Evaluating Classification Machine Learning Models for Identifying External Factors Influencing Student Choices in Virtual Learning Environments

Bagher Javadi[1,*], Suwimon Kooptiwoot[2], Chaisri Tharasawatpipat[2], Sivapan Choo-in[2], Pantip Kayee[2], and Duongdearn Suwanjinda[3]

[1]Department of Sciences, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand
[2]Department of Applied Sciences, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand
[3]School of Educational Studies, Sukhothai Thammathirat Open University, Nonthaburi, Thailand
[*]Corresponding author: javadi.ba@ssru.ac.th (B.J.)
Manuscript received May 22, 2025; revised June 13, 2025; accepted September 1, 2025; published January 20, 2026

***Abstract*—The integration of machine learning techniques into the educational landscape has opened new avenues for analyzing and improving learning experiences. This research investigates the predictive capability of classification algorithms in determining how external conditions affect student preferences for online learning. By analyzing variables such as internet Disruption, device availability, and psychological stress during the COVID-19 pandemic, we developed several classification models to uncover the patterns driving these preferences. The study applied a range of supervised learning algorithms—namely, random forest, logistic regression, gradient boosting, support vector machines, k-nearest neighbors, naïve Bayes, and decision trees—to identify the most accurate predictive approach. Among the evaluated classification algorithms, K-Nearest Neighbors achieved the highest accuracy (0.798) and F1-score (0.883), with strong recall (0.979). Support Vector Machine obtained the highest recall (1.000) but had a lower ROC-AUC (0.409). Logistic Regression, Naive Bayes, Random Forest, and Gradient Boosting showed balanced performance, with F1-scores ranging from 0.844 to 0.860. Decision Tree yielded the lowest accuracy (0.712) but maintained competitive precision (0.824). Overall, K-Nearest Neighbors and Support Vector Machine demonstrated superior recall, while K-Nearest Neighbors provided the best overall classification performance. To gain deeper interpretability of feature contributions, we employed SHAP (SHapley Additive exPlanations), which highlighted stress as the most influential factor. The findings offer actionable insights into how non-academic influences shape learning modality choices, supporting data-driven strategies to adapt online education to diverse student needs during crisis conditions and beyond.**

***Keywords*—classification algorithms, virtual education, learner decision-making, external influences, predictive modeling, Learning Management Systems (LMS) platforms, educational data analytics**

I. Introduction

As machine learning and data analytics continue to revolutionize numerous industries, the education sector is also increasingly embracing these technologies in order to boost student engagement and academic outcomes [1–3]. A foundational step in improving educational systems involves the accurate identification of influential variables that shape learning effectiveness. Analytical modeling plays a key role in uncovering such variables and informing strategic interventions [4, 5].

Robust predictive algorithms and advanced data mining techniques are essential to capturing subtle patterns within complex educational datasets [6]. When applied effectively, these tools can extract significant behavioral insights, identify hidden factors, and support the development of adaptive learning environments [7]. Moreover, predictive modeling can play a proactive role in early detection of students at risk of underperforming, allowing for timely pedagogical adjustments to support learning progression [8].

The outbreak of COVID-19 posed unprecedented global challenges, triggering health emergencies, economic instability, and significant disruptions to daily life [9, 10]. In response, governments enforced lockdowns and social distancing policies, which, while necessary for public health, contributed to increased stress and social disruption [11]. The educational domain was notably affected, with traditional classroom instruction rapidly replaced by online learning environments [12–14].

Although this transition posed multiple obstacles, it also accelerated the adoption of web-based virtual learning environments such as Coursera, and Khan Academy which became central to remote learning [15, 16]. Still, the shift highlighted critical limitations of virtual instruction and emphasized the urgent need for more effective and inclusive learning frameworks [17].

This study addresses these challenges by employing various classification-based machine learning models to investigate how external conditions—such as technology access and psychological stress—affect students' preferences for online education. By analyzing these factors during the pandemic, the research seeks to uncover actionable insights that can improves the design and delivery of distance learning systems [18, 19]. A variety of machine learning classification techniques were utilized in this investigation to unravel the complex interplay between diverse external variables that shape students' preferences for remote education. This strategy allows for a nuanced analysis of the ways in which elements such as psychological stress, ease of use of learning platforms, and the ability to adapt to digital environments collectively impact decision-making among learners. Through this analytical lens, the study addresses existing knowledge gaps and contributes practical insights aimed at improving student support systems during crisis-driven educational transitions.

Among the algorithms applied, both regression trees and classification-based models offer distinct analytical capabilities suited for different types of educational data. Regression decision trees, for instance, are designed to forecast continuous values by iteratively dividing data into

doi: 10.18178/ijiet.2026.16.1.2501

subsets that minimize internal variability. A practical application of this model might involve estimating the number of weekly study hours a student devotes based on demographic and infrastructural variables like age, internet access, and academic level [20]. In contrast, classification models are adept at handling categorical outcomes—such as determining whether a student is likely to prefer online learning or not. Logistic regression, a foundational classification method, estimates class membership probabilities via linear predictors. Ensemble methods such as Random Forest and Gradient Boosting build on this by integrating numerous decision trees to improve predictive reliability and reduce overfitting. Meanwhile, Support Vector Machines (SVM) operate by identifying the best separating boundary between classes in high-dimensional space. K-Nearest Neighbors (KNN) bases its classification decisions on the proximity of data points to previously labeled samples, and the Naïve Bayes classifier offers a probabilistic approach that assumes conditional independence between predictors for faster computations. To facilitate this analysis, a structured survey was conducted, gathering data on multiple independent features, while student receptivity to online learning served as the primary binary outcome variable—categorized as either high or low. Regression models were employed to examine continuous outcomes like time spent on virtual study sessions, whereas classification models such as SVM, Logistic Regression, and Random Forest were used to identify factors contributing to categorical differences in learning adaptability. This dual approach underscored the complementary strengths of regression and classification frameworks in capturing the multifaceted nature of student learning behaviors.

The overarching goal of this study is to design and compare predictive frameworks using various supervised learning algorithms, with an emphasis on determining which model offers the highest accuracy and interpretability. Furthermore, the research explores how each external feature contributes to model outputs, thereby offering deeper insight into how extrinsic conditions influence learning preferences in digitally mediated environments.

II. Research Methodology

To collect data on student preferences regarding Online versus Onsite Learning (OL/OS), a structured questionnaire was developed and distributed using Google Forms. This platform enabled efficient data gathering, incorporating various extrinsic factors influencing students' learning mode preferences. Once collected, the responses were exported in CSV format and imported into a Google Colab environment for in-depth analysis. Google Colab, a collaborative and interactive coding platform, facilitated data preprocessing, feature engineering, model training, and evaluation, supporting the implementation of machine learning algorithms tailored to binary classification tasks. This methodological framework allowed for the identification of critical predictors influencing students' choices between online and traditional classroom learning, offering strategic insights for educational stakeholders.

The survey, conducted in February 2024, yielded a dataset of 120 responses from undergraduate students. Each instance comprised a binary target variable (OL/OS) and multiple independent attributes, namely Internet Disruption (ID), Cognitive Demands (CD), Learning Platform (LP), Illness (I), Related Illness (RI), Vaccination (V), Infection Severity (IS), Environmental Sensitivity (ES), and Protective Adaptability (PA). All responses were numerically encoded to facilitate computational analysis. To remove scale-related biases and enhance model stability, all features were normalized using standard scaling, transforming them to have zero mean and unit variance. Model training and evaluation were first performed using a conventional 70/30 train–test split, followed by 5-fold and 10-fold cross-validation for each classification algorithm. The use of cross-validation mitigated overfitting and improved generalization performance, which was particularly critical given the relatively small sample size. we employed 5 and 10-fold cross-validation, a widely accepted resampling method that partitions the dataset into training and validation folds, ensuring that all instances contribute to both training and evaluation [21]. This approach provides more reliable performance metrics than a single split, particularly for limited datasets, by reducing variance in model evaluation [22]. Cross-validation is especially valuable in educational research, where collecting large student datasets can be challenging, thus enhancing generalizability of findings [23].

Data preprocessing steps included cleaning, transformation, and feature scaling, followed by feature selection tailored to the model requirements. The analysis adopted a supervised learning approach, as the target variable was known and binary in nature. A total of seven machine learning classification algorithms were employed to construct predictive models and evaluate their performance: Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Regression Trees. Each model was trained to learn patterns from the feature set and predict students' OL/OS preference.

Logistic Regression was used to model the probability of online learning preference based on combinations of input features. Random Forest, an ensemble method, constructed multiple decision trees and utilized majority voting to enhance prediction accuracy. Gradient Boosting followed a sequential learning approach, where successive models corrected the errors of their predecessors using gradient descent to optimize performance iteratively. SVM aimed to identify the hyperplane that best separated the binary classes with maximum margin, while KNN classified data points based on the dominant class among their nearest neighbors. Naive Bayes, grounded in Bayes' theorem and the assumption of feature independence, was applied for its computational efficiency and robustness to noisy data. Regression Trees segmented the dataset into homogenous regions through recursive binary splits, enabling the capture of non-linear patterns.

To enhance model explainability, SHAP (SHapley Additive exPlanations) values were calculated. SHAP offers a game-theoretic approach to interpret model outputs, assigning importance scores to each feature by quantifying its marginal contribution to the prediction. This facilitated a transparent understanding of which factors most significantly influenced students' preferences, thus aiding the interpretation of complex model behavior.

Model evaluation was performed using four key performance metrics: accuracy, precision, recall, and F1-score. Accuracy provided a general measure of correctness, while precision evaluated the proportion of correct positive predictions. Recall measured the ability to identify all actual positive cases, and the F1-score served as a balanced metric combining both precision and recall. The dataset was divided into training and testing subsets, and each model was trained on the training data and validated on the test set. Performance metrics were computed using functions from the Scikit-learn library, ensuring consistency and reliability across all evaluations.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F - \text{measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The combined use of classification models, and SHAP analysis enabled a comprehensive exploration of student behavior in the context of learning modality selection. This integrative methodology not only offered high predictive performance but also supported interpretability, aligning with the broader goal of generating actionable insights for educational improvement. As highlighted in earlier studies [24–28], the use of machine learning in education research has proven to be effective in identifying key influencing variables, and this study extends that capability with a focus on simplicity, transparency, and practical relevance (Fig. 1).

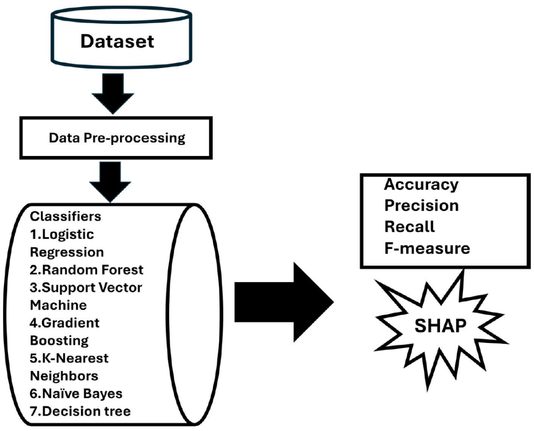

Fig. 1. This diagram represents a machine learning pipeline for evaluating online education adaptability. The dataset undergoes preprocessing, followed by feature importance analysis using SHAP and classification using various algorithms like Logistic Regression and Random Forest. The models are evaluated using performance metrics such as accuracy, precision, recall, and F-measure.

III. Results

The F-Value feature offers a statistical measure derived from the F-test, commonly used to evaluate the significance of individual predictors in a model. This metric evaluates whether there is a substantial difference in the averages across two or more groups. In feature selection, the F-Value measures how well each feature distinguishes between categories or accounts for variation in the data. A larger F-Value suggests a more significant connection between the features—namely Vaccination (V), Learning Platform (LP), and Internet Disruption (ID)—and the dependent variable (Online/Onsite, or ON/OS), thus implying greater predictive power. This metric proves particularly valuable in identifying influential predictors during model development, especially in scenarios involving Analysis of Variance (ANOVA) and regression analysis (Fig. 2).

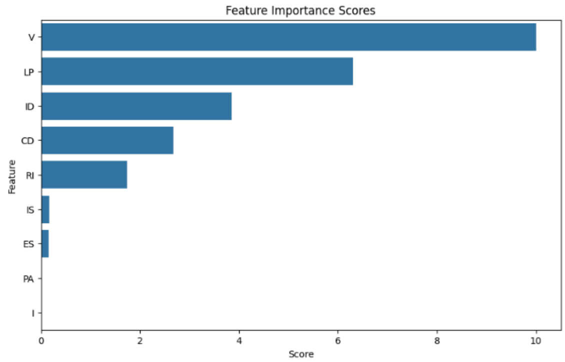

Fig. 2. This bar chart illustrates the F-value scores used to assess the importance of input variables in predicting online learning adaptability during the COVID-19 pandemic. Vaccination (V) is identified as the most influential factor, followed by Learning Platform (LP), Internet Disruption (ID), and Cognitive Demands (CD), all of which significantly impact model performance. In contrast, variables such as Infection Severity (IS), Environmental Sensitivity (ES), Protective Adaptability (PA), and Illness (I) show relatively low contribution to the predictive outcome.

Using a single train–test split (70/30), the Decision Tree classifier achieved the highest accuracy (0.583), demonstrating relatively better predictive capability in this evaluation setting. The Support Vector Machine (SVM) obtained a perfect recall score (1.0), indicating its ability to correctly identify all positive instances without false negatives. Gradient Boosting achieved the highest precision (0.550), reflecting its reduced false-positive rate. Furthermore, the SVM attained the highest F1-score (0.703), suggesting a strong balance between precision and recall and positioning it as the most balanced model in this evaluation. In contrast, the K-Nearest Neighbors (KNN) algorithm recorded the lowest accuracy (0.458), though it maintained moderate recall and F1-score performance. To further assess model robustness and mitigate potential biases from a single data split, all classifiers were subsequently evaluated using 5-fold and 10-fold cross-validation.

These findings highlight the trade-offs involved in selecting classifiers. While accuracy measures overall correctness, it may obscure important performance nuances in imbalanced datasets or cost-sensitive applications. For example, although the SVM attained a high F1-score due to its perfect recall, its relatively lower precision suggests an increased risk of false positives. The Decision Tree, with its higher accuracy, appears to be a solid choice but carries a moderate risk of misclassification as indicated by its precision. Random Forest, Gradient Boosting, and Naive Bayes showed comparable performance across the majority of measures, showcasing their adaptability. KNN, despite its lower accuracy, retained a respectable recall rate, suggesting utility in detecting relevant instances despite a higher misclassification rate. Ultimately, classifier selection must be guided by specific application goals and tolerance for different types of errors.

Table 1. Performance of seven supervised classification algorithms in predicting student preferences for online learning using 10-fold cross-validation. K-Nearest Neighbors achieved the highest accuracy and F1-score, while Support Vector Machine showed perfect recall but lower probability discrimination (ROC-AUC)

Classifier	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.757	0.781	0.958	0.859	0.561
Decision Tree	0.712	0.824	0.804	0.809	0.603
Random Forest	0.746	0.809	0.891	0.844	0.586
Gradient Boosting	0.747	0.807	0.892	0.844	0.586
Support Vector Machine	0.782	0.782	1.000	0.877	0.409
K-Nearest Neighbors	0.798	0.807	0.979	0.883	0.517
Naive Bayes	0.764	0.799	0.934	0.860	0.575

To ensure robust and unbiased model evaluation, we compared three performance estimation methods: a single 70/30 train-test split, 5-fold cross-validation, and 10-fold cross-validation. The single train-test split, though commonly used, may suffer from random sampling bias, especially with smaller datasets. In contrast, cross-validation distributes the data across multiple iterations, producing more stable and reliable estimates of classifier performance. As expected, both 5-fold and 10-fold cross-validation yielded higher and more consistent performance metrics compared to the single split. The 5-fold cross-validation results showed substantial improvements in accuracy and F1-score across nearly all classifiers. For instance, Random Forest improved from 0.50 accuracy (split) to 0.722 (5-fold), while Naive Bayes increased its F1-score from 0.647 to 0.855. These gains suggest that some models may have been underestimated using a single train-test partition. The 10-fold validation, while computationally more intensive, offered marginal yet consistent boosts for several models. K-Nearest Neighbors, in particular, benefited the most from increased folds, reaching the highest F1-score of 0.883 and accuracy of 0.798 under 10-fold CV. While 10-fold cross-validation provides slightly better average metrics, 5-fold also offers a reliable and efficient evaluation standard, especially for small datasets. Given the consistent trends across both cross-validation strategies, we report the full 10-fold CV results in Table 1 to highlight the best-performing models. For transparency, we also compare outcomes from the train-test split and 5-fold CV in this section. This comprehensive reporting allows for a more nuanced understanding of classifier behavior and demonstrates the robustness of the findings across multiple evaluation techniques.

The ROC curves indicate that while all classifiers perform slightly better than random guessing in distinguishing between students who prefer online vs. onsite learning, their probabilistic discrimination power is modest. Gradient Boosting showed the highest average AUC (0.61 ± 0.13), suggesting a limited but consistent ability to rank predictions. The relatively large standard deviations in AUC scores, particularly for Random Forest and SVM, highlight variability in model performance across different data splits, which can be attributed to the dataset's small size and class imbalance. These results emphasize the need for larger and more balanced datasets in future research to improve classification reliability and calibration (Fig. 3).

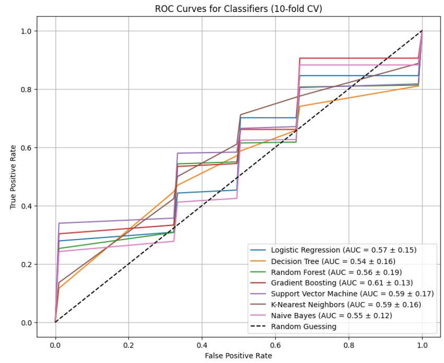

Fig. 3. ROC curves of classification models evaluated using 10-fold cross-validation. The curves plot the true positive rate (TPR) against the false positive rate (FPR), with the diagonal line representing random guessing (AUC = 0.5). Gradient Boosting and K-Nearest Neighbors show the best area under the curve (AUC), indicating moderate probabilistic discrimination. Standard deviations reflect model variability across folds.

In a Random Forest model, feature importance is generally determined by assessing how much a feature contributes to reducing impurity, like Gini impurity or entropy, during the data splitting process in the decision trees of the ensemble. A greater reduction in impurity indicates a more impactful feature. Visualizing feature importance helps identify which features exert the greatest influence on predictions, offering crucial insights for model refinement (Fig. 4).

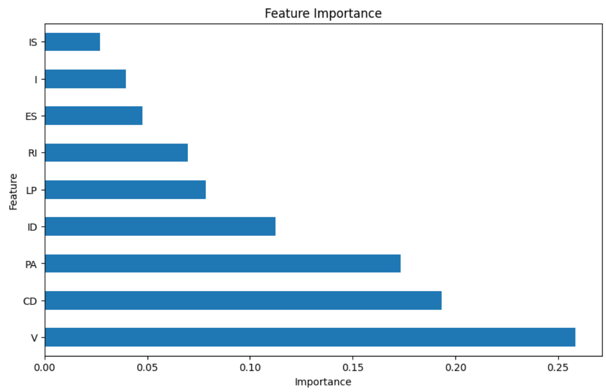

Fig. 4. This bar chart presents feature importance scores from the Random Forest model in assessing online learning adaptability during COVID-19. Vaccination (V), Cognitive Demands (CD), and Protective Adaptability (PA) emerged as the top predictors, contributing significantly to the model's performance. Features like Infection Severity (IS) and Illness (I) showed the least impact, indicating lower relevance in prediction.

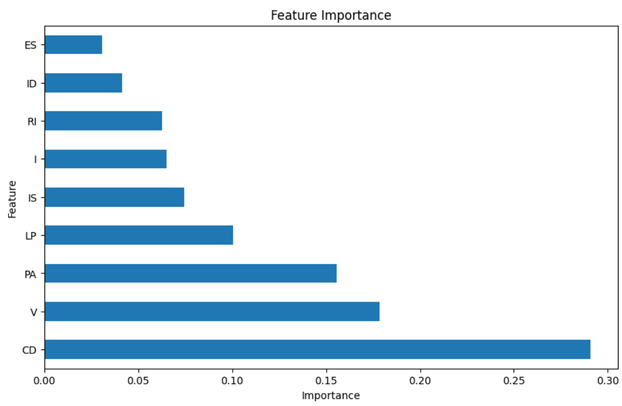

Fig. 5. Feature importance scores from the Decision Tree model. This bar chart illustrates feature importance scores from the Decision Tree model for evaluating online learning adaptability during COVID-19. The most influential factors were Cognitive Demands (CD), Vaccination (V), and Protective Adaptability (PA). Less impactful variables included Infection Sensitivity (ES), Internet Disruption (ID), Related Illness (RI), suggesting minimal predictive power.

Similarly, in a Decision Tree model, feature importance is determined by assessing the reduction in variance or "node purity" achieved through splits. A graphical representation of feature importance highlights which variables contribute most significantly to reducing model error, guiding feature selection and improving model interpretability and accuracy (Fig. 5).

The construction of a Random Forest involves generating multiple Decision Trees, each trained on a random subset of both the data and the features. This randomness helps mitigate overfitting and enhances generalization. The first tree in the forest reflects the model's initial attempt to learn from the data, influenced by randomly selected subsets. The final tree, built after extensive ensemble learning, represents a more refined structure, having benefited from the accumulated learning across trees. The contrast between these trees illustrates the power of ensemble methods to average out individual biases and enhance robustness (Figs. 6 and 7).

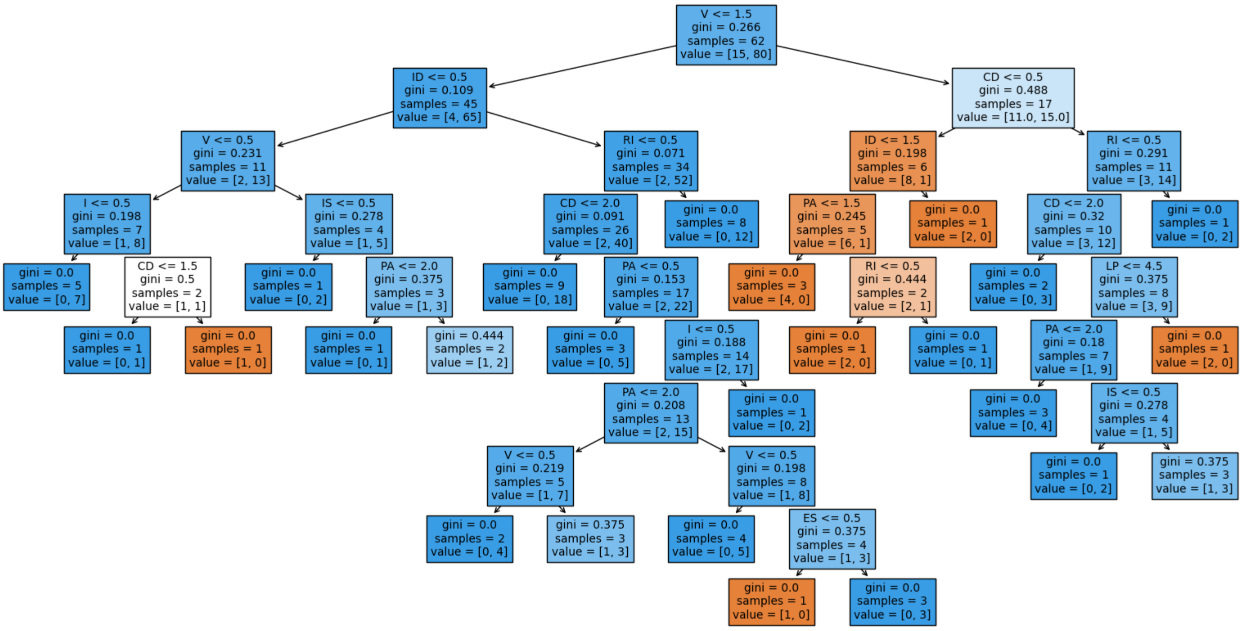

Fig. 6. To build the first tree in a Random Forest, a random subset of the training data is selected with replacement (bootstrap sampling), and at each node, a random subset of features is chosen to determine the best split. This randomness introduces diversity among trees, enhancing the model's overall robustness and reducing overfitting.

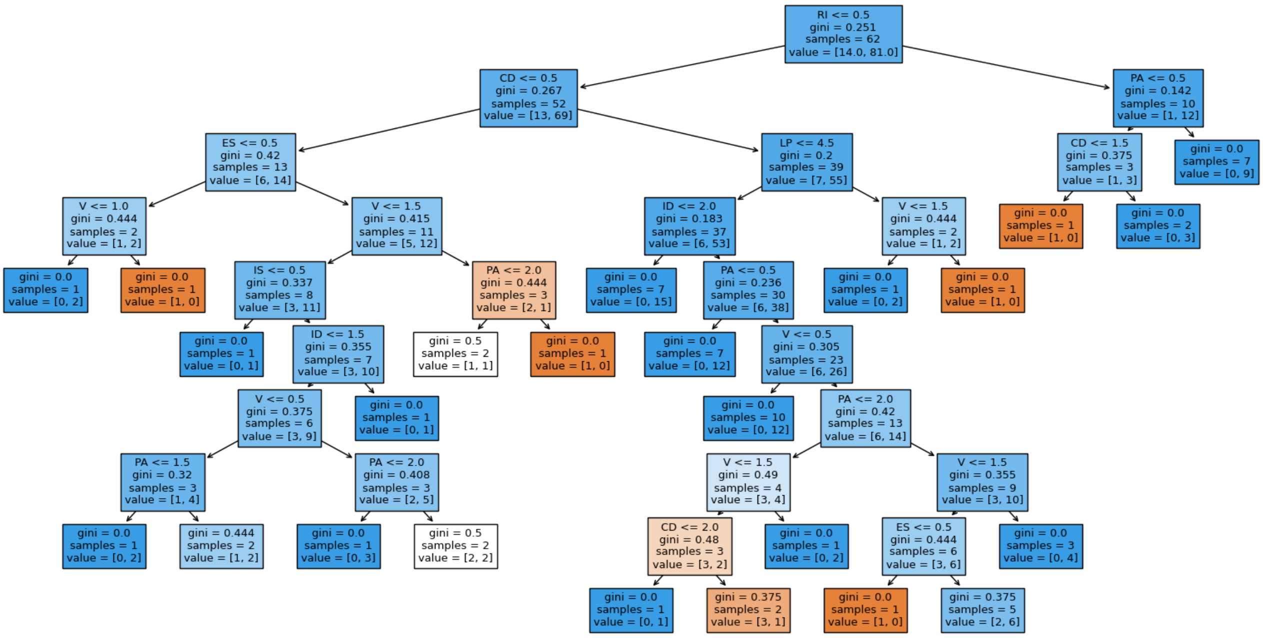

Fig. 7. The final tree in the Random Forest model is one of many decision trees trained on different random subsets of data and features. While each tree may vary in structure and predictions, the final output of the Random Forest is based on the aggregated results—typically using majority voting (classification)—from all trees including this last one.

A classification tree segments the data based on decisions at internal nodes, using metrics such as Gini impurity to select optimal splits. Each path from the root to a leaf node represents a rule derived from the input features. Classification trees are valued for their interpretability and efficacy in modeling complex, non-linear relationships (Fig. 8).

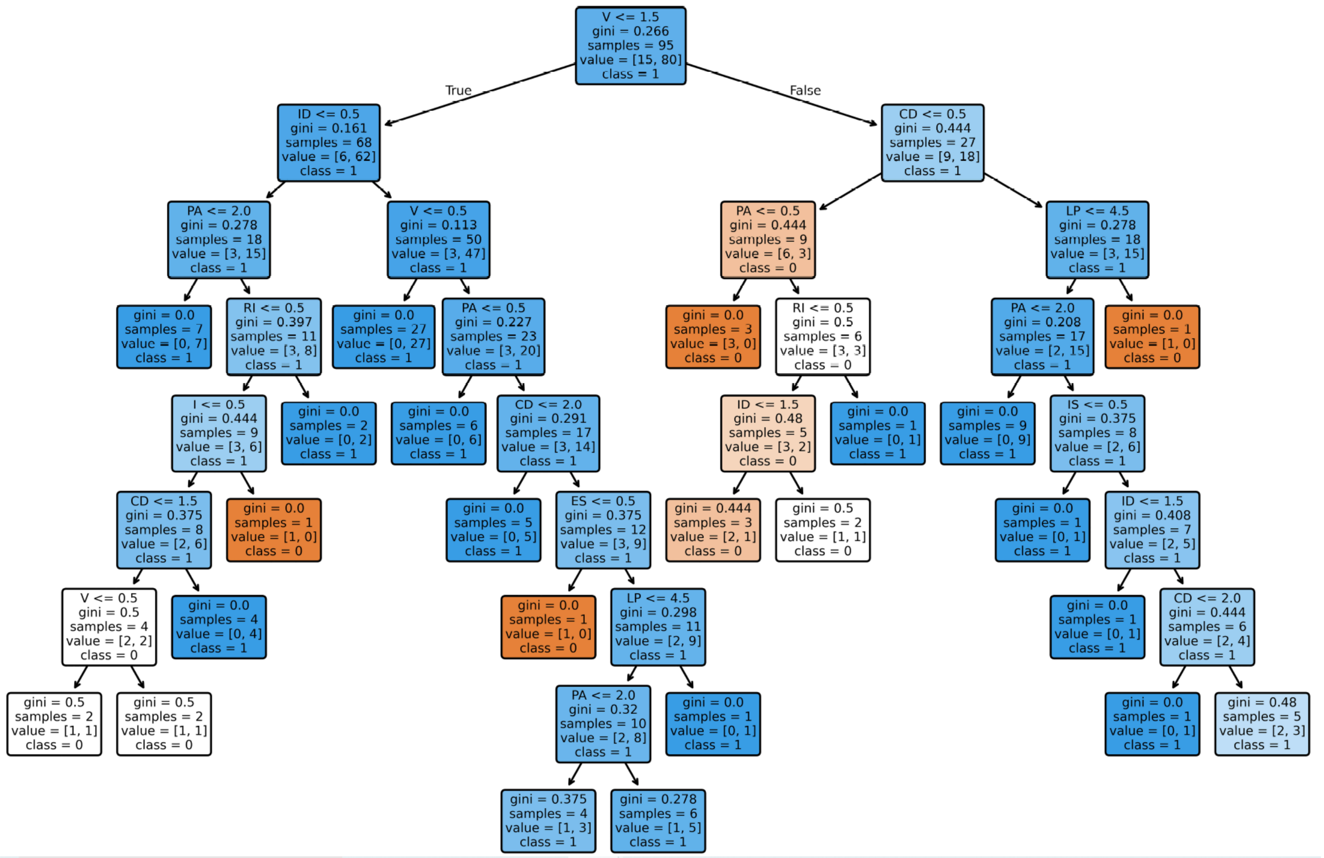

Fig. 8. A decision tree is a machine learning model that splits data into subsets based on feature values, resulting in a tree-like structure where each leaf node represents a classification outcome. It is easy to interpret and visualize the complex data. However (Fig. 7) random Forest differs from a decision tree by using an ensemble of multiple decision trees, trained on random subsets of the data and features, improving accuracy and generalization while reducing overfitting.

SHAP (SHapley Additive exPlanations) is a powerful method to explain individual predictions by assigning each feature a contribution value based on cooperative game theory. SHAP values quantify whether a feature has a positive or negative effect on a model's prediction and to what extent. The SHAP summary plot presents a visual overview of feature impacts across all samples. It displays features on the y-axis and SHAP values on the x-axis, where the color denotes the feature's actual value. In this study, variables such as Vaccination (V) and Cognitive Demands (CD) were identified as having significant effects on prediction outcomes (Fig. 9).

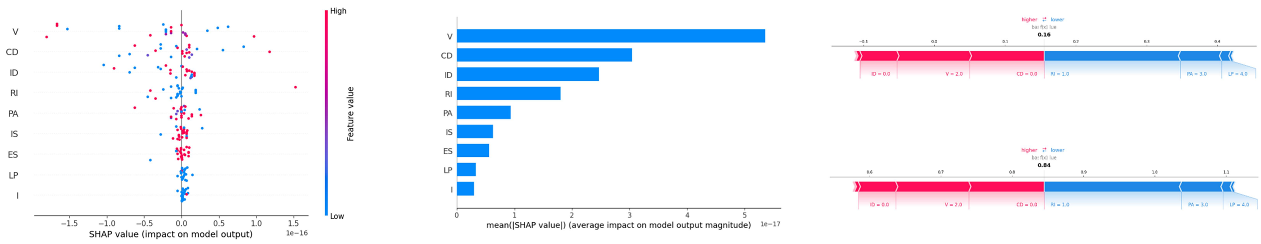

Fig. 9. This image displays SHAP (SHapley Additive exPlanations) results: The left and middle plots show feature importance and their average impact on model output, highlighting those features like Vaccination (V) and Cognitive Demands (CD) contribute most to prediction variation. The right side shows force plots for individual predictions, illustrating how each feature pushes the prediction toward higher or lower values.

To improve interpretability, we present a SHAP summary plot (left) and individual SHAP force plots (right), which show how specific values (high stress, low internet access) contribute to the final prediction. For instance, high vaccination confidence generally pushes predictions toward online preference, whereas internet disruption pushes predictions toward onsite.

A survey conducted during the study revealed that 90 students preferred online learning as their primary mode of education (Fig. 10). Feature importance analysis across models indicated that vaccination status had a significant influence on this preference. Most students favoring online learning were vaccinated. However, students who did not prefer online learning often cited unstable internet connectivity (Internet Disruption) as a major challenge. These students also experienced obstacles such as lack of internet access at home and familial responsibilities, which discouraged participation in online learning.

These results emphasize the need to consider extrinsic variables—particularly Internet Disruption and psychological stress—when evaluating online learning environments. While previous models often prioritized intrinsic learner characteristics, this study demonstrates that external conditions play a pivotal role in shaping learning preferences. These insights can inform the development of technology acceptance models and classroom management strategies aimed at improving the online learning experience under uncertain or challenging conditions.

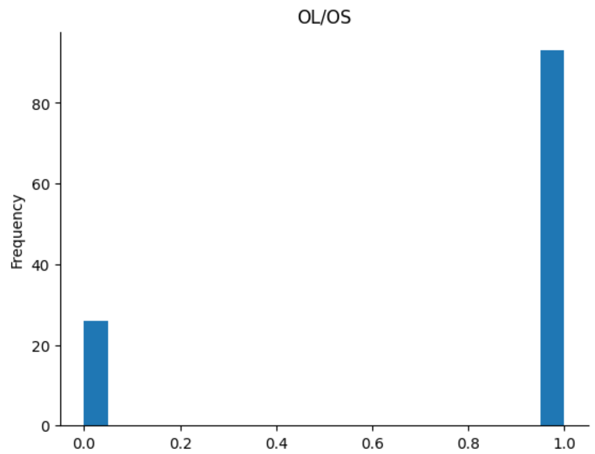

Fig. 10. This bar chart shows the distribution of student preferences between online (1) and offline (0) learning modes. The majority of students (90) prefer online learning, while a smaller group (25) prefers offline.

IV. Discussion

In this research, seven machine learning techniques — Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, Naive Bayes, Support Vector Machines, and K-Nearest Neighbors— implemented to build predictive models for student adaptability to online learning. These algorithms were chosen due to their demonstrated effectiveness across diverse classification problems. We evaluated each model using standard performance metrics: accuracy, precision, recall, and F1-score [6].

Previous research on student performance has largely focused on intrinsic factors. Nghe and colleagues [29] compared Bayesian Networks and Decision Trees for predicting academic outcomes, finding that Decision Trees generally outperformed Bayesian Networks. Similarly, Cortez and Silva [30] predicted secondary school grades using classification models, again showing superior performance from tree-based methods. Other researchers, including Mayilvaganan and Kalpanadevi [31, 32], explored similar models to predict outcomes such as academic performance and student retention. Lykourentzou *et al*. [33] extended this work by integrating data from Learning Management Systems (LMS) to predict early dropout rates.

To ensure robust and unbiased performance assessment, this study compared three evaluation strategies: a single 70/30 train-test split, 5-fold cross-validation, and 10-fold cross-validation. The traditional train-test split, while commonly used, may be sensitive to random variation—especially in small datasets—and may not reliably reflect real-world performance. Cross-validation, in contrast, divides the data into multiple folds, allowing every data point to serve as both training and test data across different iterations. This provides a more stable and generalizable estimate of model effectiveness. The results clearly demonstrated that cross-validation produced consistently higher performance metrics across all models compared to the single split. For example, the accuracy of Random Forest increased from 0.50 (split) to 0.746 (10-fold CV), while K-Nearest Neighbors improved from 0.458 to 0.798. These improvements indicate that train-test splits may underestimate model capabilities due to class imbalance or unlucky partitions, while cross-validation helps mitigate this risk by providing a fairer distribution of classes across training and testing sets. Among the classifiers tested, K-Nearest Neighbors emerged as the top performer under 10-fold cross-validation, achieving the highest accuracy (0.798) and F1-score (0.883). Support Vector Machine also performed strongly, particularly in terms of recall (1.000), making it effective at identifying students who prefer online learning. However, its relatively low precision and ROC-AUC suggest a higher false positive rate and weaker confidence calibration. Gradient Boosting and Random Forest both demonstrated a balanced trade-off between precision and recall, each attaining F1-scores of 0.844, making them suitable choices when model interpretability and robustness are both valued. Naive Bayes, though based on strong independence assumptions, showed competitive performance, indicating that even simpler models can capture meaningful patterns in student behavior when appropriately tuned and evaluated. These results underscore the value of using multiple metrics—accuracy, F1-score, and recall—to assess classifier effectiveness comprehensively, particularly when dealing with imbalanced or limited educational datasets. ROC curve analysis was conducted to evaluate each model's ability to rank predictions based on confidence scores. The area under the ROC curve (AUC) provides a measure of this capability, with values closer to 1.0 indicating superior class separation. In this study, AUC values for all models ranged from 0.54 to 0.61, only modestly above the 0.50 baseline that represents random guessing. Gradient Boosting achieved the highest AUC (0.61 ± 0.13), followed by K-Nearest Neighbors and Support Vector Machine (both at 0.59), but all models exhibited relatively high standard deviation across folds. This variability reflects the small sample size (n = 120) and the binary nature of the classification task, which limits the models' probabilistic discrimination. Notably, while SVM achieved perfect recall, its AUC was lower than expected, suggesting that it produces confident predictions without consistently ranking them well across true and false positives. These findings indicate that although some models can predict outcomes reliably, their ability to express calibrated prediction probabilities remains limited—a point that should be addressed in future work using larger, more diverse datasets.

Although the AUC values in this study are relatively low (ranging from 0.409 to 0.603), this does not invalidate the applicability of the models. In small and imbalanced educational datasets, AUC can underestimate model utility, particularly when the primary goal is classification rather than probability estimation [34]. Prior studies have shown that models with modest AUC can still provide actionable insights when recall, precision, or F1-scores are strong, especially in high-stakes decision-making contexts where correctly identifying positive cases is critical [35, 36]. In this research, metrics such as recall and F1-score consistently demonstrated high performance, underscoring that low AUC values primarily reflect limitations in probabilistic discrimination rather than outright ineffectiveness of the classifiers. Therefore, while AUC highlights an area for improvement, the findings remain meaningful for guiding online learning strategies.

Classification, a supervised learning approach, trains models on labeled data to predict classes of unseen instances [37]. It involves two phases: training, where patterns are learned from labeled data, and testing, where performance is evaluated using a confusion matrix of true/false positives and negatives to derive metrics such as accuracy, precision, recall, and F1-score [38, 39]. The COVID-19 pandemic accelerated the adoption of online education, demanding greater digital literacy and adaptability to online platforms [40]. This shift also created challenges for instructors and learners, including complex workflows and technological barriers [41]. Our study emphasizes how pandemic-related stressors shaped students' preferences for online versus onsite learning.

While traditional models like the Technology Acceptance Model (TAM) emphasize perceived usefulness and ease of use [42], our results highlight the significant impact of extrinsic stressors. Challenges such as internet instability, computational limitations, infection severity, vaccination status, and environmental sensitivity influence students' preferences and learning behaviors. These findings align with behavioral theories underscoring the role of external conditions in shaping educational engagement [43]. Moreover, online learning success is closely tied to infrastructure—particularly reliable internet access—and the digital literacy of both students and instructors [44, 45]. Asynchronous learning offers flexibility, helping to reduce stress from real-time participation and connectivity issues. Additionally, game-based approaches, including serious games and MMORPGs, have shown promise in alleviating stress while enhancing motivation [46].

Online learning, especially in STEM fields, demands not only technical resources but also well-designed pedagogical frameworks that integrate computational tools with effective instructional design [47–49]. Our study suggests that remote learners face challenges not just in content acquisition, but also in synthesizing and applying information in the absence of traditional classroom interactions [50]. We examined several extrinsic variables—Internet Disruption (ID), Cognitive Demands (CD), Learning Platform (LP), Illness (I), Related Illness (RI), Vaccination (V), Infection Severity (IS), Environmental Sensitivity (ES), and Protective Adaptability (PA)—to understand their impact on learning preferences. These factors influence the outcome variable: Preference for Online or Onsite Learning (OL/OS). According to TAM, difficulties related to ID and CD directly affect perceived ease of use. Meanwhile, health and environmental factors contribute to psychological stress, influencing students' adaptability and learning decisions.

To support effective online learning, robust models of instruction, scheduling, and resource management are essential [51]. Basic proficiency in multimedia tools and digital platforms is critical for educators and learners to ensure seamless delivery and engagement. Strategies such as well-structured courses, visualization tools, and simulation software help clarify complex concepts, improve student confidence, and enhance learning outcomes [52]. From a psychological perspective, motivation is crucial in helping students overcome learning barriers. The concept of constructive alignment—connecting Intended Learning Outcomes (ILOs), Teaching-Learning Activities (TLAs), and Assessment Tasks (ATs)—can greatly enhance student motivation and provide clearer understanding [53, 54]. Our findings also reinforce the value of structured guidance and accessible educational tools for both teachers and learners [55]. Technology-Enhanced Learning (TEL) further expands this perspective by integrating content design, multimedia, instructional delivery, and iterative refinement [56]. Guided by the cognitive theory of multimedia learning, effective use of visual and auditory materials can enhance comprehension by engaging multiple memory systems. Our findings suggest that combining such tools with attention to external stressors and learner preferences is essential for building resilient and effective online education systems.

Recent advances in educational data science highlight the evolving role of machine learning in modeling student outcomes and preferences across diverse educational contexts. Selvakumar *et al*. [57] applied machine learning classifiers to primary and middle-school students' modality preferences, revealing K-Nearest Neighbors as most accurate in predicting online learning choices. Brigato and Iocchi [58] demonstrated that, under limited data conditions, simpler neural network architectures, along with data augmentation, often outperform complex deep learning models. In parallel, Althnian *et al*. [59] explored the impact of dataset size within medical classification tasks, identifying AdaBoost and Naive Bayes as robust options when data availability is constrained, underscoring dataset representativeness over sheer volum. Li [60] introduced a model for analyzing shifts in student learning preferences through educational big data, linking cognitive development and preference dynamics over time. In higher education, a Chilean case study used machine learning to predict student dropout across academic years, with Random Forest offering the strongest performance and socioeconomic factors influencing retention [61]. A systematic evaluation of learning algorithms on image classification tasks further revealed that model performance on small datasets is highly sensitive to algorithmic choice and augmentation strategies [62, 63]. Kokol *et al*. [64] leveraged synthetic knowledge synthesis to map research on machine learning in small-sample scenarios, synthesizing evidence from various domains to address challenges inherent in limited data. Finally, in adaptive learning contexts, personalized systems using educational big data have shown promise in tailoring learning experiences dynamically, though often require sophisticated modeling and feature mapping [65].

Compared to recent work in educational data mining, this study stands out by integrating pandemic-related external stressors (e.g., vaccination status, infection severity, psychological stress) with machine learning classification. Furthermore, the use of SHAP enhances interpretability—rarely applied in student preference modeling under crisis conditions—positioning our study at the intersection of explainable AI and educational technology research.

Limitations and Future Work: The limited dataset size (n = 120) constrains generalizability and may lead to overfitting. Future research should include a more diverse and larger student population across multiple institutions and regions. The binary classification of learning preference could be extended to multi-class or multi-label frameworks to better capture nuanced preferences. Moreover, linking predictors to actual academic outcomes (e.g., GPA) and conducting longitudinal analyses could enrich understanding of temporal learning behaviors.

V. Conclusion

This research evaluated various machine learning classification models using survey data collected to explore online learning preferences during covid-19 pandemic. The analysis focused on estimating how different factors influence these preferences, identifying key external variables that have the most significant impact. These findings provide crucial insights into student behavior and can be used to predict preferences and develop effective methods for managing online education. Additionally, the study revealed that the goals and design of the learning system play a critical role in shaping student preferences, emphasizing the need for educational strategies that are closely aligned with students' needs and expectations.

Although the findings are primarily based on data from a specific region, they hold potential applicability in broader contexts. However, variations in technological literacy and internet access across regions may require contextual adjustments. Educators and policymakers should therefore consider local conditions when applying these insights elsewhere. Overall, this research enhances the understanding of the factors that influence student preferences in online learning and lays the groundwork for developing targeted educational interventions to improve the global online learning experience.

Institutional Review Board Statement

All procedures performed in this study involving human participants were in accordance with the ethical standards of the national research committee. Participation was voluntary, and no personally identifiable information was collected. The survey instrument was approved by the Ethical Committee of Suan Sunandha Rajabhat University (Ref. No. COE.2-012/2024). Informed consent was obtained from all participants, and all procedures adhered to the institutional and national ethical guidelines involving human subjects.

Competing Interests

The authors declare no competing interests.

Authors' Contributions

Writing—original draft, B. J.; Data curation and Formal analysis S. K., C.T., S.C., P.K., D.S. and B.J.; Supervisor, Conceptualization, writing—review & editing, B.J. The authors (C.T., S.C., P.K. and S.K) contributed equally to this paper. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

The authors would like to thank the Faculty of Science and Technology, Suan Sunandha Rajabhat University and Sukhothai Thammathirat Open University, Thailand.

References

[1] E. Kurilovas, "Advanced machine learning approaches to personalise learning: learning analytics and decision making," *Behav. Inf. Technol.*, vol. 38, no. 4, pp. 410–421, 2019.

[2] A. Moubayed *et al*., "E-learning: Challenges and research opportunities using machine learning & data analytics," *IEEE Access*, vol. 6, pp. 39117–39138, 2018.

[3] S. J. Qin and L. H. Chiang, "Advances and opportunities in machine learning for process data analytics," *Comput. Chem. Eng*., vol. 126, pp. 465–473, 2019.

[4] M. Archer, *Social Origins of Educational Systems*, Abingdon, U.K.: Routledge, 2013.

[5] V. I. Tinyakova *et al*., "Monitoring of human resources and a new educational structure for training specialists as key factors to reactivate the system of consumer cooperation in Russia," *Amazonia Investiga*, vol. 7, no. 17, pp. 353–359, 2018.

[6] S. Hilbert *et al*., "Machine learning for the educational sciences," *Rev. Educ.*, vol. 9, no. 3, e3310, 2021.

[7] D.-N. Lu, H.-Q. Le, and T.-H. Vu, "The factors affecting acceptance of e-learning: A machine learning algorithm approach," *Educ. Sci*., vol. 10, no. 10, p. 270, 2020.

[8] L. Bognár and T. Fauszt, "Factors and conditions that affect the goodness of machine learning models for predicting the success of learning," *Comput. Educ., Artif. Intell.*, vol. 3, 100100, 2022.

[9] S. Abu-Kaf *et al*., "Emotional distress among the Bedouin Arab and Jewish elderly in Israel: The roles of gender, discrimination, and self-esteem," *Psychiatry Res*., vol. 291, 113203, 2020.

[10] C. Atchison *et al.*, "Early perceptions and behavioural responses during the COVID-19 pandemic: A cross-sectional survey of UK adults," *BMJ Open*, vol. 11, no. 1, e043577, 2021.

[11] A. A. Alkhamees *et al*., "The psychological impact of COVID-19 pandemic on the general population of Saudi Arabia," *Compr. Psychiatry*, vol. 102, 152192, 2020.

[12] A. Joshi, M. Vinay, and P. Bhaskar, "Impact of coronavirus pandemic on the Indian education sector: Perspectives of teachers on online teaching and assessments," *Interact. Technol. Smart Educ*., vol. 18, no. 2, pp. 205–226, 2021.

[13] F. Martin, K. Xie, and D. U. Bolliger, "Engaging learners in the emergency transition to online learning during the COVID-19 pandemic," *J. Res. Technol. Educ*., vol. 54, no. sup1, pp. S1–S13, 2022.

[14] J. Psotka, "Exemplary online education: For whom online learning can work better," *J. Comput. Assist. Learn*., vol. 38, no. 1, pp. 199–201, 2022.

[15] P. Bradford *et al.*, "The Blackboard learning system: The be all and end all in educational instruction?" *J. Educ. Technol. Syst*., vol. 35, no. 3, pp. 301–314, 2007.

[16] I. Ketut Sudarsana *et al*., "The use of Google classroom in the learning process," *J. Phys., Conf. Ser*., vol. 1175, 012165, 2019.

[17] M. N. Y. Utomo, M. Sudayanto, and K. Saddhono, "Tools and strategy for distance learning to respond COVID-19 pandemic in Indonesia," *Ing. Syst. Inf.*, vol. 25, no. 3, pp. 383–390, 2020.

[18] S. Kaparthi and D. Bumblauskas, "Designing predictive maintenance systems using decision tree-based machine learning techniques," *Int. J. Qual. Rel. Manag*., vol. 37, no. 4, pp. 659–686, 2020.

[19] M. Somvanshi *et al*., "A review of machine learning techniques using decision tree and support vector machine," in *Proc. Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Pune, India, 2016, pp. 1–7.

[20] S. Kooptiwoot, S. Kooptiwoot, and B. Javadi, "Application of regression decision tree and machine learning algorithms to examine students' online learning preferences during COVID-19 pandemic," *Int. J. Educ. Pract*., vol. 12, no. 1, pp. 82–94, 2024.

[21] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Jr. Conf. Artif. Intell. (IJCAI)*, Montreal, QC, Canada, 1995, pp. 1137–1143.

[22] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds., Boston, MA, USA: Springer, 2009, pp. 532–538.

[23] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan *et al*., Eds., Oxford, U.K.: Academic Press, 2019, pp. 542–545.

[24] S. Chen, Y.-J. J. Goo, and Z.-D. Shen, "A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree for forecasting fraudulent financial statements," *Sci. World J.*, vol. 2014, 968712, 2014.

[25] S. Hou *et al*., "Research on C5.0 algorithm improvement and the test in lightning disaster statistics," *Int. J. Control Autom*., vol. 7, no. 1, pp. 181–190, 2014.

[26] T.-T. Huynh-Cam, L.-S. Chen, and H. Le, "Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance," *Algorithms*, vol. 14, no. 11, p. 318, 2021.

[27] G. Wang and N. Wu, "A comparative study on contract recommendation model: Using Macao mobile phone datasets," *IEEE Access*, vol. 8, pp. 39747–39757, 2020.

[28] L. Zhou and H. Fujita, "Posterior probability based ensemble strategy using optimizing decision directed acyclic graph for multi-class classification," *Inf. Sci.*, vol. 400, pp. 142–156, 2017.

[29] N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of

techniques for predicting academic performance," in *Proc. 37th Annu. Front. Educ. Conf. (FIE)*, Milwaukee, WI, USA, 2007, pp. T2G-7–T2G-12.
[30] A. Silva, "Using data mining to predict secondary school student performance," M.S. thesis, Univ. Porto, Porto, Portugal, 2008.
[31] M. Mayilvaganan and D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of students academic environment," in *Proc. Int. Conf. Commun. Netw. Technol. (ICCNT)*, Sivakasi, India, 2014, pp. 168–173.
[32] G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, "Predicting students drop out: A case study," in *Proc. 2nd Int. Conf. Educ. Data Min. (EDM)*, Cordoba, Spain, 2009, pp. 41–50.
[33] I. Lykourentzou *et al*., "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Comput. Educ*., vol. 53, no. 3, pp. 950–965, 2009.
[34] D. J. Hand, "Measuring classifier performance: A coherent alternative to the area under the ROC curve," *Mach. Learn.*, vol. 77, no. 1, pp. 103–123, 2009.
[35] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Trans. Knowl. Data Eng*., vol. 17, no. 3, pp. 299–310, 2005.
[36] F. Provost and T. Fawcett, "Robust classification for imprecise environments," *Mach. Learn*., vol. 42, no. 3, pp. 203–231, 2001.
[37] H. Luan and C.-C. Tsai, "A review of using machine learning approaches for precision education," *Educ. Technol. Soc*., vol. 24, no. 1, pp. 250–266, 2021.
[38] I. Issah *et al*., "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," *Decis. Anal. J*., vol. 7, 100204, 2023.
[39] H. Munir, B. Vogel, and A. Jacobsson, "Artificial intelligence and machine learning approaches in digital education: A systematic revision," *Information*, vol. 13, no. 4, p. 203, 2022.
[40] R. O. Wissing *et al*., "Peer relationships buffer the negative association of online education with education satisfaction and subsequently with study engagement among undergraduate medical students," *BMC Med. Educ*., vol. 22, no. 1, p. 276, 2022.
[41] K. Lobos *et al*., "Expectations and experiences with online education during the COVID-19 pandemic in university students," *Front. Psychol*., vol. 12, 815564, 2022.
[42] A. Granić and N. Marangunić, "Technology acceptance model in educational context: A systematic literature review," *Br. J. Educ. Technol*., vol. 50, no. 5, pp. 2572–2593, 2019.
[43] M. A. Gurban and A. S. Almogren, "Students' actual use of e-learning in higher education during the COVID-19 pandemic," *SAGE Open*, vol. 12, no. 2, 21582440221091250, 2022.
[44] C. Jie and N. M. Ali, "COVID-19: What are the challenges of online learning? A literature review," *Int. J. Adv. Res. Future Ready Learn. Educ*., vol. 23, no. 1, pp. 23–29, 2021.
[45] P. Sepulveda-Escobar and A. Morrison, "Online teaching placement during the COVID-19 pandemic in Chile: challenges and opportunities," *Eur. J. Teach. Educ*., vol. 43, no. 4, pp. 587–607, 2020.
[46] N. Tavares, "The use and impact of game-based learning on the learning experience and knowledge retention of nursing undergraduate students: A systematic literature review," *Nurse Educ. Today*, vol. 117, 105484, 2022.
[47] R. W. Bybee, *The Case for STEM Education: Challenges and Opportunities*, Arlington, VA, USA: NSTA Press, 2013.
[48] J. Cavanaugh, S. J. Jacquemin, and C. R. Junker, "Variation in student perceptions of higher education course quality and difficulty as a result of widespread implementation of online education during the COVID-19 pandemic," *Technol. Knowl. Learn*., vol. 28, no. 4, pp. 1787–1802, 2023.
[49] L. Zhai, "A inquiry teaching mode based on STEM education," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 14, no. 17, p. 44, 2019.
[50] M. Honey, *Design, Make, Play: Growing the Next Generation of STEM Innovators*, New York, NY, USA: Routledge, 2013.
[51] H. Allam *et al*., "Medical Undergraduate Students' Perception about Online Education during the COVID-19 Pandemic," *Open Access Maced. J. Med. Sci*., vol. 10, no. E, pp. 213–218, 2022.
[52] Y. Mor and N. Winters, "Design approaches in technology-enhanced learning," *Interact. Learn. Environ*., vol. 15, no. 1, pp. 61–75, 2007.
[53] J. Biggs and K. Collis, "Towards a model of school-based curriculum development and assessment using the SOLO taxonomy," *Aust. J. Educ*., vol. 33, no. 2, pp. 151–163, 1989.
[54] P. A. Alexander and P. H. Winne, *Handbook of Educational Psychology*, 2nd ed., New York, NY, USA: Routledge, 2012.
[55] S. Abaci *et al*., "Supporting school teachers' rapid engagement with online education," *Educ. Technol. Res. Develop*., vol. 69, no. 1, pp. 29–34, 2021.
[56] Y. Y. Cho and H. Woo, "Factors in evaluating online learning in higher education in the era of a new normal derived from an Analytic Hierarchy Process (AHP) based survey in South Korea," *Sustainability*, vol. 14, no. 5, p. 3066, 2022.
[57] V. Selvakumar *et al*., "Predicting primary and middle-school students' preferences for online learning with machine learning," *S. Afr. J. Child. Educ*., vol. 13, no. 1, p. 1206, 2023.
[58] L. Brigato and L. Iocchi, "A close look at deep learning with small data," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Milan, Italy, 2021, pp. 2490–2497.
[59] A. Althnian *et al*., "Impact of dataset size on classification performance: an empirical evaluation in the medical domain," *Appl. Sci*., vol. 11, no. 2, p. 796, 2021.
[60] M. Li, "A model to predict and analyze students' learning preferences and their cognitive development through educational big data," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 18, no. 16, pp. 4–20, 2023.
[61] C. A. Palacios et al., "Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile," *Entropy*, vol. 23, no. 4, p. 485, 2021.
[62] J. Kabathova and M. Drlik, "Towards predicting student's dropout in university courses using different machine learning techniques," *Appl. Sci.*, vol. 11, no. 7, p. 3130, 2021.
[63] I. Iqbal *et al*., "Comparative investigation of learning algorithms for image classification with small dataset," *Appl. Artif. Intell*., vol. 35, no. 10, pp. 697–716, 2021.
[64] P. Kokol, M. Kokol, and S. Zagoranski, "Machine learning on small size samples: A synthetic knowledge synthesis," *Sci. Prog*., vol. 105, no. 1, 00368504211029777, 2022.
[65] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review," *IEEE Access*, vol. 11, pp. 48392–48409, 2023.