Academic Performance Prediction Using Machine Learning Approaches: A Survey

Jialun Pan, Zhanzhan Zhao, and Dongkun Han

***Abstract*—Properly predicting students'academic performance is crucial for elevating educational outcomes in various disciplines. Through precise performance prediction, schools can quickly pinpoint students facing challenges and provide customized educational materials suited to their specific learning needs. The reliance on teachers' experience to predict students' academic performance has proven to be less accurate and efficient than desired. Consequently, the past decade has witnessed a marked surge in employing machine learning and data mining techniques to forecast students' performance. However, the academic community has yet to agree on the most effective algorithm for predicting academic outcomes. Nonetheless, conducting an analysis and comparison of the existing algorithms in this field remains meaningful. Furthermore, recommendations for selecting an appropriate algorithm will be provided to interested researchers and educators based on their specific requirements. This article reviews the state-of-the-art literature on academic performance predictions using machine learning approaches in recent years. It details the variables analyzed, the algorithms implemented, the datasets utilized, and the evaluation metrics applied to assess model efficacy. What makes this work different is that relevant surveys in the past 10 years are also analyzed and compared, highlighting their contributions and review methods. In addition, we compared the accuracy of various machine learning models using popular open-access datasets and determined the best-performing algorithms among them. Our dataset and source codes are released for future algorithm comparisons and evaluations in this community.**

***Index Terms*—Artificial intelligence (AI) in education, learning analytics, machine learning, neural networks (NNs), performance prediction, students' performance, supervised learning, survey.**

Abbreviation

AUC	Area under the curve.
CUD	Chinese University dataset.
DT	Decision tree.
EDM	Educational data mining.
FN	False negative.
FP	False positive.
GBT	Gradient boosted tree.
KNN	K-nearest neighbor.
LR	Logistic regression.
MAE	Mean absolute error.
MLR	Multiclass logistic regression.
MSE	Mean squared error.
NB	Naive Bayes.
NN	Neural network.
OULAD	Open University learning analytics dataset.
PRISMA	Preferred reporting items for systematic reviews and metaanalyses.
RF	Random forest.
RMSE	Root mean squared error.
ROC	Receiver operating characteristics.
SVM	Support vector machine.
SVM-L	Support vector machine—linear.
SVM-P	Support vector machine—polynomial.
SVM-R	Support vector machine—RBF.
SVM-S	Support vector machine—sigmoid.
TN	True negative.
TP	True positive.

Received 2 March 2024; revised 15 July 2024, 5 November 2024, 2 February 2025, and 10 March 2025; accepted 13 March 2025. Date of publication 24 March 2025; date of current version 18 April 2025. This work was supported by the Hong Kong Teaching Development and Language Enhancement Grant under Grant TDLEG-4170989. *(Corresponding author: Dongkun Han.)*

Jialun Pan and Dongkun Han are with the Department of Mechanical and Automation Engineering, Chinese University of Hong Kong, Hong Kong (e-mail: allen06@link.cuhk.edu.hk; dkhan@mae.cuhk.edu.hk).

Zhanzhan Zhao is with the School of Humanities and Social Science, Chinese University of Hong Kong (Shenzhen), Shenzhen, China (e-mail: zhanzhanzhao@cuhk.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TLT.2025.3554174, provided by the authors.

Digital Object Identifier 10.1109/TLT.2025.3554174

I. Introduction

Academic performance encompasses a student's learning activities. It is reflected in a variety of assessment methods, including homework, quizzes, examinations, and projects throughout the academic term [1]. Timely insights into students' academic abilities enable teachers to adjust their instructional plans dynamically and help prevent dropouts, which can affect career prospects, strain public resources, and remain a significant concern, especially with high dropout rates on online platforms [2], [3], [4], [5], [6], [7], [8].

Traditionally, teachers assess the risk of academic failure in their students based on previous performance in homework and exams. However, limited knowledge about students' backgrounds makes teachers rely mainly on past records for performance predictions.

Recently, artificial intelligence (AI)-driven education has progressed rapidly, with EDM leveraging data analysis to gain insights into learning processes [9], [10], [11]. By utilizing these techniques, universities can improve student performance predictions and track academic progress more effectively [12], [13]. Studies show that various forecasting methods benefit both educators and administrators: educators can identify struggling students early and adjust their instructional strategies, while administrators can enhance resource allocation and program planning based on these insights [14]. These techniques also extend to online education, fostering AI integration in virtual learning and higher education [15].

The purpose of this survey is to examine recent research to address four specific research questions. These research questions will be thoroughly discussed in Section III-A.

1) What are the popular machine learning approaches used in recent years?
2) What are the factors used in predicting students' performance?
3) What are the datasets used to train the prediction model, and if they are open access?
4) What is the evaluation metric used when assessing the performance of the prediction model?

The authors focus particularly on the algorithms used, key factors, widely used datasets, evaluation methods, and their application scenarios. The findings are subsequently validated through simulations in Section V using the most widely recognized dataset in this field. In addition, all the algorithms and datasets used in the simulation are openly accessible. Note that this study does not solely focus on EDM. For readers interested in learning analytics and AI in education, this study provides a clear guide for selecting appropriate datasets and the most effective algorithms. Furthermore, the authors contribute a high-quality dataset from the Chinese University of Hong Kong to the field of research.

In the subsequent paragraphs, we provide a comprehensive summary of the practical applications of EDM based on academic performance prediction and outline the structure of our survey. Section I-A discusses a quintessential EMD application of the early warning system, which utilizes machine learning techniques to discern patterns and determinants of student dropout, thereby enabling early interventions to bolster retention rates. Section I-B explores the implementation of early warning systems that leverage predictive analytics to identify students at risk, permitting educators to deploy prompt support measures aimed at promoting academic achievement. Section I-C illustrates how EDM has broadened to encompass a variety of predictive objectives, including scholarship allocation, identification of at-risk students, postgraduation aspirations, and university admissions processes, with machine learning methods augmenting decision-making processes and predictive precision. Section I-D delineates the structure and organization of the subsequent text, setting the stage for a detailed discussion.

A. Dropout Predictions

While EDM has numerous applications, one of its most crucial tasks is to predict student progression and learning outcomes, such as dropout rates, performance, and course grades [16]. In predicting student dropout, Chung and Lee [2] employed the RF algorithm, achieving a 95% accuracy rate when analyzing data from a Korean high school. Hegde and Prageeth [3] highlighted the significant threat that student dropout poses to educational institutions and used the NB algorithm to predict the probability of dropout, identifying key contributing factors, such as academics, demographics, psychology, and health. Yaacob et al. [5] compared various algorithms, including DT, LR, RF, KNN, and NN. It found that LR performed best on a dataset of 64 Malaysian university students. Similarly, Pérez et al. [6] conducted a case study on students in systems engineering at a Colombian university, comparing DTs, LR, NB, and RF algorithms.

B. Early Warning System

In addition to predicting students dropout, the development of early warning systems has also attracted significant interest from researchers and educators. In the study of [17], a regression model was used for feature engineering to identify the most critical factors that influence students' academic performance. Utilizing the selected features, the learning analytics framework successfully predicted students' final grades as early as one-third of the way through the semester. In [18], multiple machine learning and deep learning algorithms were tested to develop a prediction model at different stages of the course. The RF emerged as the most effective approach for predictions. Through feature engineering, the model achieved an average accuracy of 79%, even when the course was only 20% complete. Akçapinar et al. [1] developed an early-warning system that successfully identified struggling students within as little as three weeks from the start of the semester. For this study, the KNN algorithm achieved an accuracy of 89% in the overall examination.

C. Other Prediction Goals

Besides the aforementioned applications, prediction goals have extended to meaningful objectives beyond academic performance. Previous work has focused on supporting scholarship programs, as proposed in [19] and [20], where C4.5 (DT) and AdaBoosting classification algorithms have proven valuable in selecting suitable candidates. The authors in [21] and [22] concentrated on identifying at-risk students in a virtual learning environment, primarily by analyzing interaction data with the online system. Imhof et al. [23] focused on analyzing student behavioral data from e-Learning platforms to identify students who are prone to procrastination. Anticipating the students predisposed to procrastination enables educators to provide timely intervention, offering essential assistance and resources to facilitate enhanced time management and bolstered academic efficiency among learners. The study conducted in [24] predicted student difficulties in a digital design course session. Fedushko and Ustyianovych [25] aimed to identify the key factors that contribute to student success and analyzed the significant factors that support specific processes. Recent research highlights the effectiveness of genetic algorithm-based prediction methods, achieving higher accuracy in identifying key attributes.

D. Survey Structure and Organization

The rest of this article is organized as follows. In Section III, we provide a summary of the included studies, outlining their search processes, search strings, and research questions. This section lays the foundation for the proposed methodology in the

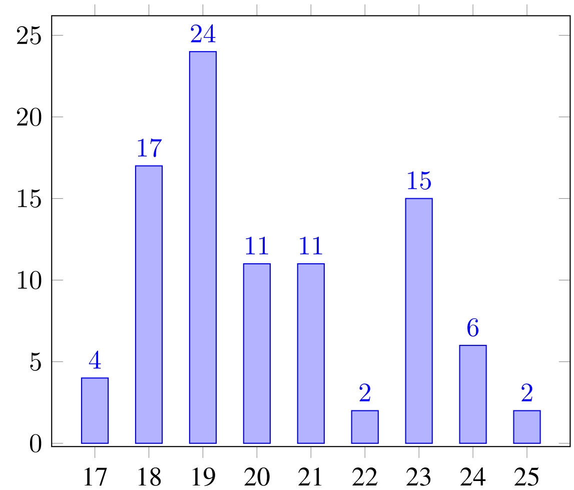

Fig. 1. Amount of performance prediction papers published.

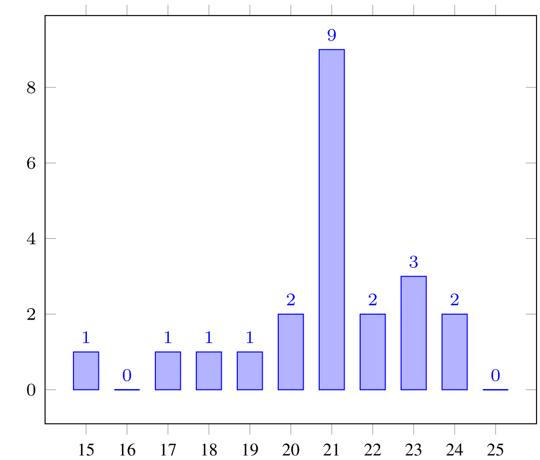

Fig. 2. Amount of survey papers published.

following section. Section III outlines the research methodology used in this survey, including the definition of research questions, the identification of search items, and the selection of pertinent literature. In Section IV, we review approximately 90 technical articles published from 2017 to the present. Based on the findings of these papers, we address the proposed research questions. In addition, we extract and synthesize datasets that are significant in this field. Section V presents our simulation results of several popular algorithms utilized in academic performance prediction. To ensure the fairness in the comparison, we work on the most widely used open-access dataset discussed in Section IV. Section VI emphasizes the contribution of the survey again and discusses the inspiration of practical applications, and finally, Section VII concludes this article.

II. Existing Reviews and Surveys

This section provides an overview of existing reviews and surveys conducted in the field of academic performance prediction between 2015 and 2025. We review the search terms, research questions, and inclusion criteria used in previous studies to identify areas for improvement. This analysis aims to provide a more objective and comprehensive methodology for gathering a broader range of technical literature. Reviewing existing literature enhances our understanding of the field. Figs. 1 and 2 illustrate the distribution of investigated publications over the years. Notably, 2021 has the highest number of identified records.

A. Search Process for Related Reviews

The search was carried out in February 2025 using multiple widely used digital libraries in the fields of computer science and education, namely IEEE Xplore, Springer, ACM Digital Library, ScienceDirect, ResearchGate, and Google Scholar. To ensure comprehensive coverage, we utilized two sets of search strings, adapting the search items as necessary to suit the specific formats of each dataset. The first set is "("student" OR "academic") AND ("performance" OR "grade") AND "prediction" AND ("survey" OR "review")," and the second set is "student" AND "prediction" AND "("survey" OR "review")." Initially, we collected 231 relevant publications, which were then manually inspected through their titles, keywords, and abstracts to eliminate irrelevant hits. It is important to note that some reviews on EDM also contribute to students' performance prediction; however, most of them primarily summarize techniques and applications. Consequently, there are only a few reviews with "EDM" in their titles included in our review. After the screening process, we retained 20 papers for further examination due to the others' lack of search items or research questions.

The studies [26], [27], [28], [29], and [30] have been excluded from the Supplementary Material (see Tables I–V in the Supplementary Material) due to their omission of relevant factors. For instance, [26] failed to systematically present the inclusion and exclusion criteria, [27] lacks both the research questions and exclusion criteria, [28] lacks the research questions and inclusion criteria, [29] did not introduce the selected keywords and inclusion/exclusion criteria, and [30] did not address its research questions. Nonetheless, their inclusion criteria will still be considered in the Section III-B.1.

B. Synthesis of Relevant Reviews

Based on our results, which can be found in the Supplementary Material (see Tables I–V in the Supplementary Material), we have identified 15 relevant papers along with their corresponding search strings and research questions. This allows us to determine the most frequently utilized set of keywords. By examining the tables, we observe the following.

Search Strings From Relevant Reviews:

1) Performance: 41 times.
2) Prediction: 30 times.
3) Data mining: 17 times.
4) Machine learning: 9 times.
5) Student: 76 times.
6) Academic: 12 times.
7) Education: 11 times.
8) System: 5 times.
9) Grade: 7 times.

TABLE I
SUMMARY OF RESEARCH OBJECTIVES AND CONTRIBUTIONS

Contribution	Details
Literature investigation	Investigate recent literature on student performance prediction from 2017 to 2025.
Methodology review	Discuss the review methodology of previous review works innovatively, and highlight research on student performance prediction in the past 8 years.
Progress presentation	Present significant progress in student performance prediction in a beginner-friendly way, serving as a quick start guideline for educators.
Dataset evaluation	Evaluate and compare popular datasets, highlighting those useful as benchmarks in student performance prediction.
Algorithm evaluation	Provide an evaluation standard among popular algorithms based on the benchmark dataset and compare their performances.
Open source codes	Supply with open source codes for the algorithms and datasets comparisons.
Dataset contribution	Contribute an open source dataset with real student information from the Faculty of Engineering at the Chinese University of Hong Kong, containing 1095 samples of students and 7 attributes each.

Regarding the proposed research questions, while each review may have its unique set of concerns and objectives, certain popular questions have consistently captured the attention of authors across different regions and time periods. It is noteworthy that surveys commonly utilize Boolean logic search queries to retrieve relevant studies. For instance, a search query might include phrases such as "factors affecting students' performance" OR "predicting students' performance." In such cases, individual words may appear multiple times and even more than the number of papers examined.

Research Questions From Relevant Reviews:

1) What machine learning method are used?: 10 times.
2) What are the most meaningful factors or features?: 10 times.
3) What is the most suitable method?: 3 times.
4) What is the evaluation metric of student performance?: 2 times.
5) What is the evaluation metric used when assessing the performance of the prediction model?: 2 times.

The frequency at which a research question is posed serves as an indicator of the level of attention dedicated to the topic. Research questions that are mentioned fewer than twice are disregarded as they often pertain to the unique objectives of individual literature. A significant issue arises from the lack of specific evaluation criteria in the majority of studies, as they primarily rely on extracting simulation results from individual works and making comparisons solely based on them. However, this approach seems unpersuasive as it fails to account for variations in simulation backgrounds and model configurations. Therefore, in the subsequent section, we devote additional efforts on addressing the limitations of previous review works and formulating our own unique survey methodology.

The main contributions of our survey are highlighted in the Table I.

III. Survey Methodology

This section provides an overview of our review methodology, which encompasses the formulation of research questions, the utilization of statistical data to define search strings with the inclusion and exclusion criteria.

A. Research Questions and Strategy

The primary objective of this article is to conduct a comprehensive review of the literature in the field of student performance prediction, providing detailed guidance for disposed educators. Therefore, we provide specific instructions for each essential step of the prediction process.

Building upon the findings from the previous section, in addition to addressing the commonly addressed topic, we have also included essential questions that are often overlooked by EDM researchers and analysts. To ensure the coherence of our contributions.

We have formulated the following *research questions*:

1) What are the popular machine learning approaches used in recent years?
2) What are the factors used in predicting students' performance?
3) What are the datasets used to train the prediction model, and if they are open access?
4) What is the evaluation metric used when assessing the performance of the prediction model?

Our search items are driven by the survey results in the subsection "synthesis of relevant reviews" and our research questions. Therefore, we conclude our search items as "("Student" OR "Academic") AND ("Performance" OR "grade" OR "level") AND ("prediction" OR "predict") AND ("data mining" OR "machine learning")" in the first stage. During the search process, we identified numerous pieces of literature with "EDM" in their titles, which may contain content related to performance prediction but did not match our search keywords. As a result, we conducted a second round of searching using the terms "educational data mining" or "EDM." We conducted our search across multiple mainstream data sources, including IEEE Xplore, Springer, ACM Digital Library, Science Direct, ResearchGate, and Google Scholar. A total of 775 papers published between 2017 and 2025 were included for consideration. In addition, we employed snowballing techniques to identify additional secondary studies [31].

B. Literature Selection

The selected articles were imported into Zotero [32],[1] a free reference management tool utilized for literature collection [33].

[1] [Online]. Available: https://www.zotero.org/

A total of 852 papers were initially included before the exclusion process. To ensure the clarity and robustness of our methodology, we reviewed and summarized the inclusion criteria based on a thorough examination of 20 relevant reviews in this field. Criteria that appeared fewer than five times were not considered. The inclusion and exclusion criteria for this work can be summarized as follows.

Inclusion Criteria:

1) Should be research papers.
2) Full text must be available.
3) Use machine learning approaches to predict the student's academic performance and introduce what methods are used.
4) Should be written in English.
5) Include feature engineering process, introduce what dataset is used and if it is open access.
6) Paper published after 2017.

Exclusion Criteria:

1) Review studies, abstracts, commentaries, book chapters, or editorials. Short papers, editorials, business posters, patents, already conducted reviews, technical reports, Wikipedia articles, survey studies, and extended papers of already reviewed papers.
2) Duplicate papers with similar contributions by the same author.
3) More than two items among factors, learning methods, dataset used, and evaluation metric are neglected or not clearly stated.

PRISMA is a research-based guideline that offers a standardized framework for reporting systematic reviews and metaanalyses. Its main focus lies in reviews that assess the impact of interventions [34]. This survey adhered to the four recommended stages outlined in the PRISMA statement [34]. These stages included identifying relevant studies, applying exclusion criteria, assessing relevance, and selecting studies for inclusion. Fig. 3 presents the PRISMA flow diagram, illustrating the sequential progression of our study.

Through the screening process, a total of 82 articles have been selected for further investigation. In the following section, we will summarize their characteristics, datasets, learning methods, and evaluation metrics.

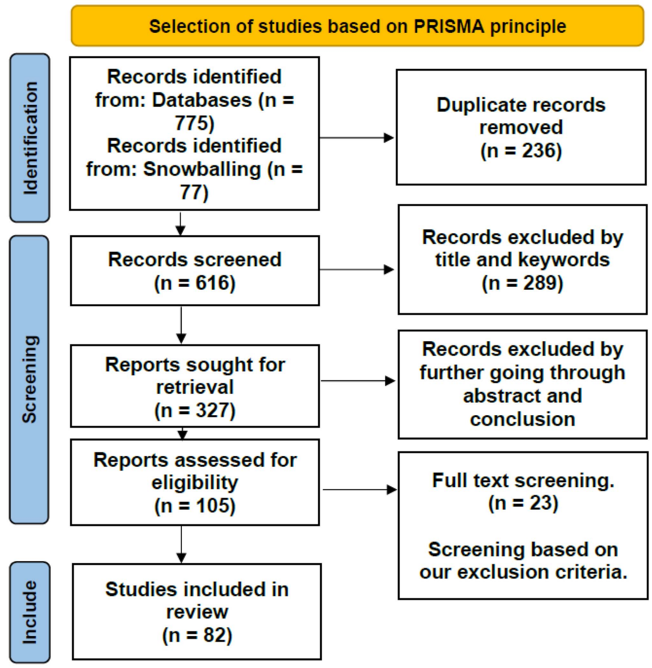

Fig. 3. Literature selection based on PRISMA.

TABLE II
PREDICTION TARGET OF REVIEWED STUDIES

Ref No.	Prediction target
[35], [36], [37], [38], [39]	Predict of students' letter grades in course taken
[36], [39], [40], [41], [22], [42], [43], [44]	Predict of students' pass or fail in course taken
[45], [46], [47], [14], [48], [49], [50]	Predict of students' GPA in the semester
[2],[3],[4],[5], [6],[7]	Predict of student's drop out
[21], [51], [52], [1], [53]	Predict at-risk students
[19], [20]	Predict of student's success in scholarship programs
[12]	Predict of student's learning time
[54]	Predict of student's active degree
[55]	Predict of student's learning rate
[43]	Predict of student's degree of participation
[23]	Predict of student's delay behavior
[24]	Predict of the difficulties encountered by students

IV. SURVEY RESULTS

This section will mainly proceed according to the sequence of the proposed research questions, encompassing the most crucial features and steps involved in predicting students' academic performance.

A. Review of Prediction Targets

Our inclusion criteria primarily center around the concept of "academic performance prediction," which encompasses various specific prediction targets, and the summary of prediction purposes can be found in Table II. Nevertheless, the simulations in this survey are constricted to the target of predicting the letter grade of a course, due to limited space and page restrictions.

B. Review of Machine Learning Techniques

In this survey, we have compiled a list of popular algorithms used for making predictions in Table III to provide educators with a clear overview of the patterns.

The pie chart in Fig. 4 demonstrates that the usage of these popular algorithms shows little variation. Among all the methods, DT accounts for the largest portion, constituting 21% of the total number. SVM, RF, NB, and KNN are also widely applied. Since multilayer perceptron (MLP) is often regarded as a type of artificial neural network (ANN) in many studies, for simplicity, we consider both MLP and ANN as NN in this study.

TABLE III
MACHINE LEARNING METHODS IN RELEVANT STUDIES

Evaluation metrics	Amount	Reference No.
Decision Tree	41	[56], [59], [60], [61], [54], [62], [36], [19], [2], [63], [13], [64], [39], [65], [46], [40], [66], [67], [20], [5], [6], [68], [69], [55], [47], [70], [71], [14], [43], [48], [72], [24], [50], [51], [52], [73], [74], [14], [8], [75], [76]
Naive Bayes	31	[56], [59], [60], [62], [36], [78]. [13], [65], [3], [39], [65], [40], [66], [41], [78], [19], [5], [6], [24], [1], [50], [51], [52]. [73], [8]
Support Vector Machine	31	[35], [59], [60], [61], [54], [62], [80], [77], [64], [81], [65], [21], [82], [67], [18], [22], [20], [79], [55], [83], [23], [49], [24], [1], [51], [52], [73], [74], [8], [77], [74]
Neural Network(with MLP)	28	[84], [35], [56], [59], [60], [12], [62], [36], [13], [64], [65], [66], [18], [49], [72], [24], [1], [52], [74], [75], [76]
Random Forest	25	[60], [61], [62], [36], [63], [77], [37], [81], [39], [65], [67], [18], [4], [20], [6], [70], [23], [48], [49], [72], [1], [50], [74], [8], [75]
K Nearest Neighbour	24	[56], [59], [54], [61], [77], [38], [41], [83], [18], [20], [5], [42], [79], [55], [14], [23], [49], [1], [51], [52], [75], [14], [8]
Logistic Regression	12	[59], [61], [77], [22], [5], [6], [48], [24], [52], [74], [8], [75]
Gradient Boosted Tree	8	[61], [85], [67], [18], [4], [20], [43], [23]

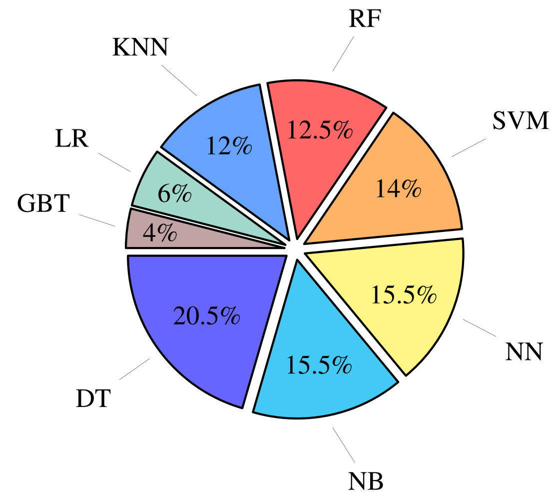

Fig. 4. Machine learning methods used for prediction.

In addition, ensembling multiple algorithms together has proven to be an effective approach for making predictions and has been shown to improve performance. For example, Rawat and Malhan [56] utilized voting to ensemble J48(DT), NB, IBK, and ANN, resulting in higher accuracy compared to individual algorithms. The implementation results in [13] also demonstrated that the bagging method significantly enhanced the performance of the DT model. As discussed in [57], there is a general consensus that combining prediction methods leads to more accurate and robust results. Therefore, it is recommended that educators not only assess the performance of individual models but also consider combinations of models to achieve a high accuracy.

C. Machine Learning Methods

Choosing an appropriate machine learning algorithm is of utmost importance in improving the performance of classifiers. Predictive models effectively analyze educational data, with traditional machine learning widely adopted for their accuracy and fast learning [58].

In essence, learning methods can be categorized into various groups, with a commonly used criterion being the classification of algorithms as supervised and unsupervised models.

The supervised learning model is capable of predicting output values based on given inputs [50]. Typically, we need to assign labels to the training data to enable the model to classify accurately. In the context of supervised algorithms, they can be categorized into classification and regression methods. The primary distinction between these approaches lies in their output: classification divides data into appropriate categories, while regression predicts numeric data. In the domain of academic performance prediction, the scenario is complex, often requiring the prediction of students' pass or fail outcomes. This typical binary classification task is best addressed using methods such as SVM and DT. On the other hand, regression methods, such as LR, are more suitable for predicting students' grades in specific learning activities.

The unsupervised methods explore unlabeled data to discover hidden patterns within a dataset, without the need to predict specific output variables [50]. Clustering methods can be further classified into hierarchical clustering and partitioned clustering, with k-means and particle swarm optimization (PSO) clustering algorithms serving as typical examples.

Interestingly, combining these two models can yield even better performance than using them individually. Research by Livieris et al. [78] has shown that by leveraging a small amount of labeled data alongside a large amount of unlabeled data, the accuracy of semisupervised methods can be significantly improved, thereby reducing the burden of labeling tasks. Studies in [86] and [87] also proved the superiority of the semisupervised machine learning method in predicting students' grades in the final exams.

Although several researchers have developed new classifiers based on baseline methods [e.g., reduced training vector (RTV)-SVM was proposed in [21] to maintain accuracy with reduced training data, the authors in [14], [42], and [67] made predictions using various types of KNN, and student academic performance predicting (SAPP) in [88] combines four-layer stacked long short-term memory network, RF, and GB], traditional unsupervised classifiers still dominate a significant portion of the field. However, multiple researchers have noted that the best algorithms for prediction may vary, and no single algorithm is universally superior [61]. A similar viewpoint was expressed in [39], where the authors suggested that each classifier has its own strengths and weaknesses, and there is no perfect algorithm. This implies that comparing different methods is essential to determine the most effective algorithms based on specific datasets. These findings align with our own observations during the review process, as the best algorithms varied across nearly every study. For instance, in [41], NB outperformed the KNN with the highest accuracy of 93.6%, while the authors in [46], [66], and [70] supported the effectiveness of DTs in handling prediction problems. RF was considered a more suitable model for classification in [9], [39], [63], and [81], and SVM outperformed the KNN and DT with the highest accuracy of 95% in [54]. However, Al-Shehri et al. [82] believed that both SVM and KNN were well-suited to the problem, while Jalota and Agrawal [60] recommended using an MLP. It is crucial for researchers to compare algorithms before applying them to predict academic performance [61].

D. Factors Used in Performance Prediction

Features are universally acknowledged as crucial factors that impact the performance of prediction models. While some individuals tend to incorporate all available variables to identify potential rules [89], others advocate for selecting relevant attributes as an effective means of enhancing performance [1], [59]. It has been observed that employing all features can even have a detrimental effect on accuracy [84]. Therefore, in the case of the latter approach, feature selection is considered essential prior to training.

Feature selection aims to identify the most relevant variables to reduce the computational cost of predictors, including CPU time and memory. By incorporating appropriate attributes, the accuracy can be efficiently improved [24]. In addition, Hajar et al. [20] emphasized the significance of feature engineering, which is widely recognized as one of the most critical components in the learning process. Consequently, various evaluators were employed to identify the attributes that have a substantial impact on students' overall performance [49]. In [4], an embedded feature selection method was utilized, which assesses the importance of each feature during model training. This approach presents a lightweight implementation of a predictive model, resulting in lower resource costs. In addition, Polyzou and Karypis [67] utilized RF to evaluate the importance of features, while Waheed et al.[22] employed a sparse feature reduction technique to identify the most critical factors. In [72], feature engineering was conducted using sequential forward selection to select the strongest pair of student characteristics. Furthermore, Alturki and Alturki [50] compared several attribute evaluators, including search-based method, correlation-based method, information gain-based method, and wrapper with NB, and made a final decision based on their respective results.

Based on the outcomes of our review, we observed that the most significant feature sets vary. This variation can be attributed not only to the availability of different datasets for learning analytics but also to the employed methods of feature engineering. Although reaching a consensus is challenging, we identified similarities among different studies by examining relevant literature. Some papers focused on the influence of individual characteristics on overall academic performance, while others aimed to identify the most significant set of attributes contributing to the final outcomes. For example, Hooshyar et al. [65] demonstrated that students' tendencies to procrastinate can directly lead to failure in online learning. Fadelelmoula [45] examined the impact of attendance on students' performance and found a positive correlation between the two. For dropout prediction, the authors in [6] and [47] found grade point average (GPA) and high school performance, especially in specialized courses, to be the most influential factors. Similar patterns emerged when predicting academic performance. The results of [48] and [50] indicated that the GPA of each semester is one of the most relevant factors associated with successful graduation. In addition, Fernandes et al. [85] revealed strong relationships between GPA, absences, and academic performance, while also considering the potential effects of school, age, and neighborhood. The significant effect of students' family background, such as the educational status of their parents, was highlighted in [90]. Furthermore, based on experimental findings, Budiman et al. [46] emphasized the importance of organizational involvement, age, and place of birth in determining students' final outcomes. Shahiri et al. [26] also supported the notion that demographic information, such as gender, has a significant impact on grades. However, Chen et al. [18] disagreed with this view, stating that there is no clear correlation between the final outcomes and demographic attributes. Instead, they indicated that assessment scores and engagement are more significant.

We conducted an analysis of research studies conducted within the past six years about predicting students' performance. We extracted the factors used to train the predictive models from these studies and compiled the results in Table IV. This table serves as a comprehensive reference for identifying the prominent attributes considered in these studies. To facilitate analysis, we categorized the factors into the following groups: demographics information (1), online study behavior (2), previous academic records (3), current academic performance (4), family background (5), attendance or class activeness (6), and others (7). Additional detailed explanations can be found in Table V.

Fig. 5 reveals that researchers predominantly consider demographics, current study performance, family background, and activeness in class as the most influential factors. This finding aligns with [26], which also emphasizes that cumulative GPA (CGPA) is the most commonly used factor.

E. Datasets

Our survey placed particular emphasis on the datasets used to train the prediction model. We observed that many literature sources lack specific descriptions of their datasets, including information on whether they are open access. This omission is detrimental to educators who wish to practice the content

TABLE IV
EXPLANATIONS ABOUT CATEGORIES

Number of categories	Explanation
1 (Demographics Information)	Gender, Age, Region, Major, Nationality, Address...
2 (Online Study Behavior)	Click stream Data, Assessment Marks, Modules Information...
3 (Previous Academic Records)	Grade of Entrance Exam, Grade of High School, High School Type or Rank...
4 (Current Academic Performance)	Pass of Fail in the School Year, GPA of each semester, Performance of Specific Course, Rank, Credits, Grade of quiz or exam...
5 (Family Background)	Father's Job, Mother's Job, Parents' Education Status, Family Size...
6 (Attendance or Activeness on Class)	Number of Hands Raising, Attendance Rate, Absence Rate...
7 (Others)	Travel Time, Study Time, Free Time, Extracurricular Activities...

TABLE V
ATTRIBUTES USED IN TRAINING PREDICTIVE MODELS (EXPLANATION OF THE NUMBERS ARE CLARIFIED IN TABLE IV)

Ref No.	1	2	3	4	5	6	7
[35], [54], [36], [13], [40]	*			*	*	*	
[56], [68]	*			*	*		
[85], [39]	*			*		*	*
[77], [50]				*			
[66], [79], [49]	*		*	*			
[41], [82], [71]	*				*		
[62]	*			*	*	*	*
[80]	*		*	*	*	*	*
[81]	*			*			
[91]	*				*	*	
[22]	*	*		*			
[47]	*		*	*		*	
[70]	*			*	*		*
[92]		*					
Total	22	2	5	20	14	11	5

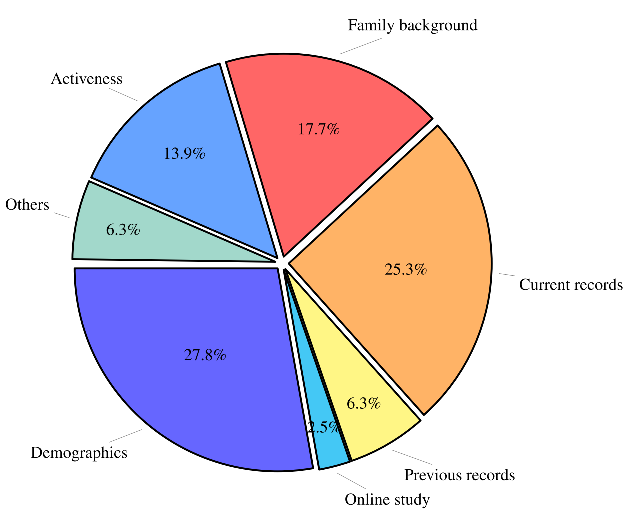

Fig. 5. Proportion of different features employed.

presented in the papers. In addition, it is objective to evaluate the performance of different classifiers using the same criteria, especially when a benchmark dataset is available.

Due to privacy and copyright concerns, the majority of existing studies utilized student information within their regions/countries, making the datasets nonopen sources. While some authors mentioned that the data are available upon request from the corresponding author [13], access is restricted to specific purposes. Table VI presents a public student dataset suitable for performance prediction, with the prominent one being the OULAD provided by Open University.

In the following part of the survey, we will introduce another dataset, which is proposed and released for the first time by this work, named "CUD" from the Faculty of Engineering, The Chinese University of Hong Kong, in Section V.

Variables and data structure: OULAD package downloaded contains several files and the information contained are listed as follows.

The "studentInfo" file contains the demographic information such their gender, region, age, and their disability, and educational background information including highest educational level, number of previous attempts, studied credits, as well their final outcomes that are divided into four categories of pass, fail, withdrawn, and distinguish. Student identifiers, such as student ID and code module.

The "student registration" file contains data pertaining to students' registration activities, including the date of registration and unregistration.

The "studentAssessment" and "assessment" files comprise information related to assessments, such as the assessment ID, assessment type, student submission date, and exam scores.

The "studentVle" and "vle" files capture the interactions between students and the online teaching system. They encompass clickstream data, timestamps of access, and the types of activities performed by students.

The "course" file serves as supplementary material, providing the presentation codes and module lengths.

In general, the OULAD has been extensively utilized as a benchmark dataset for academic performance prediction. It is readily accessible on the internet and user-friendly, making it an ideal resource for educators to begin their exploration. In Section V, our experiment will also be conducted using OULAD, enabling us to provide a comprehensive comparison of various popular algorithms.

F. Evaluation on the Prediction Model

Besides the algorithms and training data, the performance of classifiers are always of great interest in prediction model selection. We conducted a comprehensive review of relevant studies and compiled a list of all studies included in our work, which can be found in Table VII. Before introducing the performance

TABLE VI
DATASETS WITH OPEN ACCESS

Ref No.	Dataset description	Source
[35], [62], [36], [82], [69], [83], [93]	The data was collected from two public schools in the Alentejo region of Portugal during the academic year 2005-2006. This dataset includes 1044 instances with 33 attributes, encompassing student grades, demographic information, social factors, and school-related features.	Cortez P, Silva A (2008) Using data mining to predict secondary school student performance. In: 15th European concurrent engineering conference 2008, ECEC 2008; 5th Future of business technology conference, FUBUTEC 2008, pp. 5-12 [94]
[59], [21], [22], [96], [43], [96], [97], [98], [99]; [100], [101], [103], [88]	Open University Learning Analytics Dataset. The dataset encompasses data from 22 courses, consisting of 32,593 students. It includes their assessment results as well as logs documenting their interactions with the Virtual Learning Environment (VLE)	J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics Dataset," Scientific data, vol. 4, p. 170171, 2017.
[61]	Mendeley Data Repository The collected dataset comprises information from 12,411 samples, with each sample containing 44 variables. These variables contain both academic assessments and personal information, including socioeconomic status represented categorically.	https://data.mendeley.com/datasets/83tcx8psxv/1
[77]	The dataset includes the academic achievement grades of 1854 students who took the Turkish Language-I course at a state university in Turkey during the fall semester of 2019-2020.	Additional File 1 at the end of the paper.
[81], [74]	The dataset on student academic performance was sourced from the UCI student performance dataset, comprising 831 samples. Each student's data encompasses 22 attributes.	https://archive.ics.uci.edu/ml/datasets/student+performance.
[83], [75]	Students Academic Performance Dataset (SAPD) Data for this study was collected from three colleges in Assam, India, namely Duliajan College, Doomdooma College, and Digboi College. The dataset comprised 300 instances, with each instance containing 24 attributes.	Hussain, S., Dahan, N.A., Ba-Alwi, F.M., and Ribata, N., 2018. Educational data mining and analysis of students' academic performance using WEKA", Indonesian Journal of Electrical Engineering and Computer Science, 9, 2, 447-459.

TABLE VII
EVALUATION METRICS USED IN RELEVANT STUDIES

Evaluation metrics	Amount	Reference No.
Accuracy	41	[84], [56], [12], [54], [62], [2], [63], [77], [13], [89], [37], [81], [39], [65], [21], [66], [41], [91], [21], [4], [78], [20], [5], [103], [42], [79], [55], [47], [50], [51], [52], [73], [74], [88]
Recall	27	[27], [33], [34], [41], [43], [45], [51], [52], [54], [57], [58], [60], [64], [65], [71], [73], [77], [77], [79], [81], [83], [84], [88], [89], [103], [133], [137]
Precision	25	[60], [62], [36], [77], [13], [64], [39], [65], [40], [66], [41], [91], [67], [22], [4], [68], [103], [69], [55], [70], [71], [7], [49], [24], [75], [88]
F-Measure	24	[59], [60], [54], [62], [36], [77], [13], [64], [39], [65], [40], [66], [91], [67], [4], [103], [69], [55], [70], [71], [7], [49], [72], [24], [1], [50], [74], [88]
Sensitivity	4	[2], [70], [1], [52]
Specificity	4	[2], [70], [71], [1]
AUC	8	[2], [77], [40], [67], [5], [6], [7], [43]
RMSE	8	[35], [59], [61], [63], [80], [47], [14], [25]
MAE	5	[63], [80], [39], [82], [47]
MSE	4	[17], [80], [82], [83]
R^2	4	[17], [61], [80], [83]
Kappa	5	[39], [40], [79], [43], [24]

TABLE VIII
EXPLAINATIONS ABOUT CATEGORIES

Classification	Actually positive	Actually negative
Predicted positive	True positive(TP)	False positive(FP)
Predicted negative	False negative(FN)	True negative(TN)

parameters, it is important to understand the concept of the confusion matrix, which is commonly used for accuracy calculation in data mining. The information within the confusion matrix is crucial for evaluating the performance of the classification model [19]. Table VIII provides a specific representation of the matrix. In this context, TP represents the number of positive students correctly predicted as positive, FP represents the number of positive students incorrectly predicted as negative, TN represents the number of negative students correctly predicted as negative, and FN represents the number of negative students incorrectly classified as positive.

Among the included technical literature, accuracy emerged as the most commonly used metric, as it provides an intuitive reflection of the number of correctly predicted samples. However, relying solely on classification accuracy is not sufficient for unbalanced classification problems, as stated in [103]. Similar sentiments were expressed in [24], where the author noted that accuracy is a less reliable performance measure than precision or recall, particularly in imbalanced datasets. Therefore, additional metrics were employed to ensure a more objective assessment. Figs. 6 and 7 present two pie charts illustrating the distribution of these metrics. Fig. 6 mainly focuses on the evaluation indices for classification models, such as SVM, KNN, and NB, with accuracy, recall, precision, and F-measure representing a significant portion.

However, less used factors, such as RMSE, ROC curve, and Cohen's kappa coefficient, can also be well utilized to assess the goodness of fit between the data and a model [24].

V. SIMULATION RESULTS AND COMPARISON AMONG ALGORITHMS

In this section, we will analyze numerical examples and compare the performance of eight popular machine learning algorithms: NN, DT, GBT, KNN, MLR, NB, RF, and SVM.

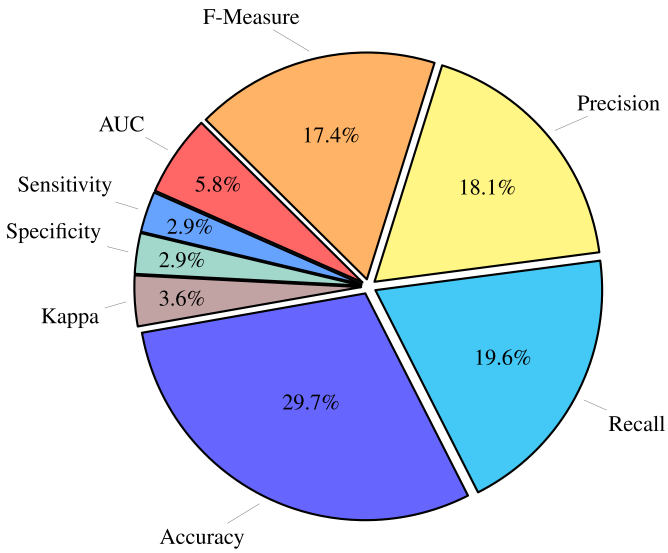

Fig. 6. Proportion of evaluation indices for classification model.

TABLE IX
FEATURE DISCRETIZATION FOR OULAD

Feature name	Discretization method
Gender	M=0, F=1
Disability	N=0, Y=1
Imd_band	0–10%=0, 20–30%=1, 30–40%=2, 40–50%=3, 50–60%=4, 60–70%=5, 70–80%=6, 80–90%=7 90–100%=8
Highest_education	HE Qualification=0, A level or Equivalent=1, Lower than A level=2, Post Graduate Qualification=3, No Formal quals=4
Age_band	0-35=0, 35-55=1, 55<=2
Region	East Anglian Region=0, Scotland=1, South East Region=2, North Western Region=3, Wales=4, West Midlands Region=5, South Region=6, South West Region=7, Yorkshire Region=8, London Region=9, East Midlands Region=10, Ireland=11, North Region=12

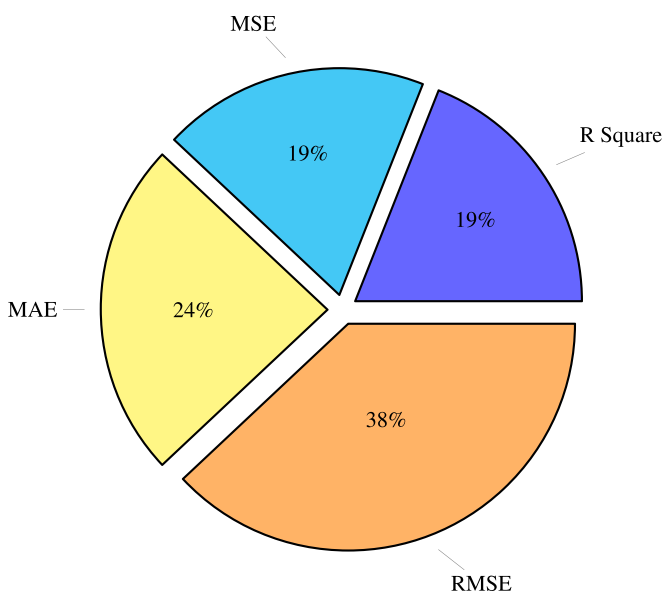

Fig. 7. Proportion of evaluation indices for regression model.

The comparison will be conducted using two datasets. The first is the OULAD [104], which is widely recognized and utilized as a benchmark dataset in the field of EDM [95]. The second dataset is from the Faculty of Engineering, Chinese University of Hong Kong. Our comparison will be conducted in two rounds to make the work more objective.

A. Experiment Dataset

OULAD has been described in detail in the previous section, while CUD was obtained from the admission department from our college, which contains the academic information of 1095 individuals from the Chinese University of Hong Kong during the period from 2019 to 2021, and for each sample, seven attributes related to their previous academic performance was recorded, such as their admission channel, secondary school information, type of entrance exam, math background, grade of MATH1510, grade of ENGG1130, and the grade of ESTR1006, all these factors have a strong relationship with students' academic performance.

Students' privacy in this work is fully considered and protected. All data were collected from the students who were allowed to use their personal course data for research purposes, and this teaching project is authorized by the Faculty of Engineering, Chinese University of Hong Kong, to use and disseminate the dataset. Moreover, all personal information, such as names and student ID have been removed to further protect students' privacy.

After careful comparison, the "studentInfo" file from OULAD was selected as the first set of training data. This choice was based on two main facts: the significantly larger number of samples compared to the other two files, and the comprehensive nature of the information it contains, encompassing students' demographics, performances, and final outcomes. To facilitate clearer feature identification by the predictive model, digitization was necessary. Currently, there are various discretization methods available, such as one-hot encoding [35] and direct discretization. During the review process, direct discretization emerged as the more popular choice, as one-hot encoding increases the data dimensionality, which can negatively impact learning model performance. Consequently, specific numerical values were assigned to each category of features, as outlined in Tables IX and X for OULAD and CUD, respectively. After removing duplicated data, the "studentInfo" file from OULAD contained information from approximately 28 000 samples, including student ID, gender, region, highest education level, index of multiple deprivation band, age band, number of previous attempts, studied credits, disability status, and their final results as Table XI shows, most of the factors here belongs to demographic features. After data processing, CUD contains 546 students' information related to their academic performance; to make the samples in the dataset more balanced, only students who took

TABLE X
FEATURE DISCRETIZATION FOR CUD

Feature name	Discretization method
Admission_Channel	JUPAS=1, Mainland (Non-JEE)=2, Mainland Fee-paying=3, Mainland Scholarship (tuition)=4, Non-JUPAS (Local)=5, Non-JUPAS (Other Region)=6, Mainland=7, International=8, JUPAS Appeal-in=9
Secondary_School	1-10 based on their level
Exam	AD YR 1=1, AD/HD completed=2, GCE-AL=3, HKDSE=4, IB=5, Mainland Gaokao (JEE)=6, SAT+SAT Subject Test=7, Others=8
Math_Background	HKDSE5**=6.5, HKDSE5*=5.83, HKDSE5=5.11, HKDSE4=4.61, HKDSE3=6, HKDSE2=0, HKDSE1=0, Mainland Gaokao (JEE):>140=6.44, Mainland Gaokao (JEE):130-139=5.87, Mainland Gaokao (JEE):120-129=6.19, Mainland Gaokao (JEE):110-119=6.19, Mainland Gaokao (JEE):<110=6

part in Mainland Gaokao and Hong Kong Diploma of Secondary Education Examination (HKDSE) are kept.

Therefore, throughout our simulation, we can also compare the relevancy of demographic information and academic information to the final prediction performance.

B. Prediction Target

In the OULAD dataset, we employ demographic factors such as age, highest level of education, disability, and academic factors including earned credits and the number of previous attempts to predict the final pass or fail outcome for the corresponding students.

In the context of CUD, academic performance factors, specifically the type of college entrance exam, entrance exam grade, and grade in MATH1510, are considered when predicting the score of ENGG1130. It is important to note that enrollment in ENGG1130 is only possible for students who have passed MATH1510, resulting in a high correlation between the grades of these two courses.

C. Predictive Models and Setups

Based on the review results in Section IV, we included all eight algorithms for comparison in a Python environment. The algorithms used were GBT, LR, DT, NB, NN, SVM, RF, and KNN. We employed the sklearn package [105] to complete the task. All of the coding was conducted in Google Colab, a cloud-based platform provided by Google for executing and sharing Jupyter Notebook environments. Google Colab offers free access to a virtual machine with preinstalled libraries and dependencies, providing a convenient environment for Python code development and execution. The hyperparameters in the simulation are set to their default values since hyperparameters vary across different prediction situations, and model tuning is a complex task that relies on experience and specific circumstances. Therefore, to maintain fairness, the hyperparameters of the predictive models remain unchanged.

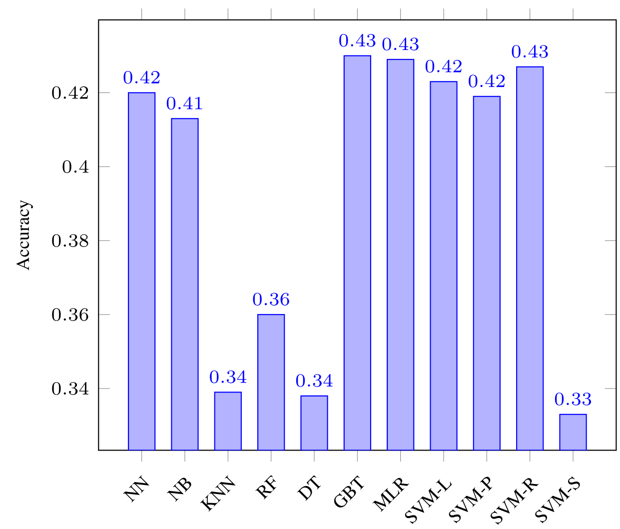

Fig. 8. Accuracy comparison for OULAD.

1) Model Setup for OULAD: In the experiment, 80% of the total data were allocated for training the model, while the remaining 20% were reserved for testing. For the KNN model, k was set to 4. In the RF model, the number of estimators was set to 200, and the criterion was set to entropy. In the GBT model, the number of estimators was set to 100, the learning rate was set to 1.0, and the maximum depth was set to 1. All other parameters were left at their default values. Due to the larger number of hyperparameters associated with deep learning, specific adjustments were made. The NN was trained using the backpropagation technique, with a three-layer network consisting of eight input nodes, four output nodes, and nine nodes in the hidden layer. The learning rate was set to 0.05, the momentum factor was 0.01, and the training iterations were set to 1000 steps.

2) Model Setup for CUD: To control variables, all the model setup keep the same as those for OULAD with the exception of BPNN, since the number of the input and output are different, we have to change the amount of nodes both in input layer and output layer, respectively. Therefore, when we make a comparison based on CUD, the number of input nodes are set to be 5, and that of output nodes are set to be 7 for the NN.

D. Experiment Discussion

The comparison results are illustrated in Figs. 8–15. It is important to note that there are four different SVM classifiers involved in the comparison based on their kernels. SVM-L represents the linear kernel, SVM-P represents the polynomial kernel, SVM-R represents the radial basis functions (RBF) kernel, and SVM-S indicates the sigmoid kernel.

Fig. 8 presents a comparison of the accuracy attained by different machine learning algorithms based on the dataset OULAD. The horizontal axis denotes the evaluated algorithms,

TABLE XI
DATA SAMPLES GIVEN BY OULAD (PART)

Code Module	Student ID	Gender	Region	Highest Education	Age Band	Studied Credits	Final Result
AAA	11391	M	East Anglian Region	HE Qualification	55<=	240	Pass
AAA	28400	F	Scotland	HE Qualification	35-55	60	Pass
AAA	30268	F	North Western Region	A Level or Equivalent	35-55	60	Withdrawn
AAA	31604	F	South East Region	A level or Equivalent	35-55	60	Pass
AAA	32885	F	West Midlands Region	Lower Than A Level	0-35	60	Pass
AAA	38053	M	Wales	A Level or Equivalent	35-55	60	Pass
AAA	45462	M	Scotland	HE Qualification	0-35	60	Pass
AAA	45642	F	North Western Region	A Level or Equivalent	0-35	120	Pass
AAA	52130	F	East Anglian Region	A Level or Equivalent	0-35	90	Pass
AAA	53025	M	North Region	Post graduate Qualification	55<=	60	Pass
AAA	57506	M	South Region	Lower Than A Level	35-55	60	Pass

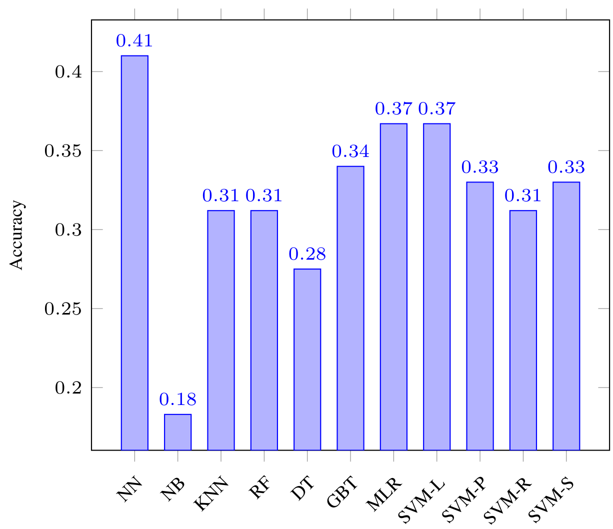

Fig. 9. Accuracy comparison for CUD.

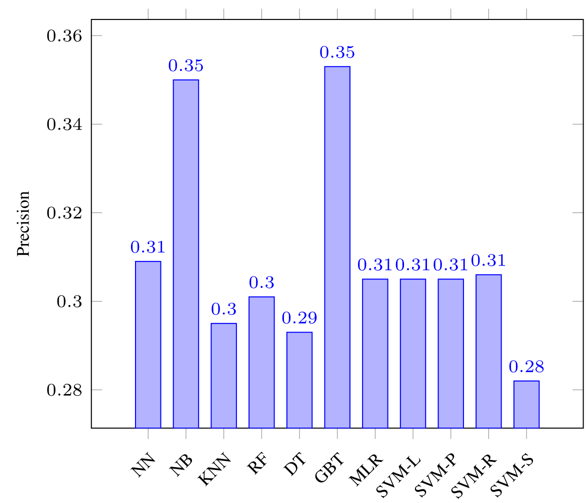

Fig. 10. Precision comparison for OULAD.

while the vertical axis represents the corresponding accuracy values. The highest accuracy is achieved by GBT with a value of 0.430, followed closely by MLR with an accuracy of 0.429. The remaining algorithms exhibit slight differences in their accuracy values, ranging from 0.33 to 0.36. Fig. 9 presents the model performance in terms of accuracy on the dataset CUD. The highest accuracy of 0.41 is achieved by the NN model, while the NB model exhibits poor performance with an accuracy of approximately 0.18.

Fig. 10 illustrates the evaluation of algorithms based on OULAD, with the x-axis representing the algorithms and the y-axis representing the corresponding precision values. It is evident that NB and GBT significantly outperform the other six predictive models. GBT attains the highest precision of 0.353, closely followed by NB with a precision of 0.350. Fig. 11 presents the precision values based on CUD. Note that the KNN achieves the highest precision of 0.38, followed by RF with a value of 0.31 and DT with a value of 0.27. Conversely, NB exhibits the poorest performance with a precision value of 0.13.

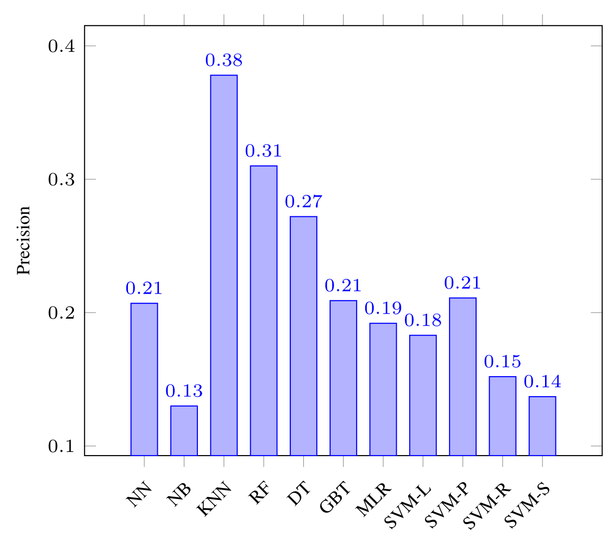

Fig. 11. Precision comparison for CUD.

Fig. 12 presents the recall values based on OULAD. Among the evaluated algorithms, GBT achieves the highest recall value of 0.313, followed by MLR with a recall of 0.305. Fig. 13 presents the model recall performance when handling CUD. Similar to the precision measures, KNN achieves the highest value once again, reaching 0.29. However, DT, GBT, and SVM

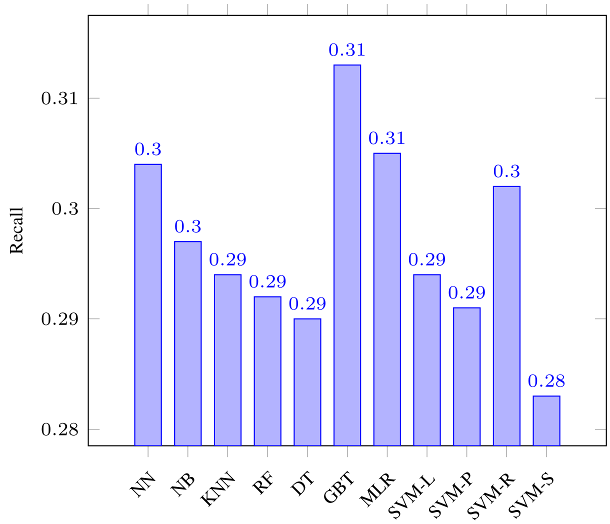

Fig. 12. Recall comparison for OULAD.

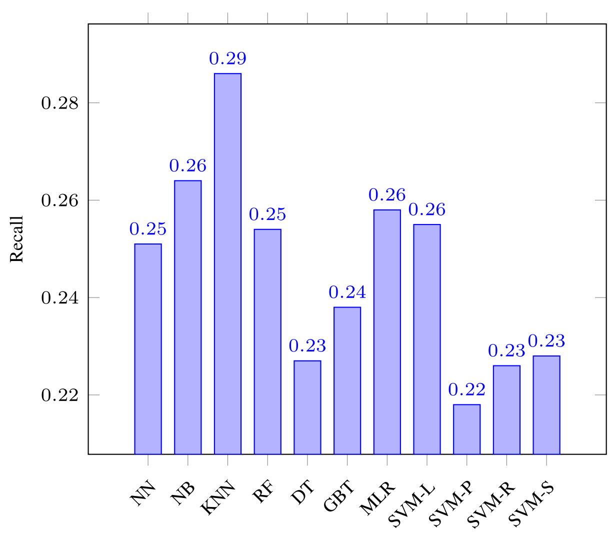

Fig. 13. Recall comparison for CUD.

with polynomial, RBF, and sigmoid kernels exhibit relatively poor performance, with recall values ranging from 0.22 to 0.24.

Fig. 14 offers a concise visual representation of the comparative performance of the evaluated machine learning algorithms in terms of F-measure based on OULAD. The results highlight that KNN, DT, GBT, and RF achieve relatively higher F-measure values, approximately 0.29. Fig. 15 depicts the F-measure performance of each algorithm, revealing that the KNN outperforms all other models with a value of 0.29, followed by RF with 0.27. Conversely, NB exhibits the lowest performance with a value of 0.12.

In summary, our experiment validates the fact that the most appropriate model for making predictions varies depending on the training dataset. When evaluating the model performance on OULAD, GBT exhibits excellent overall performance. However, it is obvious that GBT does not attain the highest F-measure, which assesses the balance between recall and precision. A lower

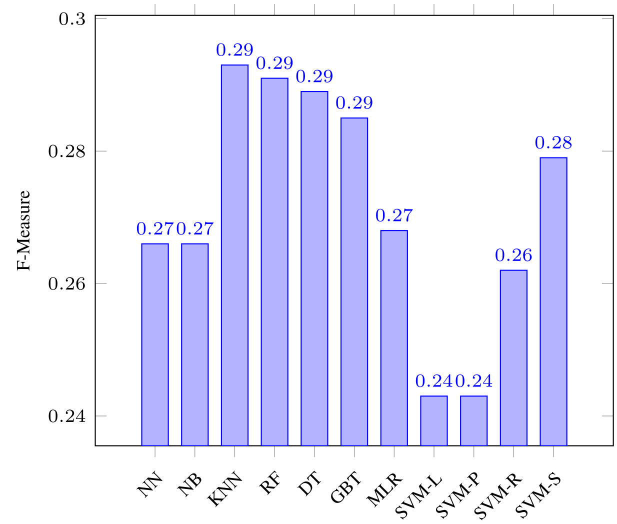

Fig. 14. F-Measure comparison for OULAD.

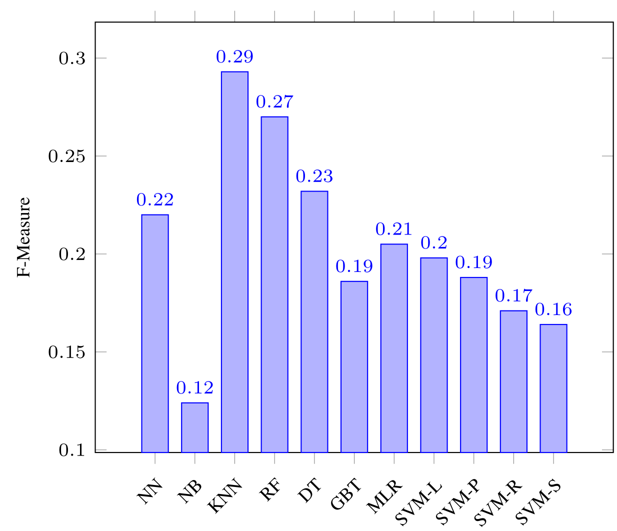

Fig. 15. F-Measure comparison for CUD

TABLE XII
OPTIMAL ALGORITHM IN DIFFERENT DATASET AND EVALUATION METRICS

OULAD	Accuracy	GBT 0.43
OULAD	Precision	GBT 0.35
OULAD	Recall	GBT 0.31
OULAD	F-Measure	KNN 0.29
CUD	Accuracy	NN 0.41
CUD	Precision	KNN 0.38
CUD	Recall	KNN 0.29
CUD	F-Measure	KNN 0.29

F-measure value indicates a greater disparity between precision and recall. On the other hand, when dealing with CUD, the KNN demonstrates a superior performance across most evaluation metrics, the optimal model in different datasets and evaluation metrics are summarized in Table XII.

In the following, we will consider the impact of the dataset scale on the model's performance based on two datasets.

Figs. 1 and 2 in the Supplementary Material reveal a similar result in terms of the variation trend observed across all algorithms. The accuracy depicted in two figures are essentially comparable to that in Figs. 8 and 9. The primary distinction lies in the total number of samples in the training dataset, or, in other words, the scale of the training dataset. Both sets employed 80% of the dataset for training and the remaining 20% for testing. The accuracy presented reflects the outcomes of the testing process. As the training data increases, the accuracy tends to be stabilized when the sample number reaches approximately 10 000 and 250, respectively. These results indicate that more training data does not necessarily lead to an improved accuracy.

Fig. 3 in the Supplementary Material highlights the significant impact of training data scale on SVML, SVMR, and SVMS. The F1 scores of the former two models are enhanced, whereas the latter exhibits a contrasting trend. The graph clearly indicates that a dataset size of 15 000 sets of data is adequate to capture the characteristics of all the models, particularly their F-measurement scores. Fig. 4 in the Supplementary Material supports the result where all the algorithms can reach their peak point before 300 sets of training data.

Similar to the accuracy analysis, the precision scores of the examined algorithms exhibit minimal correlation with the dataset scale. Figs. 5 and 6 in the Supplementary Material illustrate that the precision scores remain consistently stable regardless of the variations in training data, even though there are some fluctuations.

When examining the final performance index presented in Fig. 7 in the Supplementary Material, the recall scores exhibit overall stability with minor fluctuations, except for SVMR, SVML, and SVMS. The first two models seem to be influenced by the number of training data, particularly when the sample size reaches 10 000, although the improvement in their scores is not substantial. Conversely, SVMS shows a consistent decline in value as the training data increases. Notably, when dealing with a small-scale dataset in Fig. 8 in the Supplementary Material, most algorithms demonstrate stable recall values with the exception of GBT and NB reaching the peak point at around 300 sets of training data.

Running time is considered as an important metric for different models since it determines their efficiency in handling large amounts of data in practical implementations. An ideal predictor should not only ensure accuracy but also minimize the CPU time usage. Figs. 9 and 10 in the Supplementary Material provide insights into the running time comparison among different machine learning methods. Generally, SVM models require more time compared to other methods. For instance, SVML takes around 2.5 min to learn from a dataset of 25 000 sets of data in a single run, while the other three SVM models take approximately 1 min. This starkly contrasts with other classifiers that complete training in less than 5 s. These findings suggest that when training models with large datasets and prioritizing running time, SVM may not be the optimal choice.

The graphs allow us to draw the conclusion that the GBT exhibits the most comprehensive and balanced performance when handling big-scale dataset such as OULAD, excelling in various aspects, such as accuracy, precision, and recall. But the KNN presents a better performance based on the dataset CUD.

Moreover, our research has revealed that the relationship between the scale of training data and model performance is not linear. Specifically, based on our findings, when handling a big amount of data, such as OULAD, a dataset comprising 15 000 sets of student information is deemed sufficient for effectively assessing the performance of a predictive model, otherwise it may display an overfitting phenomenon such as the prediction of SVMS based on OULAD.

VI. Discussion

The utilization of machine learning techniques has a significant influence on modern education. Looking ahead, AI possesses the potential to function as an effective teaching assistant, aiding not only in predicting students' academic performance but also in tailoring the learning experience to their unique needs, facilitating more profound instruction, and even matching teachers and students based on their individual learning styles [107]. The integration of machine learning techniques into EDM is poised to revolutionize modern education in the near future.

However, there are still certain limitations that need to be addressed before the widespread adoption of learning methods can be achieved.

To begin with, ML models are only as good as the data they are trained on, while data can be biased due to various factors such as socioeconomic backgrounds, cultural differences, and the educational institution itself. In recent years, federated learning methods have emerged as an effective solution for addressing data bias issues. Federated learning is a distributed machine learning approach where multiple participants collaboratively train a global model without sharing raw data. Researchers in [108] and [109] employed federated learning in EDM and demonstrated its potential to effectively utilize educational data from diverse backgrounds while maintaining data privacy.

Second, the restricted accessibility of open-source datasets poses a significant obstacle to their utilization in training predictive models and for evaluation purposes. This scarcity primarily stems from privacy concerns, as most centralized model training architectures require pooling data at a centralized location, such as a cloud server. This approach necessitates transferring student data across the network, leading to privacy issues [110]. Distributed learning techniques, such as federated learning, can effectively address these concerns by keeping raw data on local devices. Consequently, a broader range of institutions and individuals can actively contribute to the education community [111], [112], ultimately leading to the establishment of a more equitable and comprehensive assessment system.

Third, besides the limitations of open-source datasets, the high sparsity of datasets also contributes to the lack of training data. The study in [106] proposes an innovative technical pathway to address this issue using data modeling and generative AI data augmentation, thereby expanding the application boundaries of generative AI in educational data augmentation.

Fourth, Models trained on data from a specific institution, curriculum, or demographic may not generalize well to other contexts. For example, a model trained to predict student success in a particular high school in the United States might not be applicable to schools in other countries with different educational systems [113]. Generalizability is also a concern when dealing with different levels of education, such as transitioning a model from predicting undergraduate student success to graduate-level courses [114], Luo et al. [115] addressed generalizability issues by developing a new classification method for blended courses. This method clusters students based on their online learning behaviors using the expectation-maximization algorithm. The results indicated that classifying blended courses based on students' online behavior improved prediction accuracy in each category.

Last but not least, ML models, especially complex ones such as deep learning networks, can act as "black boxes," making it difficult to understand how they arrive at their predictions. This lack of interpretability is problematic in educational settings because educators and policymakers need to understand the basis of the predictions to make informed decisions [116], while model explanation techniques generate insights into why a student is likely to fail [117]; it is crucial to know the factors contributing to that prediction to provide the appropriate support. Increasingly, researchers and educators are focusing on this area. According to [118], the RF model outperforms other models with an accuracy of 90.4%, highlighting the importance of having models that are both interpretable and transparent, thereby demonstrating the potential of explainable AI in education. The methodology in [102] compared DT, an intrinsically explainable model, with ANNs, which require post hoc explainability techniques. The results show that the ANN has higher accuracy, indicating that improving the interpretability of predictive models remains a challenge for future research.

VII. Conclusion

This survey investigated more than 100 existing studies conducted in past ten years on academic performance prediction. It encompassed various aspects, including popular machine learning algorithms, significant factors influencing students' performance, datasets and their accessibility, evaluation methods, and application scenarios.

Based on our results, it shows that DTs, NB, NNs, SVMs, RF, and KNNs are the most commonly employed algorithms for building prediction models [119], [120], [121]. In terms of evaluating prediction performance, accuracy, recall, precision, and F-measure are all studied as important metrics for conducting a comprehensive assessment [122], [123], [124].

In addition, numerical simulations and comparisons were conducted among several popular machine learning algorithms using the OULAD and CUD. The simulation results in Section V indicates that for datasets primarily consisting of demographic information, such as OULAD, GBTs have the potential to outperform other methods. Conversely, for predictions based on academic-related information, KNN demonstrated the best capability. It is important to note that various factors, including data scale and data imbalance, can influence the situation, and different models may have divergent results.

Algorithms and datasets employed in this work can be accessed online.[2]

Acknowledgment

The authors would like to appreciate the valuable inputs from Prof. L. Tan and Ms. J. Yan at the Chinese University of Hong Kong. The authors would also like to appreciate the inspiring comments and constructive suggestions from the editor and reviewers.

References

[1] G. Akçapinar, A. Altun, and P. Aşkar, "Using learning analytics to develop early-warning system for at-risk students," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, pp. 1–20, 2019.

[2] J. Y. Chung and S. Lee, "Dropout early warning systems for high school students using machine learning," *Child. Youth Serv. Rev.*, vol. 96, pp. 346–353, 2019.

[3] V. Hegde and P. P. Prageeth, "Higher education student dropout prediction and analysis through educational data mining," in *Proc. Int. Conf. Inventive Syst. Control*. 2018, pp. 694–699.

[4] A. Alamri et al., "Predicting MOOCs dropout using only two easily obtainable features from the first week's activities," in *Intelligent Tutoring Systems*. Berlin, Germany: Springer, 2019, pp. 163–173.

[5] W. F. W. Yaacob, N. M. Sobri, S. A. M. Nasir, W. F. W. Yaacob, N. D. Norshahidi, and W. Z. W. Husin, "Predicting student drop-out in higher institution using data mining techniques," *J. Phys., Conf. Ser.*, vol. 1496, no. 1, 2020, Art. no. 012005.

[6] B. Pérez, C. Castellanos, and D. Correal, "Predicting student drop-out rates using data mining techniques: A case study," in *Applications of Computational Intelligence*. Berlin, Germany: Springer, 2018.

[7] L. Qiu, Y. Liu, Q. Hu, and Y. Liu, "Student dropout prediction in massive open online courses by convolutional neural networks," *Soft Comput.*, vol. 23, no. 20, pp. 10287–10301, 2019.

[8] Y. T. Badal and R. K. Sungkur, "Predictive modelling and analytics of students' grades using machine learning algorithms," *Educ. Inf. Technol.*, vol. 28, no. 3, pp. 3027–3057, 2023.

[9] F. Alshareef, "Educational data mining applications and techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, pp. 729–734, 2020.

[10] Z. Wang, Y. Tian, R. Chen, and L. Kong, "Research and application of AI-enabled education," in *Data Science*. Berlin, Germany: Springer, 2023.

[11] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. J. D. Baker, Eds., *Handbook of Educational Data Mining*, 1st ed. Boca Raton, FL, USA: CRC Press, 2010.

[12] P. Prasetyawan, I. Ahmad, R. I. Borman, Y. A. Ardiansyah Pahlevi, and D. E. Kurniawan, "Classification of the period undergraduate study using back-propagation neural network," in *Proc. Int. Conf. Appl. Eng.*, 2018, pp. 1–5.

[13] M. Ragab, A. M. K. Abdel Aal, A. O. Jifri, and N. F. Omran, "Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques," *Wireless Commun. Mobile Comput.*, vol. 2021, 2021, Art. no. e6241676.

[14] S. Hussain and M. Q. Khan, "Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning," *Ann. Data Sci.*, vol. 10, no. 3, pp. 637–655, 2023.

[15] J. E. Raffaghelli, M. E. Rodríguez, A.-E. Guerrero-Roldán, and D. Bañeres, "Applying the UTAUT model to explain the students' acceptance of an early warning system in higher education," *Comput. Educ.*, vol. 182, 2022, Art. no. 104468.

[16] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Implementing AutoML in educational data mining for prediction tasks," *Appl. Sci.*, vol. 10, no. 1, , pp. 1–27, 2020.

[2] [Online]. Available: Github_link_TLT24a

[17] O. H. T. Lu et al., “Applying learning analytics for the early prediction of students’ academic performance in blended learning,” *J. Educ. Technol. Soc.*, vol. 21, no. 2, pp. 220–232, 2018.
[18] W. Chen, C. G. Brinton, D. Cao, A. Mason-Singh, C. Lu, and M. Chiang, “Early detection prediction of learning outcomes in online short-courses via learning behaviors,” *IEEE Trans. Learn. Technol.*, vol. 12, no. 1, pp. 44–58, Jan.–Mar. 2019.
[19] E. Sugiyarti, K. A. Jasmi, B. Basiron, and M. Huda, “Decision support system of scholarship grantee selection using data mining,” *Int. J. Pure Appl. Math.*, vol. 119, no. 15, pp. 2239–2249, 2018.
[20] M. Hajar, J. Adil, Y. Ali, and A. Z. Khalid, “Predicting Student Success in a Scholarship Program,” in *Digital Technologies and Applications*. Berlin, Germany: Springer, 2022, pp. 333–341.
[21] K. T. Chui, D. C. L. Fung, M. D. Lytras, and T. M. Lam, “Predicting at-risk university students in a virtual learning environment via a machine learning algorithm,” *Comput. Hum. Behav.*, vol. 107, 2020, Art. no. 105584.
[22] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, “Predicting academic performance of students from VLE Big Data using deep learning models,” *Comput. Hum. Behav.*, vol. 104, 2020, Art. no. 106189.
[23] C. Imhof, I.-S. Comsa, M. Hlosta, B. Parsaeifard, I. Moser, and P. Bergamin, “Prediction of dilatory behavior in eLearning: A comparison of multiple machine learning models,” *IEEE Trans. Learn. Technol.*, vol. 16, no. 5, pp. 648–663, Oct. 2023.
[24] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, “Using machine learning to predict student difficulties from learning session data,” *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381–407, 2019.
[25] S. Fedushko and T. Ustyianovych, “Predicting pupil’s successfulness factors using machine learning algorithms and mathematical modelling methods,” in *Advances in Computer Science for Engineering and Education II*. Berlin, Germany: Springer, 2020, pp. 625–636.
[26] A. M. Shahiri, W. Husain, and N. A. Rashid, “A review on predicting student’s performance using data mining techniques,” *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015.
[27] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, “Analyzing and predicting students’ performance by means of machine learning: A review,” *Appl. Sci.*, vol. 10, no. 3, pp. 1–16, 2020.
[28] E. Alyahyan and D. Düştegör, “Predicting academic success in higher education: Literature review and best practices,” *Int. J. Educ. Technol. Higher Educ.*, vol. 17, no. 1, pp. 1–21, 2020.
[29] A. Khan and S. K. Ghosh, “Student performance analysis and prediction in classroom learning: A review of educational data mining studies,” *Educ. Inf. Technol.*, vol. 26, no. 1, pp. 205–240, 2021.
[30] B. Sekeroglu, R. Abiyev, A. Ilhan, M. Arslan, and J. B. Idoko, “Systematic literature review on machine learning and student performance prediction: Critical gaps and possible remedies,” *Appl. Sci.*, vol. 11, no. 22, 2021, Art. no. 10907.
[31] R. Alamri and B. Alharbi, “Explainable student performance prediction models: A systematic review,” *IEEE Access*, vol. 9, pp. 33132–33143, 2021.
[32] Zotero, “GitHub repository”, 2011. [Online.] Avaialble: https://github.com/zotero/zotero
[33] N. Sghir, A. Adadi, and M. Lahmer, “Recent advances in predictive learning analytics: A decade systematic review (2012–2022),” *Educ. Inf. Technol.*, vol. 28, no. 7, pp. 8299–8333, 2023.
[34] A. Liberati et al., “The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration,” *Ann. Intern. Med.*, vol. 151, no. 4, 2009, Art. no. W-65.
[35] C. Huang, J. Zhou, J. Chen, J. Yang, K. Clawson, and Y. Peng, “A feature weighted support vector machine and artificial neural network algorithm for academic course performance prediction,” *Neural Comput. Appl.*, vol. 35, no. 16, pp. 11517–11529, 2023.
[36] C.-C. Kiu, “Data mining analysis on student’s academic performance through exploration of student’s background and social activities,” in *Proc. Int. Conf. Adv. Comput., Commun. Automat.*, 2018, pp. 1–5.
[37] S. Gaonkar, “Grade prediction of educational institute using ML algorithm,” *Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol.*, vol. 6, pp. 569–575, 2020.
[38] J. M. Brown, “Predicting math test scores using K-nearest neighbor,” in *Proc. IEEE Integr. STEM Educ. Conf.*, 2017, pp. 104–106.
[39] K. Nahar, B. I. Shova, T. Ria, H. B. Rashid, and A. H. M. S. Islam, “Mining educational data to predict students performance,” *Educ. Inf. Technol.*, vol. 26, no. 5, pp. 6051–6067, 2021.
[40] S. Helal et al., “Predicting academic performance by considering student heterogeneity,” *Knowl.-Based Syst.*, vol. 161, pp. 134–146, 2018.
[41] I. A. Abu Amra and A. Y. A. Maghari, “Students performance prediction using KNN and Naïve Bayesian,” in *Proc. Int. Conf. Inf. Technol.*, 2017, pp. 909–913.
[42] A. Maghari, “Prediction of student’s performance using modified KNN classifiers,” in *Proc. Int. Conf. Eng. Future Technol.*, 2018, pp. 143–150.
[43] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, “Student engagement predictions in an e-learning system and their impact on student course assessment scores,” *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–21, 2018.
[44] V. T. N. Chau and N. H. Phung, “ST_OS: An effective semisupervised learning method for course-level early predictions,” *IEEE Trans. Learn. Technol.*, vol. 14, no. 2, pp. 238–256, Apr. 2021.
[45] T. Fadelelmoula, “The impact of class attendance on student performance,” *Int. Res. J. Med. Med. Sci.*, vol. 6, pp. 47–49, 2018.
[46] E. Budiman, H. Haviluddin, N. Dengan, A. H. Kridalaksana, M. Wati, and Purnawansyah, “Performance of decision tree C4.5 algorithm in student academic evaluation,” in *Computational Science and Technology*. Berlin, Germany: Springer, 2018.
[47] M. Nachouki, E. A. Mohamed, R. Mehdi, and M. A. Naaj, “Student course grade prediction using the random forest algorithm: Analysis of predictors’ importance,” *Trends Neurosci. Educ.*, vol. 33, 2023, Art. no. 100214.
[48] A. I. Adekitan and O. Salau, “The impact of engineering students’ performance in the first three years on their graduation result using educational data mining,” *Heliyon*, vol. 5, no. 2, 2019, Art. no. e01250.
[49] B. A. Breiki, N. Zaki, and E. A. Mohamed, “Using educational data mining techniques to predict student performance,” in *Proc. Int. Conf. Elect. Comput. Technol. Appl.*, 2019, pp. 1–5.
[50] S. Alturki and N. Alturki, “Using educational data mining to predict students’ academic performance for applying early interventions,” *J. Inf. Technol. Educ., Innov. Pract.*, vol. 20, pp. 121–137, 2021.
[51] D. Baneres, M. E. Rodríguez-Gonzalez, and M. Serra, “An early feedback prediction system for learners at-risk within a first-year higher education course,” *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 249–263, Apr.-Jun. 2019.
[52] R. Alcaraz, A. Martínez-Rodrigo, R. Zangróniz, and J. J. Rieta, “Early prediction of students at risk of failing a face-to-face course in power electronic systems,” *IEEE Trans. Learn. Technol.*, vol. 14, no. 5, pp. 590–603, Oct. 2021.
[53] D. Bañeres, M. E. Rodríguez, A. E. Guerrero-Roldán, and A. Karadeniz, “An early warning system to detect at-risk students in online higher education,” *Appl. Sci.*, vol. 10, no. 13, 2020, Art. no. 4427.
[54] S. Wiyono and T. Abidin, “Comparative study of machine learning KNN, SVM, and decision tree algorithm to predict student’s performance,” *Int. J. Res.-Granthaalayah*, vol. 7, no. 1, pp. 190–196, 2019.
[55] T. P. Vital, K. Sangeeta, and K. K. Kumar, “Student classification based on cognitive abilities and predicting learning performances using machine learning models,” *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 63–75, 2021.
[56] K. S. Rawat and I. V. Malhan, “A hybrid classification method based on machine learning classifiers to predict performance in educational data mining,” in *Proc. Int. Conf. Commun., Comput., Netw.*, 2019, pp. 677–684.
[57] G. Seni and J. Elder, *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*. San Rafael, CA, USA: Morgan & Claypool, 2010.
[58] X. Bai et al., “Educational big data: Predictions, applications and challenges,” *Big Data Res.*, vol. 26, 2021, Art. no. 100270.
[59] N. Tomasevic, N. Gvozdenovic, and S. Vranes, “An overview and comparison of supervised data mining techniques for student exam performance prediction,” *Comput. Educ.*, vol. 143, 2020, Art. no. 103676.
[60] C. Jalota and R. Agrawal, “Analysis of educational data mining using classification,” in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput.*, 2019, pp. 243–247.
[61] M. A. Widowati, A. F. Wibowo, and B. Warsito, “Comparative analysis on educational data mining algorithm to predict academic performance,” in *Proc. Int. Seminar Appl. Technol. Inf. Commun.*, 2021, pp. 173–178.
[62] R. Ghorbani and R. Ghousi, “Comparing different resampling methods in predicting students’ performance using machine learning techniques,” *IEEE Access*, vol. 8, pp. 67899–67911, 2020.
[63] S. Hussain, N. A. Dahan, F. Ba-Alwib, and R. Najoua, “Educational data mining and analysis of students’ academic performance using WEKA,” *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 9, pp. 447–459, 2018.

[64] E. B. Costa, B. Fonseca, M. A. Santana, F. F. De Araújo, and J. Rego, “Evaluating the effectiveness of educational data mining techniques for early prediction of students’ academic failure in introductory programming courses,” *Comput. Hum. Behav.*, vol. 73, pp. 247–256, 2017.
[65] D. Hooshyar, M. Pedaste, and Y. Yang, “Mining educational data to predict students’ performance through procrastination behavior,” *Entropy*, vol. 22, no. 1, 2020, Art. no. 1.
[66] R. M. Ahmed, F. Nahla, and A. A. Ali, “Predicting and analysis of students’ academic performance using data mining techniques,” *Int. J. Comput. Appl.*, vol. 182, pp. 1–6, 2018.
[67] A. Polyzou and G. Karypis, “Feature extraction for next-term prediction of poor student performance,” *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 237–248, Apr.–Jun. 2019.
[68] A. Hamoud, A. S. Hashim, and W. A. Awadh, “Predicting student performance in higher education institutions using decision tree analysis.,” *Int. J. Interactive Multimedia Artif. Intell.*, vol. 5, pp. 26–31, 2023.
[69] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, “Student academic performance prediction using supervised learning techniques,” *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 14, pp. 92–104, 2019.
[70] F. J. Kaunang and R. Rotikan, “Students’ academic performance prediction using data mining,” in *Proc. Int. Conf. Inform. Comput.*, 2018, pp. 1–5.
[71] Y. K. Saheed, T. O. Oladele, A. O. Akanni, and W. M. Ibrahim, “Student performance prediction based on data mining classification techniques,” *Nigerian J. Technol.*, vol. 37, no. 4, 2018, Art. no. 4.
[72] C. C. Gray and D. Perkins, “Utilizing early engagement and machine learning to predict student outcomes,” *Comput. Educ.*, vol. 131, pp. 22–32, 2019.
[73] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz, and K. Phasinam, “Classification and prediction of student performance data using various machine learning algorithms,” *Mater. Today, Proc.*, vol. 80, pp. 3782–3785, 2023.
[74] Y. Chen and L. Zhai, “A comparative study on student performance prediction using machine learning,” *Educ. Inf. Technol.*, vol. 28, no. 9, pp. 12039–12057, 2023.
[75] S. Chen and Y. Ding, “A machine learning approach to predicting academic performance in Pennsylvania’s schools,” *Social Sci.*, vol. 12, no. 3, 2023, Art. no. 118.
[76] B. K. Verma, N. Srivastava, and A. Kumar Bharti, “Advanced models for predicting student performance in higher education using machine learning and data mining techniques,” in *Proc. Int. Conf. Syst. Model. Advance. Res. Trends*, 2024, pp. 600–605.
[77] M. Yağci, “Educational data mining: Prediction of students’ academic performance using machine learning algorithms,” *Smart Learn. Environ.*, vol. 9, no. 1, pp. 1–19, 2022.
[78] I. E. Livieris, K. Drakopoulou, V. T. Tampakas, T. A. Mikropoulos, and P. Pintelas, “Predicting secondary school students’ performance utilizing a semi-supervised learning approach,” *J. Educ. Comput. Res.*, vol. 57, no. 2, pp. 448–470, 2019.
[79] A. Zohair and L. Mahmoud, “Prediction of student’s performance by modelling small dataset size,” *Int. J. Educ. Technol. Higher Educ.*, vol. 16, pp. 1–18, 2019.
[80] P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, “Educational data mining for predicting students’ academic performance using machine learning algorithms,” *Mater. Today, Proc.*, vol. 47, pp. 5260–5267, 2021.
[81] S. Rai, K. A. Shastry, S. Pratap, S. Kishore, P. Mishra, and H. A. Sanjay, “Machine learning approach for student academic performance prediction,” in *Evolution in Computational Intelligence*. Berlin, Germany: Springer, 2021.
[82] H. Al-Shehri et al., “Student performance prediction using support vector machine and K-nearest neighbor,” in *Proc. IEEE Can. Conf. Elect. Comput. Eng.*, 2017, pp. 1–4.
[83] B. Sekeroglu, K. Dimililer, and K. Tuncal, “Student performance prediction and classification using machine learning algorithms,” in *Proc. Int. Conf. Educ. Inf. Technol., Assoc. Comput. Machinery*, 2019, pp. 7–11.
[84] I. Khan, A. Ahmad, N. Jabeur, and M. Mahdi, “A conceptual framework to aid attribute selection in machine learning student performance prediction models,” *Int. J. Interactive Mobile Technol.*, vol. 15, pp. 4–19, 2021.
[85] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. V. Erven, “Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil,” *J. Bus. Res.*, vol. 94, pp. 335–343, 2019.
[86] G. Kostopoulos, S. Kotsiantis, and P. Pintelas, “Predicting student performance in distance higher education using semi-supervised techniques,” in *Model and Data Engineering*, L. Bellatreche and Y. Manolopoulos, Eds. Cham, Switzerland: Springer, 2015.
[87] I. E. Livieris, K. Drakopoulou, T. A. Mikropoulos, V. Tampakas, and P. Pintelas, “An ensemble-based semi-supervised approach for predicting students’ performance,” in *Research on E-Learning and ICT in Education: Technological, Pedagogical and Instructional Perspectives*, T. A. Mikropoulos, Ed. Cham, Switzerland: Springer, 2018.
[88] A. Kukkar, R. Mohana, A. Sharma, and A. Nayyar, “Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms,” *Educ. Inf. Technol.*, vol. 28, no. 8, pp. 9655–9684, 2023.
[89] H. Zeineddine, U. Braendle, and A. Farah, “Enhancing prediction of student success: Automated machine learning approach,” *Comput. Elect. Eng.*, vol. 89, 2021, Art. no. 106903.
[90] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, and D. J. Murray, “Identifying key factors of student academic performance by subgroup discovery,” *Int. J. Data Sci. Analytics*, vol. 7, no. 3, pp. 227–245, 2019.
[91] B. K. Francis and S. S. Babu, “Predicting academic performance of students using a hybrid data mining approach,” *J. Med. Syst.*, vol. 43, no. 6, 2019, Art. no. 162.
[92] O. R. Yürüm, T. Taşkaya-Temizel, and S. Yildirim, “The use of video clickstream data to predict university students’ test performance: A comprehensive educational data mining approach,” *Educ. Inf. Technol.*, vol. 28, no. 5, pp. 5209–5240, 2023.
[93] C. Huang, J. Zhou, J. Chen, J. Yang, K. Clawson, and Y. Peng, “A feature weighted support vector machine and artificial neural network algorithm for academic course performance prediction,” *Neural Comput. Appl.*, vol. 35, no. 16, pp. 11517–11529, 2023.
[94] P. Cortez and A. Silva, “Using data mining to predict secondary school student performance,” in *Proc. Eur. Concurrent Eng. Conf.*, 2008.
[95] F. M. Alnassar, “Predicting student performance on virtual learning environment,” Doctoral, Goldsmiths, University of London, 2023.
[96] E. N. Azizah, U. Pujianto, E. Nugraha, and Darusalam, “Comparative performance between C4.5 and naive Bayes classifiers in predicting student academic performance in a virtual learning environment,” in *Proc. Int. Conf. Educ. Technol.*, 2018, pp. 18–22.
[97] S.-U. Hassan, H. Waheed, N. R. Aljohani, M. Ali, S. Ventura, and F. Herrera, “Virtual learning environment to predict withdrawal by leveraging deep learning,” *Int. J. Intell. Syst.*, vol. 34, no. 8, pp. 1935–1952, 2019.
[98] M. Hlosta, Z. Zdrahal, and J. Zendulka Ouroboros, “Early identification of at-risk students without models based on legacy data,” in *Proc. Int. Learn. Analytics Knowl. Conf.*, 2017, pp. 6–15.
[99] J. Kuzilek, J. Vaclavek, V. Fuglik, and Z. Zdrahal, “Student drop-out modelling using virtual learning environment behaviour data,” in *Lifelong Technology-Enhanced Learning*, Berlin, Germany: Springer, 2018.
[100] S. Rizvi, B. Rienties, and S. A. Khoja, “The role of demographics in online learning; a decision tree based approach,” *Comput. Educ.*, vol. 137, pp. 32–47, 2019.
[101] M. Wasif et al., “Understanding student learning behavior and predicting their performance,” in *Cognitive Computing in Technology-Enhanced Learning*. Hershey, PA, USA: IGI Global, 2019, pp. 1–28.
[102] S. Gunasekara and M. Saarela, “Explainable AI in education: Techniques and qualitative assessment,” *Appl. Sci.*, vol. 15, no. 3, Jan. 2025, Art. no. 1239.
[103] Z. Wang, X. Zhu, J. Huang, X. Li, and Y. Ji, “International Educational Data Mining Society,” in *Proc. Int. Conf. Edu. Data Mining*, Raleigh, NC, Jul. 2018.
[104] J. Kuzilek, M. Hlosta, and Z. Zdrahal, “Open university learning analytics dataset,” *Sci. Data*, vol. 4, no. 1, 2017, Art. no. 170171.
[105] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[106] L. Zhang et al., “Data augmentation for sparse multidimensional learning performance data using generative AI,” *IEEE Trans. Learn. Technol.*, vol. 18, pp. 145–164, 2025.
[107] N. A. Jalil, H. J. Hwang, and N. M. Dawi, “Machines learning trends, perspectives and prospects in education sector,” in *Proc. Int. Conf. Educ. Multimedia Technol.*, 2019, pp. 201–205.
[108] U. Farooq et al., “Transforming educational insights: Strategic integration of federated learning for enhanced prediction of student learning outcomes,” *J. Supercomput.*, vol. 80, no. 11, pp. 16334–16367, 2024.
[109] Y.-W. Chu et al., “Multi-layer personalized federated learning for mitigating biases in student predictive analytics,” *IEEE Trans. Emerg. Topics Comput.*, 2024, to be published, doi: 10.1109/TETC.2024.3407716.

[110] A. P. Hridi, R. Sahay, S. Hosseinalipour, and B. Akram, "Revolutionizing AI-Assisted education with federated learning: A pathway to distributed, privacy-preserving, and debiased learning ecosystems," in *Proc. AAAI Symp. Ser.*, vol. 3, no. 1, 2024, pp. 297–303.
[111] A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, and I. I. Bittencourt, Eds., "International conference on artificial intelligence in education," in *Lecture Notes in Computer Sci.*, vol. 14830. Berlin, Germany: Springer, 2024.
[112] C. Zhan, S. Joksimović, D. Ladjal, T. Rakotoarivelo, R. Marshall, and A. Pardo, "Preserving both privacy and utility in learning analytics," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 1615–1627, 2024.
[113] J. Gardner and C. Brooks, "Student success prediction in MOOCs," *IEEE Trans. Learn. Technol.*, vol. 11, pp. 456–467, 2018.
[114] M. Berland et al., "Modeling student success in MOOCs with natural language," *IEEE Trans. Learn. Technol.*, vol. 9, pp. 123–135, 2016.
[115] Y. Luo, X. Han, and C. Zhang, "Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses," *Asia Pacific Educ. Rev.*, vol. 25, no. 2, pp. 267–285, 2024.
[116] P. Blikstein and M. Worsley, "Ethics of AI in education: Towards a community-wide framework," *Int. J. Artif. Intell. Educ.*, pp. 1–23, 2020.
[117] J. G. C. Krüger, A. D. S. Britto, and J. P. Barddal, "An explainable machine learning approach for student dropout prediction," *Expert Syst. Appl.*, vol. 233, 2023, Art. no. 120933.
[118] S. J. Pachouly and D. S. Bormane, "Exploring the predictive power of explainable AI in student performance forecasting using educational data mining techniques," in *Recent Advances in Sciences, Engineering, Information Technology & Management*. Boca Raton, FL, USA: CRC Press, 2025, pp. 362–370.
[119] A. Pardo, F. Han, and R. A. Ellis, "Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 82–92, Jan.–Mar. 2017.
[120] S. Kleter, U. Matzat, and R. Conijn, "Boundary conditions of generalizing predictive models for academic performance: Within cohort versus within course," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 2129–2140, 2024.
[121] A. Zanellati, S. P. Zingaro, and M. Gabbrielli, "Balancing performance and explainability in academic dropout prediction," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 2086–2099, 2024.
[122] C. Cui et al., "Tri-branch convolutional neural networks for top-k focused academic performance prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 439–450, Jan. 2024.
[123] J. So et al., "Analytic study for predictor development on student participation in generic competence development activities based on academic performance," *IEEE Trans. Learn. Technol.*, vol. 16, no. 5, pp. 790–803, Oct. 2023.
[124] W. Qiu, A. W. H. Khong, S. Supraja, and W. Tang, "A dual-mode grade prediction architecture for identifying at-risk students," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 803–814, 2024.
[125] N. D. Lynn and A. W. R. Emanuel, "Using data mining techniques to predict students' performance. A review," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1096, no. 1, 2021, Art. no. 012083.
[126] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, 2021, Art. no. 237.
[127] M. Kumar, A. J. Singh, and D. Handa, "Literature survey on student's performance prediction in education using data mining techniques," *Int. J. Educ. Manage. Eng.*, vol. 7, no. 6, pp. 40–49, 2017.
[128] S. D. Abdul Bujang et al., "Imbalanced classification methods for student grade prediction: A systematic literature review," *IEEE Access*, vol. 11, pp. 1970–1989, 2023.
[129] H. Roslan and C. J. Chen, "Educational data mining for student performance prediction: A systematic literature review (2015–2021)," *Int. J. Emerg. Technol. Learn.*, vol. 17, pp. 147–179, 2022.
[130] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, 2021, Art. no. 552.
[131] J. López-Zambrano, J. A. L. Torralbo, and C. Romero, "Early prediction of student learning performance through data mining: A systematic review," *Psicothema*, vol. 33, no. 3, pp. 456–465, 2021.
[132] P. Balaji, S. Alelyani, A. Qahmash, and M. Mohana, "Contributions of machine learning models towards student academic performance prediction: A systematic review," *Appl. Sci.*, vol. 11, no. 21, 2021, Art. no. 10007.
[133] A. Hellas et al., "Predicting academic performance: A systematic literature review," in *Proc. Companion Annu. ACM Conf. Innov. Technol. Comput. Sci. Educ.*, 2018, pp 175–199.
[134] A. Abu Saa, M. Al-Emran, and K. Shaalan, "Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques," *Technol., Knowl. Learn.*, vol. 24, no. 4, pp. 567–598, 2019.
[135] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student's performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, 2021, Art. no 552.
[136] S. Ahmad, M. A. El-Affendi, M. S. Anwar, and R. Iqbal, "Potential future directions in optimization of students' performance prediction system," *Comput. Intell. Neurosci.*, vol. 2022, 2022, Art. no. e6864955.

Jialun Pan received the B.E. degree in automotive engineering from the Wuhan University of Technology, Wuhan, China, in 2022, and the M.S. degree in mechanical and automation engineering from the Chinese University of Hong Kong, Hong Kong, in 2023.

He is currently a Research Assistant with the Department of Mechanical and Automation Engineering, Chinese University of Hong Kong.

Zhanzhan Zhao received the bachelor's degree in flight vehicle design and engineering from the Harbin Institute of Technology, Harbin, China, in 2016, the master's degree in autonomous systems and control from the University of Michigan, Ann Arbor, MI, USA, in 2018, and the Ph.D. degree in machine learning from the Georgia Institute of Technology, Atlanta, GA, USA, in 2023.

She was a Postdoctoral Researcher at the Mathematical Sciences Research Institute, Berkeley, CA, USA, before continuing her work at Boston Childrens Hospital and Harvard Medical School. She is currently an Assistant Professor in computational social science with The Chinese University of Hong Kong (Shenzhen), Hong Kong. Her research focuses on algorithmic governance, artificial intelligence, and its societal implications.

Dongkun Han received the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2014.

He was a Research Fellow at the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI, USA. He was a Research Associate at the Department of Informatics, Technical University of Munich, Munich, Germany. He was an Exchange Research Student at the Information Systems Laboratory, Stanford University, Stanford, CA, USA. He was a Visiting Academic Faculty at the Department of Electrical Engineering, German Institute of Science and University of Munich Asia, Singapore. He is currently a Lecturer with the Department of Mechanical and Automation Engineering, Chinese University of Hong Kong, Hong Kong. His research interests include optimization and control of complex systems, with applications in education, power grids, and robotic networks.