ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

Research on the Application of Machine Learning in the Educational and Educational Management of Colleges and Universities: A Case Study of Student Academic Early Warning and Intervention System

Yonglin Zhao[1,*]

[1] Academic Affairs Office, Zhengzhou University of Science and Technology, Zhengzhou 450064, Henan, China

SUMMARY: *This study critically examines the application of machine learning technologies in higher education and academic management, with particular emphasis on their use as early warning and intervention systems for identifying students at risk of academic dropout. Employing a case study approach, the research focuses on a Chinese public-sector university recognised as a pioneer in the adoption of digital learning technologies and learning management systems (LMS). The empirical analysis is based on observational data and academic records of 141 undergraduate (BS) students enrolled across multiple programmes and disciplines. The results indicate that approximately 17% of the analysed students were classified as at risk of dropout. Risk identification was conducted using machine learning models that integrated traditional academic indicators, such as cumulative GPA and current GPA, alongside behavioural and engagement-related variables. Key behavioural factors included assignment submission timeliness, attendance rates, LMS login frequency, and completed credit hours, with higher values in these indicators corresponding to a lower likelihood of dropout. Additionally, student engagement, for which online forum participation benchmark has been adopted in this study, reflects a better assessment of levels of academic engagement, as students exhibiting higher levels of online interaction with peers and instructors demonstrated a reduced risk of attrition. Among the various machine learning models evaluated, XGBoost emerged as the most effective in predicting at-risk students, achieving superior accuracy and precision compared to alternative approaches.*

KEYWORDS: *Machine Learning; Early Warning Systems; Student Dropout Prediction; Higher Education Management Dropout Predictions; XGBoost Model*

1 Introduction

Digital transformation and technological changes taking place these days have been creating immense opportunities in different sectors [1]. The Chinese education sector is no exception, as significant digital transformation has been taking place within the sector. In particular, after the initiation and execution of the 'Education Informatisation Policy 2.0 Action Plan' pursued by the Chinese government, best practices found in different parts of the world have been introduced in the Chinese education sector [2]. This includes the use of artificial intelligence (AI) technology, which has not been leveraged for the sake of teaching, but the technology is widely adopted for diverse objectives, including research, learning management and

*zhaoyonglin791502@126.com
https://doi.org/10.65102/is20261227
Published: 10 June, 2026

institutional governance [3].

An overwhelming number of Chinese universities and colleges have been benefiting from the learning management system (LMS), which has been introduced with the aim of enhancing the quality of education in the country [4]. LMS has not just been used as a student information system, but also increasingly, it has been taking the form of campus data centres, which in turn has been facilitating decision-making processes, leading to the better realisation of the quality education objectives that educational institutes are pursuing these days [5]. Improvement in this regard has been largely realised through the use of machine learning technology, where LMS systems are integrated with the machine learning, which carefully analyses trends and data and generates important reports and intelligence that could have otherwise remained unnoticed [6].

Traditionally, educational institutes have been relying on the teacher-centric evaluation of student learning activities, where the teacher used his/her subjective judgement coupled with marks scored by the student in different exams and retrospective evaluation [7]. However, the process thus used is considered an outdated system, as such a system led to delayed feedback, which many times resulted in students' academic failure and ultimate dropout. Unlike such a traditional approach, machine learning technology is increasingly incorporated by leading educational institutions as an 'early warning and intervention system' [8]. Due to the rapid adoption of digital technologies, opportunities are emerging through the use of machine learning, as the technology could be effectively adopted for proactive insight that could support effective and timely feedback [9]. Machine learning technology could not only analyse complex datasets, but also the technology could be used to more effectively predict and forecast, which may lead to evidence management of student learning activities [10]. In particular, [11] are of the view that machine learning technology has been adopted in student academic early warning and intervention systems, enabling educational institutes to make timely interventions for at-risk students. The authors have pointed out that machine learning could analyse data like attendance records, historical performance of students, participation in online discussions and online learning behaviour to identify patterns and recommend timely intervention strategies.

As Chinese educational institutes are confronting an increasing pressure regarding quality education, various stakeholders are debating the administrative efficiency. It is believed that the use of machine learning technology offered promising opportunities in terms of administrative efficiency, as it could certainly improve educational management activities [12]. In particular, the technology could be adopted for generating predictive insight, thus better developing early warning and intervention systems [13]. In this regard, a significant amount of research has been conducted in different parts of the world that stressed the efficiency of machine learning for better educational management; however, empirical evidence in this regard from China is lacking. This research aimed to critically analyse the application of machine learning in the educational and educational management of colleges and universities and how the technology could be leveraged as an early warning and intervention system for student academic learning. The research identifies how machine learning could be leveraged for identifying key student behaviour, academic, and engagement features that influence student academic performance, thus predicting at-risk students early in the semester.

2 Materials and Methods

2.1 Research Strategy

The findings regarding the application of machine learning in the educational and educational management of colleges and universities and how the technology could be leveraged as an early warning and intervention system for student academic learning are based on a case study

approach. This means that the findings of the research are based on an actual and real-life educational institute, whose name has been anonymised in this research. The university is a public sector university operating in a large metropolitan city of the country and having more than 10,000 enrolments each year. The university has the credit for executing one of the early online programmes in the country for system and learning management systems, which provide online learning support to students. In addition to LMS, the university also has a Student Information System and a digital attendance and card-swipe system. The university officials agreed to provide the desired data on the condition that full anonymity and confidentiality of the university will be maintained throughout the research.

2.2 Participants and Sample of the Study

Although there are more than 10,000 students studying in the selected university, the data for a carefully selected sample has been selected and analysed in the study. The study is based on undergraduate students enrolled in the selected university, who were enrolled in one of the BS programmes for the academic sessions 2021-2024. The major inclusion criteria for the participants' inclusion in the study were that the individuals should be full-time undergraduates who had a full record in the SIS and LMS of the university. Like inclusion criteria, exclusion criteria have also been observed, as students who were exchange students or had incomplete records or were studying in a non-credit programme of the university have been excluded from the study. A total of 144 students who satisfy the inclusion and exclusion criteria were thus selected for this study, whose data have been used for generating findings regarding the application of machine learning in the educational and educational management of colleges and universities and how the technology could be leveraged as an early warning and intervention system for student academic learning.

2.3 Data Sources

Data has been collected from the university that has been selected for this study. Data has been largely collected regarding dependent and independent variables under consideration. The major dependent variable in this study is academic performance, which has been used to identify students at risk. The variable has been defined as a binary variable, where students having a GPA of < 2.0 have been categorised as ‘at-risk’ students, while students scoring > 2.0 GPA have been defined as ‘not-at-risk’ students. On the other hand, there are different independent variables in this study, which include students’ demographics, academic history, attendance metrics, LMS engagement, and behavioural patterns. The core data that has been selected in this regard include academic records, attendance records, LMS behavioural data, and intervention logs.

2.4 Data Analysis

There are different data analysis techniques used in this study. This includes the descriptive statistical analysis process, including means, standard deviation, and frequency calculation process. Furthermore, inferential statistical techniques have been widely adopted in this study. This includes Pearson correlation analysis that has been largely used to analyse the relations between different variables under consideration. The following formula has been used for the sake of calculating the Pearson’s bivariate correlation analysis (r):

In the above formula:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \cdot \sum(Y_i - \bar{Y})^2}}$$

In the above formula, Xi denotes individual values variable X, Yi denotes the individual values of variable Y, $\bar{X}$ denotes mean of variable X, and $\bar{Y}$ denotes the mean score of variable Y.

3 Results

3.1 Demographic Profile of the Participants Observed

The demographic characteristics of the participants of the students are analysed and summarised in the following Table 1:

Table 1: Demographic Profile of the Participants of the Research

Description	Variable	Number	Percentage
Gender	Male	75	52.1%
	Female	69	47.9%
Age	Less than 19 Years	28	19.4%
	20 to 21 Years	82	56.9%
	More than 22 Years	34	23.6%
Year of Study	Year 1	36	25.0%
	Year 2	42	29.2%
	Year 3	39	27.1%
	Year 4	27	18.8%
Academic Major	Humanities and Social Sciences	62	43.1%
	Science and Engineering	48	33.3%
	Business and Economics	34	23.6%

From the analysis of demographic data exhibited in the above Table 1, it is clear that the participants observed in this study have been carefully selected, as individuals sharing diverse backgrounds have been selected for the study. In terms of gender, about 52% of the participants whose performance has been observed in the study are male, while 48% of the participants observed were female students. In terms of age, the observation was made for students having different ages. In this regard, about 19% of the participants observed were less than 19 years old, 57% were 20 to 21 years old, while 24% were more than 22 years old. Furthermore, the students whose performance has been observed were studying at different levels in the selected university, which include 25% students in the 1st year of BS, 29% in the second year, 27% in the third year and 19% in the final year of their BS programme at the selected university. Additionally, the participants observed in the study were studying in different programmes. In this regard, 43% of the selected students were in the humanities and social sciences disciplines, 33% were in the science and engineering disciplines, while about 24% of the students observed were studying in the BS business and economics.

3.2 Descriptive Statistics of the Academic Performance of Students

The observations regarding the academic performance and engagement of the students have been summarised in the following Table 2:

Table 2: Descriptive Statistics of the Academic Performance of Students

Description	Min	Max	Mean	Standard Deviation
Prior Semester GPA	1.50	4.0	2.78	0.56
Current Semester GPA	1.40	4.0	2.85	0.58
Credit Hours Completed	18	72	42.3	11.7
Attendance percentage	60	100	88.2%	9.1
LMS Login (Per Week)	1	15	7.6	3.2
LMS Time-on-Task (Hours Per Week)	1.2	11.3	5.4	2.1
Assignment on-Time Submission %	50	100	85.1%	10.4
Forum Participation (Post Per Semester)	0	35	12.3	8.2

From the analysis of the above Table 2, there is different information that is worth noting. In this regard, the prior GPA secured by the students exhibits the potential risks, as a lower GPA in the previous semesters increased the likelihood of students being at risk of dropout. The mean calculated in this study is 2.78, which indicates that the majority of the students whose records have been observed in this study were not at risk of dropping out, as they had better GPA scores in the previous semesters. The current GPA of the students is more stable than the past record, as the mean score calculated for the selected 141 students in this study is 2.85, which surfaces the prior semester GPA. Using the binary classifications in this regard, 17% of the total participants observed (24 out of 141 students) in this study had a GPA of less than 2.0, indicating that these students were at risk, while the rest of the students whose performance has been analysed in this study were not at risk. The students who are at risk include both moderate-risk and high-risk students. Students could be classified as low-risk, moderate-risk and high-risk students. Using these three classifications, Figure 1 classifies the 141 students observed in this study into three categories.

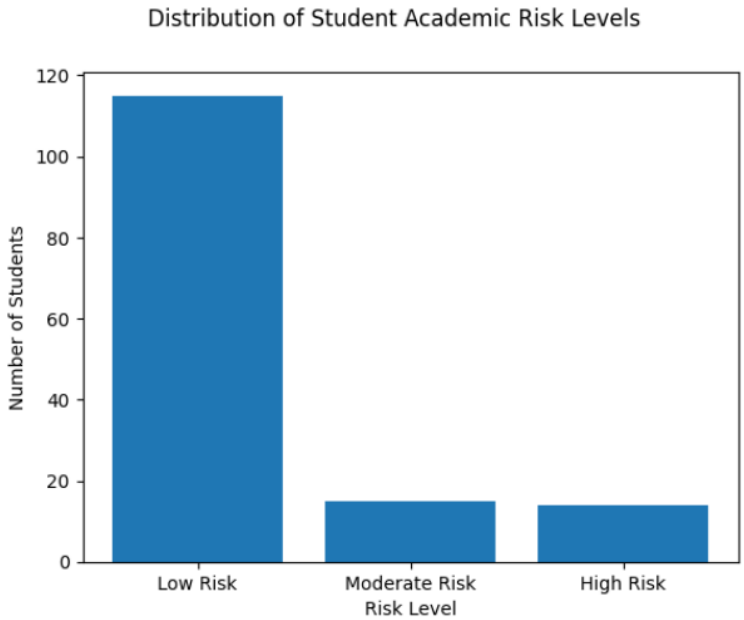

Figure 1: Distribution of Student Academic Risk Levels

Furthermore, in terms of the average credit hours of study that the participants of the study have completed, it is 42.3 hours. On average, the selected 141 students were logging into the system 7.6 times a week, and on average, they were completing about 5.4 tasks a week,

indicating more stable LMS performance. The students who were able to submit their assignments on time were about 85.1% of the total number of students observed in the study. Furthermore, the students were also having better engagement on the LMS, for which the discussion forum average of the students has been observed. Of the selected students, 12.3 posts were made on a per-semester basis, indicating above-average engagement. Usually when the students have a low degree of engagement and participation in the online forum discussions on LMS, this increases their risks of dropout. From the analysis of these different values, although student GPA is the primary determinant of the students at risk, the other four elements identified in this study also have important bearings, particularly in the early identification of students at risk.

3.3 Pearson's Correlations Analysis

The following Table 3 highlights Pearson coefficient analysis of the key construct of the study, exhibiting the relations between different variables.

Table 3: Pearson's Correlations Analysis

Description	Academic Risk	Prior GPA	Attendance	LMS Login	Assignment on-time
Academic Risk	1	-0.62**	-0.48**	-0.42**	-0.55**
Prior GPA	-0.62**	1	0.36**	0.31**	0.49
Attendance Rate	-0.48**	0.36**	1	0.28**	0.42**
LMS Logins	-0.42**	0.31**	0.28**	1	0.30
Assignment on-Time	0.55**	0.49**	0.42**	0.30**	1

The above Table 3 highlights the Pearson coefficient, indicating the academic risks that the selected 141 students have been encountering and the core categories accounting for the risks encountered. In this regard, the academic risks encountered by the students are negatively correlated with the GPA scored by the students. This in turn means that as the students scored a large GPA, this in turn decreased their probability of dropping out from the university. Similarly, the attendance rate and the assignment on-time are the other core factors that also reflect negative correlations. On the other hand, the LMS login history of the students, which indicates students' engagement, has also been found to be statistically negatively correlated with the academic risks that the students encountered in the current study.

3.4 Machine Learning Model Performance

As there are different models of machine learning that could be adopted to predict students' dropout, the accuracy of these models needs to be established first. The following Table 4 computes different statistics that could be used to analyse machine learning model performance.

Table 4: Machine Learning Model Performance

Description	Accuracy	Precision	Recall	F1-Score	AUC
Logistics Regression	0.81	0.70	0.76	0.73	0.84
Random Forest	0.86	0.78	0.82	0.80	0.90
LightGBM	0.87	0.79	0.84	0.81	0.91
XGBoost	0.88	0.81	0.85	0.83	0.92

The above Table 4 compares and contrasts different models that could be used for predicting students at risk. The logistic regression model is the traditionally used model, which has an accuracy of 81%. This could be considered as a baseline; thus, the machine learning models that have an accuracy of below 81% could not be considered as an alternative. There are three different machine learning models that have been analysed in this study for the sake of accuracy, which include Random Forest, LightGBM and XGBoost. Of these three models, XGBoost emerged as the best alternative, as it has far better accuracy than other competing machine learning models as well as the traditionally used logistic regression model. The accuracy of XGBoost is 88%, with 81% accuracy, 85% recall, 83% F1-score and 92% AUC. The AUC score of the XGBoost is of particular significance, as the machine learning model could accurately predict student dropout with 92% accuracy. This in turn indicates the superior capabilities of XGBoost to effectively identify the students at risk of dropout, as the model could be used as an early warning system in the university, thus preventing the circumstances that could lead towards student dropout. Furthermore, unlike the traditional regression model, the machine learning models have greater effectiveness, which in turn suggests the superiority of machine learning models to effectively identify and predict student dropout in a complex environment found in educational institutes. Using the XGBoost model, students who are at risk could be identified. The ROC curve of the XGBoost model has been exhibited in the following Figure 2:

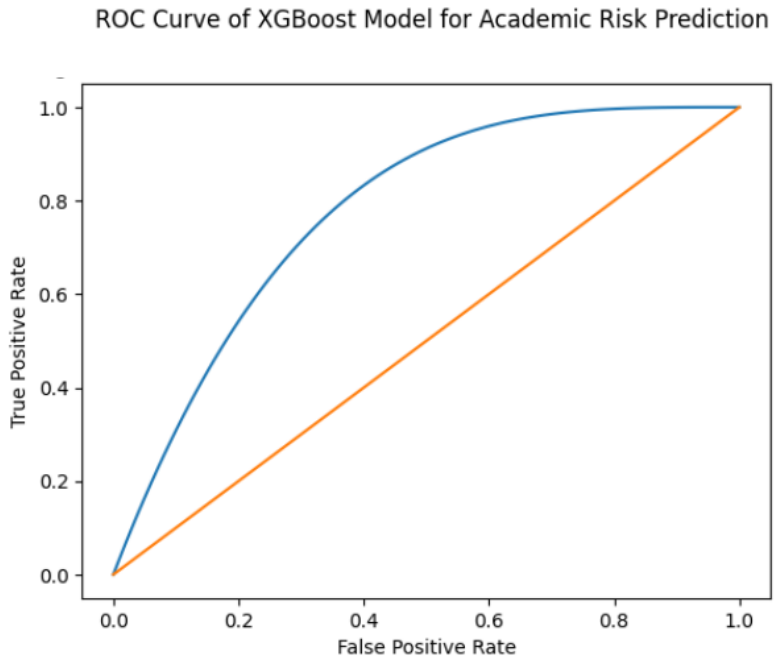

Figure 2: ROC Curve of XGBoost Model for Academic Risks Prediction

3.5 Interventions System

Taking into consideration the challenges regarding the students thus identified, an intervention process has been initiated by the university officials. The intervention strategies executed by different departments largely depend on the risk level that the students were encountering. The intervention strategies include email and short-message reminders for the low-risk students, advisor check-ins for the moderate-risk students and additional tutoring and counselling for the students categorised as high-risk students. The result of the intervention strategies thus executed by the university has been summarised in the following Table 5:

Table 5: Intervention Strategies and their Impact

Description	Number of Students	Mean GPA Before	Mean GPA After	Change in GPA	% Improvement
Low-Risk Students	15	2.50	2.60	+0.10	67%
Moderate Risk Students	12	2.10	2.40	+0.30	80%
High-Risk Students	12	1.80	2.30	0.50	86%

Of the different intervention strategies analysed and summarised in Table 5 that the selected university official initiated, the counselling and tutoring strategies emerged as the most effective strategy as 86% improvement in the GPA has been witnessed as a result of executing the strategies in the high-risk students, thereby significantly cutting down the risks of dropout that such students were confronting.

4 Discussions

4.1 At-Risk Students

This study found that out of the 141 students whose academic performance and different benchmarks were observed in this study, 17% of the students could be classified as at-risk students, which includes both moderate-risk students and high-risk students. If an immediate intervention strategy is not initiated for these students, they are confronting future dropout in the selected university. From the analysis presented in this study, it is evident that students at risk could not be determined using a single benchmark; rather, a range of factors have to be considered while determining such students at risk. In this regard, the traditional logistics correlation benchmarks have been traditionally used by academic institutes; however, such a logistics correlation process could be used for determining linear relations, while increasingly complex variables need to be considered while determining students at risk. In this regard, machine learning technology has been emerging as a potential solution, as it could use a multitude of factors, including historical figures, as well as multiple analysis and prediction processes that could more effectively determine the students at risk in academic institutions. Although there are different machine learning models that could be adopted in this regard, the present study found that XGBoost emerged as the best model that could be adopted because of its greater precision, accuracy, F-1 score and AUC.

4.2 Key Determinants of At-Risk Students

There are different variables that could be used to identify students who are at risk. Traditionally, academic performance exhibited in the form of current and prior GPA has been considered as the core determinant of students at risk. However, this study identifies that in addition to the academic performance, there are other variables that need to be considered, whereas behavioural engagement emerged as the primary determinant of students at risk in academic institutions. There are different key variables identified for identifying students at risk, which include punctuality of assignment submission, attendance rate, and participation in the LMS activities, particularly online discussions on forums. These factors go beyond the traditional academic performance factor and could be termed as more dynamic factors that shaped the contemporary and future learning behaviour of students. The range of factors identified in the present study are of more importance, as these factors reflect the changing digital educational environment in which students these days are participating. The range of factors identified in

the current study could more effectively predict the future activities of students and could identify the students at risk, who could be potential dropouts. On the basis of appropriate identification of different behavioural attributes, an appropriate intervention strategy could be launched by an educational institution to overcome the risks of dropouts that such students are encountering. The findings of the current study thus stand in line with the findings of [14] and [15], who have also stressed the behavioural and engagement variables in their studies as the primary factors that could be used to identify students at risk.

4.3 Predictive Powers of Machine Learning Models

Unlike the traditional statistical correlations model, this study found that machine learning models have far better predictive accuracy, besides having the capability to leverage multidimensional factors to predict students at risk. In this regard, three different machine learning models were considered in the present study, which include Random Forest, LightGBM, and XGBoost. Of these three models, the XGBoost emerged as the best model due to its relatively higher accuracy, precision, recall, F-1 score and AUC. Unlike other machine learning models, XGBoost is based on ensemble learning, which is considered a more suitable model that could more effectively predict non-linear relations that are based on behavioural and engagement variables. Through the machine learning model, more effective and accurate predictions of students at risk could be conducted through the XGBoost model. The findings of the current study in this regard are in line with the findings of [16] and [17], who have also stressed that XGBoost has stronger discriminatory capabilities to identify students at risk in the academic institutions using multiple variables.

5 Conclusion

This study critically analysed the application of machine learning technology in the educational and educational management of colleges and universities as an early warning and intervention system for student academic learning by identifying students at risk of dropout. The findings of the study are based on a case study for which a Chinese public-sector university has been selected, which is considered one of the pioneer universities in the execution of digital learning technologies and LMS systems. The findings of the study are based on the observation and analysis of the record of 141 BS students who were students in different programmes and disciplines in the university. The study found that out of the total students analysed, 17% of the total students were at risk. The identification in this regard is based on machine learning technology, whereas the traditional GPA, current GPA, and a range of behavioural and student engagement factors have been considered in the identification of such at-risk students within the observed students. Key behavioural factors considered in this regard include assignment on time, attendance rate, LMS logins and credit hours completed. Higher figures in these different behavioural factors indicate lower risks of dropouts and could thus be used for more effectively predicting the at-risk students. On the other hand, engagement factors like online forum participation could be used to predict the online engagement of students. Students who are usually having more online interaction with classmates and teaching faculty usually confront lower risks of dropping out as compared to students who are having lower online interaction. Furthermore, the study found that although there are different machine learning models that could be adopted for predicting students at risk using a multitude of factors, XGBoost emerged as the best model in this regard due to its higher precision and accuracy.

Funding

This research was supported by the Postgraduate Education Reform Project of Henan Province (Grant No.2023SJGLX382Y).

About the Author

Yonglin Zhao was born in1978 in Xinyang City, Henan Province, China. He earned a Master's degree from Henan University. He is currently employed in the Academic Affairs Office at Zhengzhou University of Science and Technology. His research focus is on educational management. E-mail: zhaoyonglin791502@126.com

References

[1] A. Abdulkadir, A. Alzubi and O. R. Adegboye, “Intelligent system for student performance prediction: An educational data mining approach using metaheuristic-optimized LightGBM with SHAP-based learning analytics,” *Applied Science, 10875,* 15(20), 20225.

[2] J. Pan, Z. Zhao and D. Han, “Academic performance prediction using machine learning approaches: A survey,” *IEEE Transactions on Learning Technologies,* 99, pp. 1-18, 2025.

[3] K. Alawali, R. Athauda, R. Chiong and I. Rnner, “Evaluating the student performance prediction and action framework through a learning analytics intervention study,” *Education and Information Technologies,* 30, pp. 2887-2916, 2025.

[4] M. Gul, W. Abbasi and m. Z. Baber, “Data driven decisions in education using a comprehensive machine learning framework for student performance prediction.,” *Discover Computing, 9585,* 153, 2025.

[5] E. Ahmad, “Student performance prediction using machine learning approaches,” *Journal of Computer Science and Technology, 4067721,* 2024.

[6] A. Angioplasties, J. Aliparntis, M. Konstandkis and A. Tsimpiris, “Predicting student performance and enhancing learning outcomes: A data-driven approach using educational data mining techniques,” *Computers, 83,* 14(3), 2025.

[7] M. N. Gul, W. Abbasi and M. Wani, “Revolutionizing educational decision-making: A robust machine learning mechanism for predicting student performance,” *Journal of Electrical Systems and Information Technology, 32,* 12, 2025.

[8] M. Yagci, “Educational data mining: Prediction of students' academic performance using machine learning algorithms,” *Smart Learning Environments,* 9(11), 2022.

[9] N. Junejo, M. Nawaz, Q. Huang, X. Dong, C. Wang and G. Zheng, “Accurate multi-category student performance forecasting at early stages of online education using neural networks,” *Scientific Reports, 16251,* 15, 2025.

[10] A. Turkmenbayev, E. Abdykermiova, S. Nurgzhayev, G. Karabassova and D. Baigozhanova, "The application of machine learning in predicting student performance in university engineering programs: A rapid review," *Frontiers in Education, 1562586,* 10, 2025.

[11] "Predicting student dropouts with machine learning: An empirical study in Finnish higher education," *Technology in Society, 102474,* 76, 2024.

[12] M. R. Marcelino, R. R. Porto, T. T. Primo, R. Targino, V. Ramos, E. M. Queiroga, R. Munoz and C. Cechinel, "Student dropout prediction through machine learning optimization: insights from Moodle log data," *Scientific Reports, 9840,* 15, 2025.

[13] J. Sandivar-Rosas, W. J. Marin-Rodriguez, E. Toro-Dextra and H. Villarreal-Torres, "Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review," *EAI Endorsed Transactions on Scalable Information System,* 10(5), 2023.

[14] M. A. Hassan, A. H. Muse and S. Nadarajah, "Predicting Student Dropout Rates Using Supervised Machine Learning: Insights from the 2022 National Education Accessibility Survey in Somaliland," *Applied Science, 7593,* 14(17), 2024.

[15] G. Davila, J. Haro, H. Gonzalez-Eras, O. R. Vivanco and D. G. Coronei, "Student Dropout Prediction in High Education, Using Machine Learning and Deep Learning Models: Case of Ecuadorian University," *International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA,* pp. 1677-1684, 2023.

[16] Y. N. Mnywami, H. M. Maziku and J. C. Mushi, "Enhanced Model for Predicting Student Dropouts in Developing Countries Using Automated Machine Learning Approach: A Case of Tanzanian's Secondary Schools," *Applied Artificial Intelligence, 2071406,* 36(1), 2022.

[17] M. Varma and H. Li, "Predicting student dropouts with machine learning: An empirical study in Finnish higher education," *Technology in Society,* 76(C), 2024.