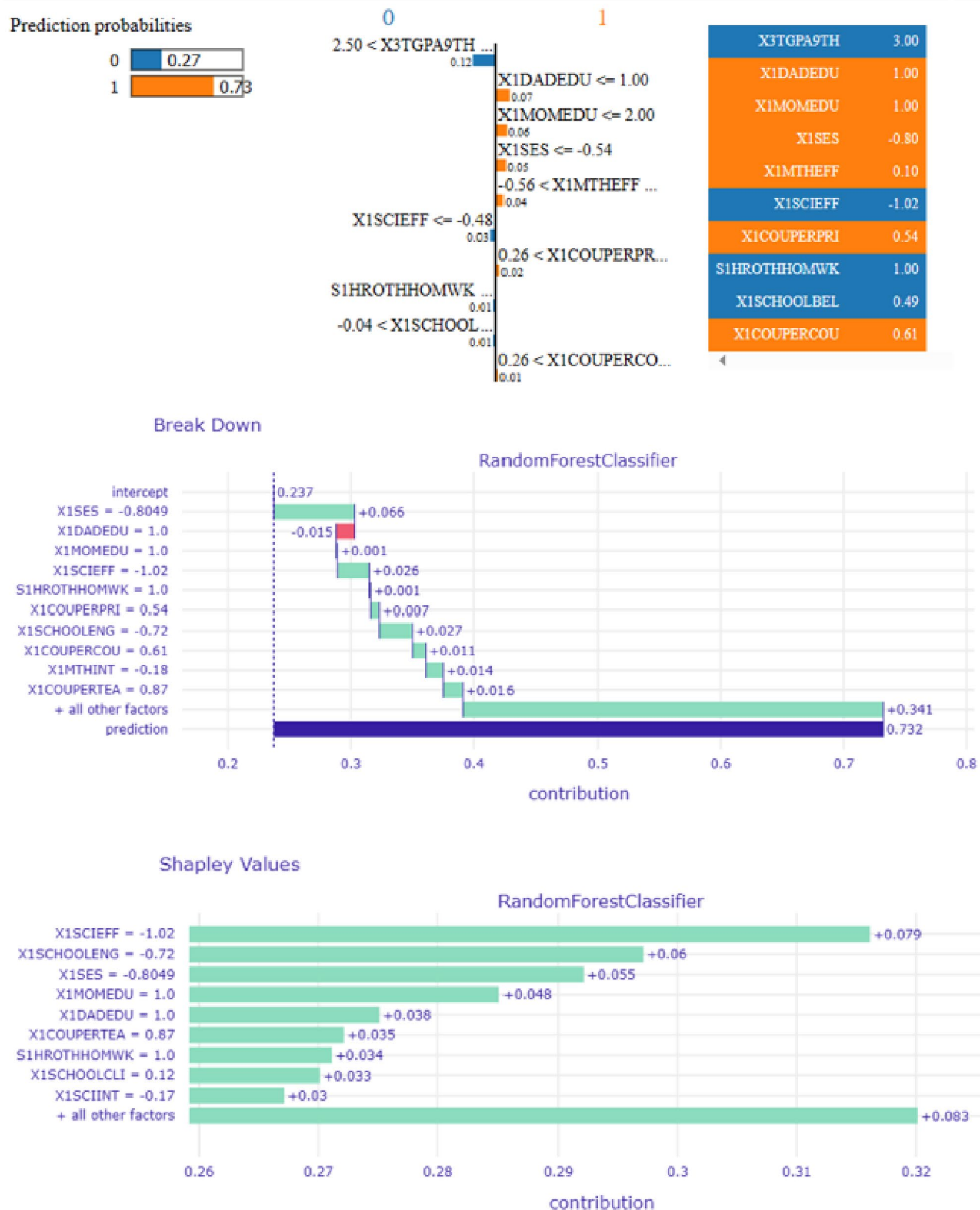


Fig. 8 The LIME explanation, the breakdown plot, and the SHAP plot of a dropout case with a moderate 9th-grade GPA

improved model transparency but also provided insights that are essential for crafting effective educational strategies.

For example, improving students' perception of science utility can be achieved by presenting the subject matter in a context that highlights its real-world applications, rather than relying solely on traditional textbook-based lectures. This approach has been shown to foster a deeper understanding and greater engagement among students [43]. Similarly, enhancing students' sense of school belonging through collaborative efforts among parents, teachers, and students can positively impact students' well-being and academic performance [44]. Furthermore, initiatives such as parent-teacher-school conferences have proven effective in creating a supportive environment, leading to improved school attendance and involvement in academic activities [45]. This collaborative intervention holds the promise of positively impacting students' overall well-being, translating into improved school attendance, enhanced academic performance, and heightened involvement in various academic activities, as elucidated by previous research [46].

This collaborative approach stands as a testament to the transformative impact that human-machine partnerships can have on shaping educational strategies and fostering student success. One noteworthy point is that the collaboration between humans and machines should not be limited to a superficial application where the machine's input is used just once. Instead, it should evolve into an iterative cycle where the insights generated by machines and humans inform and validate each other. In this approach, human insights can dynamically refine and guide the modeling process, continually incorporating and responding to preliminary results. In the context of high school dropout, we anticipate that different stakeholders in education (e.g., teachers, school administrators, and parents) can leverage the insights derived from predictive models to address risk factors associated with high school dropout and build a more supportive learning environment for students.

5.1 Limitations and future research

This study has several limitations worth noting. First, both predictive models utilized in this study exhibited low recall, suggesting room for improvement in identifying dropout cases [35]. A high prediction accuracy with low recall indicates that while the model may not identify all dropout instances, it is reliable when it does [35]. This limitation could be a consequence of the class imbalance in the targeted variable, despite being mitigated by the resampling process [15]. Future studies could explore various class imbalance mitigation techniques to identify the most effective strategy for dropout prediction. In this study, we opted for a single technique as our focus was not on testing multiple imbalance mitigation strategies, which is typically the aim of methodological papers.

Second, there are some conflicts in the prediction that may happen due to the interaction of the predictors. For instance, consider the dropout case with a low GPA. The LIME explainer and breakdown plot suggest that students' negative self-efficacy in mathematics serves as evidence against the dropout prediction, albeit with low influence, a result that may seem counterintuitive from a theoretical perspective. In contrast, its SHAP plot indicates a stronger positive contribution of negative math self-efficacy to the dropout prediction, a suggestion aligned with the existing literature.

Another discrepancy is in the impact of students' hours spent on homework/studying on a typical school day (S1HROTHOMWK) in predicting students' dropout. In the case of a student with a low 9th-grade GPA (refer to Sect. 4.3.2), the variable counteracts the dropout prediction. Conversely, in the case of a student with a moderate 9th-grade GPA (refer to Sect. 4.3.3), it contributes to the prediction of a dropout. Interestingly, this variable is absent in the SHAP output for the former scenario, suggesting that its influence might be less significant compared to the latter scenario. This discrepancy can be attributed to the thoroughness of the SHAP analysis compared to the LIME and the breakdown method. SHAP analyzes every possible combination of predictors, accounting for predictor interactions, albeit at the expense of longer computational time. Consequently, suggestions from SHAP may take precedence over the other two analyses when supported by theory and when they align more logically with the broader context. This limitation could serve as a guideline for future studies to consider results from multiple analyses in interpreting the predictions of a model.

5.2 Conclusion

This study underscores the paramount importance of cultivating a collective intelligence framework, wherein human involvement remains pivotal in validating the outcomes of predictive models before initiating any actionable measures. Despite the remarkable advancements achieved through AI and machine learning techniques, the indispensable role of human validation cannot be overstated. Establishing a synergy between AI and human expertise ensures a comprehensive and nuanced understanding of the intricate factors influencing high school dropout predictions. In the case of school dropout, as illustrated in this study, the utilization of information derived from the machine learning algorithm can offer insights into students' interest in a subject or their engagement in class. These variables may not be immediately

apparent through simple observation, yet they can offer invaluable contributions to the decision-making process led by humans. By incorporating human judgment into the validation process, we mitigate the risks of potential biases or oversights inherent in purely automated approaches. This emphasis on collective intelligence not only instills a sense of accountability and reliability in the predictive models but also reinforces the idea that technology should complement and augment human capabilities rather than replace them. In educational contexts, this approach promotes a harmonious collaboration that maximizes the strengths of both human and machine intelligence, fostering more informed and responsible decision-making processes.

Author contributions Conceptualization: Okan Bulut, Tarid Wongvorachan, Surina He, and Soo Lee; Methodology: Okan Bulut and Tarid Wongvorachan; Formal analysis and software: Tarid Wongvorachan, Writing—original draft preparation: Okan Bulut, Tarid Wongvorachan, Surina He, and Soo Lee; Writing—review and editing: Okan Bulut, Tarid Wongvorachan, and Surina He, Supervision: Okan Bulut.

Funding The authors did not receive support from any organization for the submitted work.

Data availability The original HSLs:09 dataset used in this study is available through the IES & NCES Datalab: <https://nces.ed.gov/datalab/onlinecodebook>. The pre-processed version of the dataset can be obtained from the corresponding author.

Code availability The Python codes employed in this study are available upon request from the corresponding author.

Declarations

Ethics approval and consent to participate No approval from research ethics committees was deemed necessary for the completion of this study, as the research adhered to a secondary data analysis approach utilizing an existing, publicly available database that contains no personal identifying information.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1. List of utilized HSLs:09 variables

Type	Variable name (variable code)
Continuous	Students' socioeconomic status composite score (X1SES)
	Students' mathematics self-efficacy (X1MTHEFF)
	Students' interest in fall 2009 math course (X1MTHINT)
	Students' perception of science utility (X1SCIUTI)
	Students' science self-efficacy (X1SCIEFF)
	Students' interest in fall 2009 math course (X1SCIINT)
	Students' sense of school belonging (X1SCHOOLBEL)
	Students' school engagement (X1SCHOOLENG)
	Scale of school climate assessment (X1SCHOOLCLI)
	Scale of counselor's perceptions of teacher's expectations (X1COUPERTEA)
	Scale of counselor's perceptions of counselor's expectations (X1COUPERCOU)
	Scale of counselor's perceptions of principal's expectations (X1COUPERPRI)
	Students' GPA in ninth grade (X3TGPA9TH)
Categorical	Hours spent on homework/studying on typical school day (S1HROTHHOMWK)
	Mother's/female guardian's highest level of education (X1MOMEDU)
	Father's/male guardian's highest level of education (X1DADEDU)
	How far in school 9th grader thinks he/she will get (X1STUEDEXPCT)

Type	Variable name (variable code)
	How far in school parent thinks 9th grader will get (X1PAREDEXPCT)
	How often 9th grader goes to class without his/her homework done (S1NOHWDN)
	How often 9th grader goes to class without pencil or paper (S1NOPAPER)
	How often 9th grader goes to class without books (S1NOBOOKS)
	How often 9th grader goes to class late (S1LATE)
	9th grader thinks studying in school rarely pays off later with a good job (S1PAYOFF)
	9th grader thinks even if he/she studies, he/she will not get into college (S1GETINTOCLG)
	9th grader thinks even if he/she studies, family cannot afford college (S1AFFORD)
	9th grader thinks working is more important for him/her than college (S1WORKING)
	9th grader's closest friend gets good grades (S1FRNDGRADES)
	9th grader's closest friend is interested in school (S1FRNDSCHOOL)
	9th grader's closest friend attends classes regularly (S1FRNDCLASS)
	9th grader's closest friend plans to go to college (S1FRNDCLG)
	Hours spent on math homework/studying on typical school day (S1HRMHOMWK)
	Hours spent on science homework/studying on typical school day (S1HRSHOMWK)
	How sure 9th grader is that he/she will graduate from high school (S1SUREHSGRAD)
	How often parent contacted by school about problem behavior (P1BEHAVE)
	How often parent contacted by school about poor attendance (P1ATTEND)
	How often parent contacted by school about poor performance (P1PERFORM)
	Ever dropped out of high school in 2016 (X4EVERDROP)

Appendix 2. Dropout loss of features in the random forest model

Variable name	Dropout loss
baseline	0.398996
X3TGPA9TH	0.366120
X1SES	0.344684
X1MOMEDU	0.336173
X1DADEDU	0.334339
X1MTHINT	0.332290
X1SCIUTI	0.331758
X1SCHOOLCLI	0.331700
_full_model_	0.331461
X1SCIINT	0.331290
X1COUPERCOU	0.331213
X1SCHOOLBEL	0.331177
X1COUPERTEA	0.331015
X1SCHOOLENG	0.330908
X1SCIEFF	0.330805
S1HROTHHOMWK	0.330707
X1MTHEFF	0.330504
X1COUPERPRI	0.330412

References

1. Yeh CYC, Cheng HNH, Chen ZH, Liao CCY, Chan TW. Enhancing achievement and interest in mathematics learning through math-island. *Res Pract Technol Enhanc Learn*. 2019;14(1):5. <https://doi.org/10.1186/s41039-019-0100-9>.

2. Russakovsky O, Li LJ, Fei-Fei L. Best of both worlds: human-machine collaboration for object annotation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 2121–31. <https://doi.org/10.1109/cvpr.2015.7298824>.
3. Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM. Explainable artificial intelligence: an analytical review. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2021;11(5): e1424. <https://doi.org/10.1002/widm.1424>.
4. Gunning D, Aha D. DARPA's explainable artificial intelligence (XAI) program. *AI Mag*. 2019;40(2):44–58. <https://doi.org/10.1145/3301275.3308446>.
5. Bowers AJ. Early warning systems and indicators of dropping out of upper secondary school: the emerging role of digital technologies. In: OECD digital education outlook 2021 pushing the frontiers with artificial intelligence, blockchain and robots. Paris: OECD Publishing; 2021. p. 173.
6. Haesevoets T, De Cremer D, Dierckx K, Van Hiel A. Human-machine collaboration in managerial decision making. *Comput Hum Behav*. 2021;119: 106730. <https://doi.org/10.1016/j.chb.2021.106730>.
7. Xiong W, Wang C, Ma L. Partner or subordinate? Sequential risky decision-making behaviors under human-machine collaboration contexts. *Comput Hum Behav*. 2023;139: 107556. <https://doi.org/10.1016/j.chb.2022.107556>.
8. Knowles JE. Of needles and haystacks: building an accurate statewide dropout early warning system in Wisconsin. *J Educ Data Min*. 2015;7(3):18–67.
9. Wongvorachan T, He S, Bulut O. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information*. 2023;14(1):54. <https://doi.org/10.3390/info14010054>.
10. Allensworth EM, Nagaoka J, Johnson DW. High school graduation and college readiness indicator systems: what we know, what we need to know. concept paper for research and practice. University of Chicago Consortium on School Research; 2018.
11. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. 2016. [arXiv:1606.05718](http://arxiv.org/abs/1606.05718). <http://arxiv.org/abs/1606.05718>.
12. Xiong W, Fan H, Ma L, Wang C. Challenges of human-machine collaboration in risky decision-making. *Front Eng Manag*. 2022;9(1):89–103. <https://doi.org/10.1007/s42524-021-0182-0>.
13. Van Buuren S. Flexible imputation of missing data. 2nd ed. Boca Raton: CRC Press; 2018.
14. Padoy N, Hager GD. Human-machine collaborative surgery using learned models. In: 2011 IEEE international conference on robotics and automation. 2011. p. 5285–92. <https://doi.org/10.1109/icra.2011.5980250>.
15. He H, Ma Y, editors. Imbalanced learning: foundations, algorithms, and applications. Hoboken: Wiley; 2013.
16. Melo E, Silva I, Costa DG, Viegas CM, Barros TM. On the use of explainable artificial intelligence to evaluate school dropout. *Educ Sci*. 2022;12(12):845. <https://doi.org/10.3390/educsci12120845>.
17. Wilson HJ, Daugherty PR. Collaborative intelligence: humans and AI are joining forces. *Harv Bus Rev*. 2018;96(4):114–23.
18. Khan O, Badhiwala JH, Grasso G, Fehlings MG. Use of machine learning and artificial intelligence to drive personalized medicine approaches for spine care. *World Neurosurg*. 2020;140:512–8. <https://doi.org/10.1016/j.wneu.2020.04.022>.
19. Paleja R, Ghuy M, Ranawaka Arachchige N, Jensen R, Gombolay M. The utility of explainable AI in ad hoc human-machine teaming. *Adv Neural Inf Process Syst*. 2021;34:610–23.
20. Pasquale F. The black box society: the secret algorithms that control money and information. Cambridge: Harvard University Press; 2015. <https://doi.org/10.4159/harvard.9780674736061>.
21. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206–15. <https://doi.org/10.1038/s42256-019-0048-x>.
22. Nagy M, Molontay R. Interpretable dropout prediction: towards XAI-based personalized intervention. *Int J Artif Intell Educ*. 2023. <https://doi.org/10.1007/s40593-023-00331-8>.
23. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 427–36. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Nguyen_Deep_Neural_Networks_2015_CVPR_paper.html.
24. Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. *Artif Intell Rev*. 2022;55(5):3503–68. <https://doi.org/10.1007/s10462-021-10088-y>.
25. Kozak A. Basic XAI with DALEX. Responsible ML having fun while building responsible ML models. 2020, October 18. <https://medium.com/responsibleml/basic-xai-with-dalex-part-1-introduction-e68f65fa2889>.
26. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020;20(1):310. <https://doi.org/10.1186/s12911-020-01332-6>.
27. Chollet F. Keras (3.0.1) [Python]. 2015. <https://keras.io>.
28. Sara NB, Halland R, Igel C, et al. High-school dropout prediction using machine learning: a Danish large-scale study. 23rd ESANN 2015 proceedings. 2015. p. 319–24.
29. Chung JY, Lee S. Dropout early warning systems for high school students using machine learning. *Child Youth Serv Rev*. 2019;96:346–53. <https://doi.org/10.1016/j.childyouth.2018.11.030>.
30. Sansone D. Beyond early warning indicators: high school dropout and machine learning. *Oxf Bull Econ Stat*. 2019;81(2):456–85. <https://doi.org/10.1111/obes.12277>.
31. Krüger JGC, de Souza Britto A Jr, Barddal JP. An explainable machine learning approach for student dropout prediction. *Expert Syst Appl*. 2023;233: 120933. <https://doi.org/10.1016/j.eswa.2023.120933>.
32. National Center for Educational Statistics [NCES]. High school longitudinal study of 2009. National Center for Educational Statistics [NCES]; 2016. <https://nces.ed.gov/surveys/hsls09/>.
33. Ben-Avie M, Darrow B Jr. Malleable and immutable student characteristics: incoming profiles and experiences on campus. *J Assess Inst Eff*. 2018;8(1–2):22–50. <https://doi.org/10.5325/jasseinsteffe.8.1-2.0022>.
34. Baniecki H, Kretowicz W, Piatyszek P, Wisniewski J, Biecek P. dalex: responsible machine learning with interactive explainability and fairness in Python. *J Mach Learn Res*. 2021;22(214):1–7.
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30. <https://doi.org/10.5555/1953048.2078195>.

36. Moolayil J. An introduction to deep learning and Keras. In: *Learn Keras for deep neural networks: a fast-track approach to modern deep learning with Python*. New York: Springer; 2019. p. 1–16.
37. Gianfagna L, Di Cecco A. *Explainable AI with python*. Cham: Springer; 2021.
38. Cadario R, Longoni C, Morewedge CK. Understanding, explaining, and utilizing medical artificial intelligence. *Nat Hum Behav*. 2021;5(12):1636–42. <https://doi.org/10.1038/s41562-021-01146-0>.
39. Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?” Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016. p. 1135–44. <https://doi.org/10.1145/2939672.2939778>.
40. Roßbach P. *Neural networks vs. random forests—Does it always have to be deep learning?* Germany: Frankfurt School of Finance and Management; 2018. <https://blog.frankfurt-school.de/wp-content/uploads/2018/10/Neural-Networks-vs-Random-Forests.pdf>.
41. Wang S, Aggarwal C, Liu H. Using a random forest to inspire a neural network and improving on it. In: *Proceedings of the 2017 SIAM international conference on data mining*. 2017. p. 1–9. <https://doi.org/10.1137/1.9781611974973.1>.
42. Bhutoria A. Personalized education and artificial intelligence in the United States, China, and India: a systematic review using a human-in-the-loop model. *Comput Educ Artif Intell*. 2022;23: 100068. <https://doi.org/10.1016/j.caeai.2022.100068>.
43. Dawson V, Carso K. Using climate change scenarios to assess high school students’ argumentation skills. *Res Sci Technol Educ*. 2017;35(1):1–16. <https://doi.org/10.1080/02635143.2016.1174932>.
44. Fong Lam U, Chen WW, Zhang J, Liang T. It feels good to learn where I belong: school belonging, academic emotions, and academic achievement in adolescents. *Sch Psychol Int*. 2015;36(4):393–409. <https://doi.org/10.1177/0143034315589649>.
45. Epping KA. The impact of parental involvement on student’s academic achievement, parental well-being, and parent–teacher relationships (Master’s thesis, University of Calgary, Calgary, Canada). <http://hdl.handle.net/1880/108726>.
46. Ahmadi S, Hassani M, Ahmadi F. Student- and school-level factors related to school belongingness among high school students. *Int J Adolesc Youth*. 2020;25(1):741–52. <https://doi.org/10.1080/02673843.2020.1730200>.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.