Cogent Education

ISSN: 2331-186X (Online) Journal homepage: www.tandfonline.com/journals/oaed20

Construction and application of evaluation models for online learning behaviour

Zhaohui Zheng , Kemin Hu , Zhongyuan Guo & Deng Cheng

To cite this article: Zhaohui Zheng , Kemin Hu , Zhongyuan Guo & Deng Cheng (2026) Construction and application of evaluation models for online learning behaviour, Cogent Education, 13:1, 2707713, DOI: 10.1080/2331186X.2026.2707713

To link to this article: https://doi.org/10.1080/2331186X.2026.2707713

Published online: 28 Aug 2026.

COGENT EDUCATION
2026, VOL. 13, NO. 1, 2707713
https://doi.org/10.1080/2331186X.2026.2707713

INFORMATION & COMMUNICATIONS TECHNOLOGY IN EDUCATION | RESEARCH ARTICLE

OPEN ACCESS Check for updates

Construction and application of evaluation models for online learning behaviour

Zhaohui Zheng[a], Kemin Hu[a], Zhongyuan Guo[b*] and Deng Cheng[c]

[a]School of Mathematics and Physics, Wuhan Institute of Technology, Wuhan, China; [b]School of Electronic and Information Engineering, Southwest University, Chongqing, China; [c]School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China

ABSTRACT

Online learning not only makes full use of online resources to meet the diverse learning needs of students but also enhances their autonomy and information acquisition abilities. Through learning analysis techniques, the quality of online learning can be improved, thereby enabling evaluation, diagnosis, prediction, and intervention in learning. Based on the Ulearning platform, this study used 792 undergraduate course score data from 2023, divided students' online learning behaviours into different indicators, and performed exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) on the students' online learning data to verify the feasibility of the learning evaluation model indicators. Then, through the demand analysis of teaching decisions, the dataset was split by K-fold cross-validation, a nonlinear support vector machine (SVM) was used to optimize the hyperparameters to establish an online learning behaviour evaluation model. The experimental results show that factor analysis effectively extracts three core factors (Platform usage intensity, Learning behavior, Effective learning intensity) affecting academic performance and greatly reduces data dimensionality. The proposed model achieves a high coefficient of determination (R^2) of 0.960 and a low root mean square error (RMSE) of 0.194, demonstrating outstanding predictive performance and strong generalization ability. This method provides a reference basis for further improving students' online learning effectiveness and can be extended to student modeling in other types of learning activities. More importantly, it can capture the interdependence between different aspects of students' learning experiences and learning outcomes, thereby guiding teachers to better carry out teaching work, make early interventions, and allocate resources more effectively.

ARTICLE HISTORY
Received 17 November 2025
Revised 13 July 2026
Accepted 18 July 2026

KEYWORDS
Online learning behaviour; learning effectiveness; factor analysis; SVM; evaluation model

SUBJECTS
Arts & Humanities; Arts; Media Communication; Humanities; Media & Film Studies; Media & Communications; Media Education; Social Sciences; Behavioral Sciences; Educational Psychology

1. Introduction

With the rapid development of social science and technology, big data have exerted a significant influence on contemporary teaching methods. Teaching and internet technologies are merging at an accelerated pace, leveraging the diverse array of online resources and vastness of opportunities presented by such technologies, and learning courses on online platforms have become an indispensable part of contemporary teaching processes, which offers substantial developmental potential for both learning and teaching. According to recent global education reports, over 70% of higher education institutions now incorporate some form of online learning, with student enrollment in digital courses increasing by approximately 15% annually. However, despite this widespread adoption, completion rates for massive open online courses (MOOCs) often fall below 10%, and nearly 60% of students report difficulties in maintaining consistent engagement throughout online courses (Allen & Seaman, 2017; Jordan, 2019).

CONTACT Kemin Hu 2220210315@stu.wit.edu.cn School of Mathematics and Physics, Wuhan Institute of Technology, Wuhan 430079, China; Zhongyuan Guo guozhongyuan@swu.edu.cn School of Computer Science and Engineering, Chongqing University of Science and Technology, Chongqing, 400715, China
*Present address: School of Computer Science and Engineering, Chongqing University of Science and Technology, Chongqing, China.

However, the challenges that students face in online learning include: maintaining focus and motivation can be more difficult in a virtual classroom setting. The lack of face-to-face interaction may affect peer learning and the development of social skills; Online learners must develop strong self-regulation skills to manage their time and learning effectively (Bernard et al., 2004; Broadbent & Poon, 2015). To address these challenges, it is necessary to analyze the effectiveness of online learning to guide teachers in further optimizing teaching strategies and improving teaching quality. By delving into students' online learning behaviour data, teachers can identify key factors that affect learning outcomes, adjust teaching content and methods accordingly, enhance teaching interaction, improve the learning environment, and provide personalized learning support to ensure that every student can achieve the best learning experience and results in online learning. The integration of online education with emerging technologies such as artificial intelligence and big data has led to the era of 'intelligence + education' (Fu et al., 2023). Many researchers have incorporated big data and artificial intelligence technologies into learning analysis, leading to the emergence of numerous learning analysis techniques and tools and providing an effective foundation for the development of future intelligent education (Bosch, 2021; Sun et al., 2021; Stojanov & Daniel, 2024).

To understand the current landscape of learning analytics, existing research has formed a relatively systematic methodological framework. The first stream focuses on prediction modeling using traditional machine learning, with a significant body of research highlighting the efficacy of Support Vector Machines (SVM). Multiple studies have consistently concluded that SVM is well-suited for small-sample and nonlinear data scenarios in the educational domain, demonstrating superior predictive performance for academic grades and learning early warnings (Yu et al., 2020; Jin & Zhuo, 2020; Zhang et al., 2017; Wang, 2022). However, these studies also reveal a common limitation of SVM: its weak adaptability to multi-class classification, which necessitates decomposing tasks into multiple binary classifications for practical implementation. To address these constraints, recent research has supplemented traditional single-classifier approaches with integrated models, verifying that ensemble methods possess stronger capabilities for educational data mining and staged prediction of at-risk students (Luo et al., 2024; Li et al., 2022). Overall, these studies prove that conventional machine learning classifiers lack sufficient hyperparameter optimization and scenario adaptability for practical teaching decision-making. The second stream explores online learning behavioral patterns through statistical analysis and human dynamics. By examining behavioral frequency, temporal features, and interactive patterns, these studies effectively characterize learning engagement and cognitive states (Sun & Li, 2022; Li et al., 2023; Touya & Fakir, 2014; Damez et al., 2005). While they successfully reveal the relationships between learning behaviors and outcomes, these approaches often rely on strict assumptions about data distribution and linearity, restricting their ability to capture the complex interdependencies inherent in real-world educational data. The third stream centers on multimodal fusion approaches, which have achieved promising performance in learning persistence detection, engagement identification, and teaching quality evaluation through temporal modeling, feature fusion, and attention mechanisms (Li et al., 2020; Huang et al., 2022; Li & Zhan, 2020; Yuan & Nie, 2021). Recent advancements in this area have further enriched the technical system by addressing specific data challenges. For instance, ensemble and deep learning models have been optimized to handle heterogeneous data, dynamic temporal dependencies, and class imbalances, significantly improving prediction accuracy and recall rates for underrepresented learner groups (Liu et al., 2024; Chen et al., 2022; Wang et al., 2023). Furthermore, studies have enhanced the objectivity of evaluation by combining multidimensional indicator weighting methods and integrating multimodal data (e.g. behavioral logs and physiological signals) (Zhang et al., 2021; Zhao et al., 2023). Despite these considerable advances in algorithm design and indicator optimization, existing research still exhibits notable gaps. Many recent methods overlook the importance of high-dimensional data preprocessing and systematic indicator validation through confirmatory factor analysis (CFA) (Li et al., 2024). Moreover, few studies have specifically optimized hyperparameters for practical teaching decision-making needs. Overall, the field lacks a unified framework that integrates factor analysis, indicator dimensionality reduction, and teaching decision-oriented model optimization. Therefore, constructing a reliable evaluation system and an interpretable prediction model for online learning behavior requires further investigation.

In order to bridge the gap, this paper, a detailed evaluative analysis of the online learning behaviour data of a subset of students obtained from the Ulearning platform is conducted, and performs exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) on the students' online learning data to verify the feasibility of the learning evaluation model indicators. Then, the hyperparameters are optimized by SVM through the demand analysis of teaching decision-making, and the online learning behaviour evaluation model is established.

2. The methods

Before constructing the online learning behaviour assessment model, it is necessary to analyse and select learning behaviour features from learners' online learning behaviour data because the analysis and selection of online learning behaviour features can significantly impact the final prediction results of learning outcomes. Therefore, it is necessary to classify and evaluate the indicators of learners' online learning behaviour data, and then establish an online learning behaviour evaluation model.

2.1. Dataset overview

The data used in this article are sourced from the Ulearning online teaching platform and comprise data of 800 student courses in the 2024 academic year. As the collected data often contain missing, redundant, or outlier values, they need to be cleaned and converted into standardized data formats before analysis. After cleaning all the collected data, a total of 792 real and valid learning behaviour data points were obtained. Finally, the online learning behaviour data that meet the criteria of completeness and uniqueness are obtained.

According to the collected data, the indicators that may affect students' performance were sorted and classified, shown in Table 1.

The first category of indicators refers to learning indicators, which are explicit features of students' online learning behaviour. These include metrics such as the total number of platform logins, total time spent online learning(seconds), number of resource views, number of resource downloads, number of online learning check-ins, duration of courseware viewing(seconds), average assignments score and tests average score. While the second category represents the learning effectiveness, namely, the average scores for midterm and final exams, which are also the categories for predicting grades.

To construct a more accurate and effective learning evaluation model, it is necessary to further refine and process these data and analyse the correlation between each online learning behaviour feature and the final output grades, as well as the interrelationships among the different behaviour features.

2.2. Design of online learning evaluation indicators

To assess whether the reliability level of the data meets the standard, this study uses Cronbach's Alpha coefficient as the standard for reliability testing of the data results. Options that do not meet the standard are considered interference options and are excluded to improve the reliability level of the data. As shown in Table 2, the overall Cronbach's Alpha coefficient for influencing factors is 0.778. The Cronbach's Alpha coefficients for the eight online learning behaviour indicators are 0.756, 0.76, 0.76,

Table 1. Data indicators of online learning behaviour.

Profile labels	Data indicators
Learning indicators	Total number of platform logins
	Total time spent online learning (seconds)
	Number of resource views
	Number of resource downloads
	Number of online learning check-ins
	Duration of courseware viewing(seconds)
	Assignments average score
	Tests average score
Learning effectiveness	Average scores for midterm and final exams

Table 2. The cronbach's alpha coefficient values of data indicators.

Data indicators	Cronbach's alpha
Total number of platform logins	0.756
Total time spent online learning	0.76
Number of resource views	0.76
Number of resource downloads	0.756
Number of online learning check-ins	0.74
Duration of courseware viewing	0.741
Assignments average score	0.771
Tests average score	0.744

Table 3. Results of KMO and bartlett sphericity test.

KMO		0.726
Bartlett test	χ^2	1972.583
	df	28
	Sig.	0

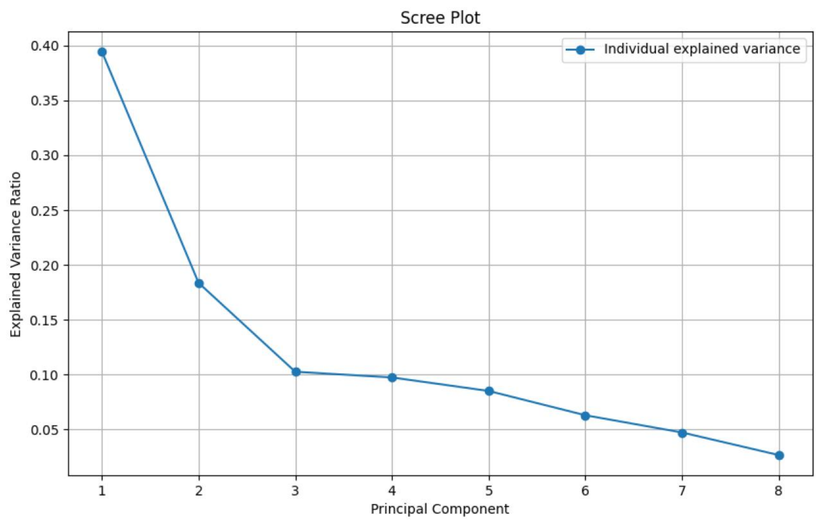

Figure 1. Scree plot of factor analysis.

0.756, 0.74, 0.741, 0.771, and 0.744, all above 0.7, indicating that the current data indicators have good reliability standards.

The construction of the online learning evaluation model is mainly divided into two parts: indicator design and weight determination. We conducted KMO and Bartlett tests to determine whether the selected evaluation indicators are suitable for EFA. The results are shown in Table 3. The KMO value is 0.726($>$0.7), indicating that the influencing factors meet the standards for factor analysis; The significance (*Sig.*) is 0($<$0.05), indicating that the significance standard is met; the approximate chi-square (χ^2) is 1972.583, and the degree of freedom (*df*) is 28. These results all meet the standards for factor analysis, indicating that this set of data is suitable for EFA.

Since the number of factors in CFA is preset by researchers and may lack objectivity and rationality, this study analyzed the dataset through EFA and obtained the visualization of the dimensional division for indicator design, as shown in the scree plot (Figure 1). The slope of the first three factors is higher, and the slope of the fourth factor becomes noticeably flatter. It can be inferred that the first three factors can mostly explain the total variance of the overall data. They were sufficient to extract the main structural information contained in the eight data indicators and could basically represent the learning level of learners.

Based on the scree plot analysis, factor analysis was used to classify the eight data indicators into three impact factors (Platform usage intensity, Learning behavior, Effective learning intensity). The factor analysis results are shown in Table 4. From the rotated component matrix, it can be seen that the factor loadings of the data indicators are all exceed 0.3, indicating that each type of data is essential for explaining the common factors.

Table 4. Factor analysis results for each dimension.

Influencing factors	Variable	Factor loading
Factor 1 (Platform usage intensity)	Total time spent online learning	0.694
	Number of resource views	0.748
Factor 2 (Learning behavior)	Total number of platform logins	0.901
	Number of resource downloads	0.906
Factor 3 (Effective learning intensity)	Number of online learning check-ins	0.705
	Duration of courseware viewing	0.614
	Assignments average score	0.733
	Tests average score	0.783

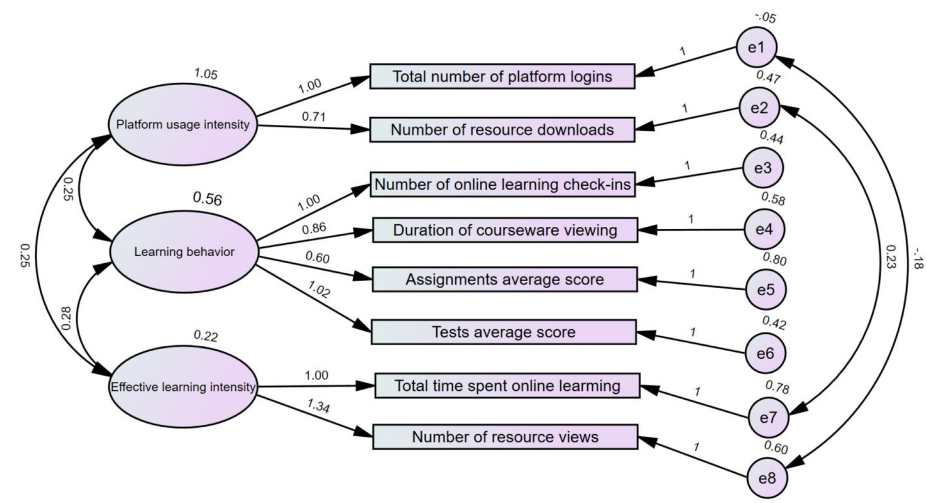

Figure 2. Confirmatory factor analysis model diagram.

Table 5. List of fitting indices.

χ^2/Df	*RMSEA*	*IFI*	*TLI*	*CFI*	
4.98	0.071	0.970	0.943	0.969	0.516

Based on the conceptual model of the influencing factors of online learning, the three dimensions mentioned above are used as factors. The structural model and path coefficients for the influencing factors of online learning are created using CFA and AMOS software, as shown in Figure 2.

Through the analysis of the data samples, we selected fit indices such as χ^2/Df, *RMSEA*, *IFI*, *TLI*, *CFI*, and *PNFI* as the basis for evaluating whether the model has scientific validity. The key fit indices in the model are shown in Table 5. The χ^2/Df is 4.98, indicating that the influencing factors have a certain degree of scientific validity. The $RMSEA = 0.071$, indicating that the model's fit level meets the reference standard. *IFI*, *TLI*, and *CFI* are greater than 0.9, and PNFI is greater than 0.5, indicating that the structural model's fit indices meet the standard. In summary, the current structural model and path coefficients for online learning have good structural validity.

2.3. SVM-based score evaluation and prediction model

After completing the indicators design, we need to determine the weights of the indicators, which will then allow us to evaluate and predict future performance. Due to the multidimensional nature of the feature data and the limited amount of data, nonlinear SVM can be used to obtain the optimal classification hyperplane that meets the classification criteria to predict learning outcomes (Cortes & Vapnik, 1995), thus enabling timely guidance for teachers to take intervention measures.

To take the three factors derived from factor analysis, namely Platform usage intensity, Learning behavior, Effective learning intensity as input features of the machine learning model, the abstract factors need to be converted into specific numerical values. The factor score F_i of each sample can be expressed as a linear combination of standardized variables Z_j. The calculation formula for the i_{th} factor is as follows:

$$F_i = \sum_{j=1}^{n} W_{ij} Z_j, i \in \{1, 2, 3\} \tag{1}$$

where F_i denotes the score of the i_{th} factor, Z_j represents the j_{th} standardized observed variable, and W_{ij} refers to the corresponding factor score coefficient. In this study, the factor loadings of all indicators listed in Table 4 are used as the factor score coefficients.

Thomson's Regression Method is adopted in this study to calculate factor scores. Based on the principle of generalized least squares, this method derives the optimal factor score coefficients by calculating the product of the inverse matrix of the correlation coefficient matrix of original variables and the factor loading matrix. Its mathematical expression is:

$$B = R^{-1} A \tag{2}$$

In the formula, R is the correlation coefficient matrix of observed variables, and A is the rotated factor loading matrix. The score coefficients calculated by this method can retain the information of original data to the greatest extent and eliminate the impact of multicollinearity among variables. Consequently, uncorrelated comprehensive factor scores are generated and taken as the input features of the subsequent SVM model (Table 5).

The establishment of SVM models for online learning features can be divided into three steps.

In the first step, the three retained features, namely, namely Platform usage intensity, Learning behavior, Effective learning intensity are standardized, and the training set and test set are divided. Construct an SVM model, which is an optimal hyperplane, so that the sum of squares of the distances from all data points to the hyperplane is minimized. This goal can be expressed by the following quadratic programming questions:

$$\text{minimize} = \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \xi_i, \xi_i \geq 0 \tag{3}$$

where C is the penalty coefficient, and ξ_i is the relaxation variable.

The constraints are:

$$\left(w^T x_i + b\right) y_i \geq 1 - \xi_i, i \in \{1, 2, \ldots, n\} \tag{4}$$

where y_i is the category label and x_i is the eigenvector.

Since there are several SVM hyperparameters and the model accuracy varies under different hyperparameters, the model hyperparameters were optimized in this study using a network search method.

In the second step, a grid search object was defined, and K-fold cross-validation was used to split the dataset into training and test sets. The entire dataset was divided into k folds, with $k-1$ folds used as the training set to train the model and the remaining 1-fold used as the validation set to score the model. This process was repeated k times. In this article, $k = 5$.

The third step involves training the model and determining the optimal parameters such as the optimal penalty coefficient, gamma coefficient and kernel function. Finally, the trained model and optimal parameters are used to predict the test set.

3. Experiments and results

We tested the model we designed in two ways: (1) the effectiveness of the model in predicting students' learning results based on online learning behaviour; (2) the construction of the model framework, that is, the contribution of each component. We used the coefficient of determination (R^2), the mean absolute error (MAE), and the root mean square error ($RMSE$) as evaluation indicators to evaluate the model's prediction results. Where R^2 evaluates the model's fitting effect by comparing the difference between the predicted value of the model and the actual value. The formula is

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{5}$$

Table 6. Evaluation metrics.

	Training set	Testing set
R^2	0.9790232076496892	0.9599262386297075
MAE	0.10875864732620538	0.12685143914901906
RMSE	0.14676758205642212	0.19366742150846983

where n is the sample size, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average of the actual values.

The MAE reflects the average difference between the predicted and actual values. The formula is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

The RMSE measures the degree of deviation between the predicted value and the actual value. The formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

The evaluation results based on SVM training and testing are shown in Table 6.

To further interpret the analytical results and clarify the correlation between online learning behaviors and academic achievement, this study conducts an in-depth discussion on core evaluation indicators and latent factors, grounded in classical learning analytics theories and prior empirical evidence. From the perspectives of self-regulated learning theory and student engagement theory, learners' online behavioral patterns can be categorized into three dimensions, namely behavioral participation, learning initiative and process-based learning performance (Sun & Li, 2022; Li et al., 2023). The three latent factors extracted via exploratory factor analysis (EFA) are highly consistent with this theoretical framework, which effectively encapsulates the essential characteristics of students' online learning status.

The high coefficient of determination ($R^2 = 0.960$) on the test set demonstrates that the composite indicators constructed from latent factors can satisfactorily explain the variance in students' academic performance. This finding corroborates the conclusions of existing relevant studies (Yu et al., 2020; Zhang et al., 2017), which verified that platform usage duration, interaction with learning resources, curriculum participation and daily academic performance are significantly correlated with final learning outcomes.

In terms of prediction errors, the optimized SVM model yields low mean absolute error (MAE) and root mean square error (RMSE) on both training and test datasets, with a marginal discrepancy between the two sets. This outcome verifies that the hyperparameter optimization strategy combined with K-fold cross-validation can effectively mitigate overfitting and enhance the generalization performance of the model on real educational datasets. Relevant conclusions have also been reported in prior studies (Jin & Zhuo, 2020; Wang, 2022).

Compared with previous research, most conventional prediction models for online learning directly adopt original behavioral variables for modeling, while neglecting reliability assessment and dimensionality reduction of evaluation indicators (Luo et al., 2024; Li et al., 2022). Additionally, traditional SVM models generally employ fixed hyperparameters, resulting in limited adaptability when dealing with small-sample and nonlinear educational data. Differing from the above approaches, this study firstly adopts EFA and confirmatory factor analysis (CFA) to verify the construct validity of evaluation indicators, and subsequently establishes an SVM model with optimized hyperparameters. The proposed methodological framework ensures the rationality of variable selection, improves the interpretability and practical value of the model in educational decision-making, and remedies the deficiencies of existing research.

4. Conclusion

With the advancement of educational network informatization, online learning has become a more popular learning mode than traditional methods. Conducting performance prediction research using online learning behaviour data is currently a crucial area of focus. In this study, student online learning behaviour data are used to conduct a factor analysis to verify the feasibility of the learning assessment model indicators.

Furthermore, a score evaluation and prediction model is established using the SVM algorithm. The results of this model can provide timely feedback on students' learning status, guide teachers in their teaching activities, and improve the efficiency of student learning. However, due to the relatively limited data sources and insufficient data volume, it is challenging to conduct a more precise analysis and research on online learning behaviours. Therefore, future research could incorporate multiple data sources and samples from different academic years. Additionally, temporal dynamics and multimodal data will be further integrated to enrich the analytical framework and empirical investigation of online learning behaviours.

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

Author contributions

CRediT: **Zhaohui Zheng**: Conceptualization, Funding acquisition, Methodology, Writing – review & editing; **Kemin Hu**: Project administration, Software, Writing – original draft; **Zhongyuan Guo**: Data curation, Formal analysis; **Deng Cheng**: Validation, Visualization.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was funded by Humanities and Social Science Foundation of the Ministry of Education in China, 25YJCZH394 and Teaching Research Project of Wuhan Institute of Technology, X2022042.

About the authors

Zhaohui Zheng received the Ph.D degree in the School of Electronic Information, Wuhan University, Wuhan, China. He is now a lecturer with the School of Mathematics and Physics, Wuhan Institute of Technology.

Kemin Hu is currently pursuing the B.S. degree with the School of Mathematics and Physics, Wuhan Institute of Technology.

Zhongyuan Guo is currently a Lecturer with the School of Computer Science and Engineering, Chongqing University of Science and Technology.

Deng Cheng is currently a Professor with the School of Computer Science and Engineering, Wuhan Institute of Technology.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors. Data is not available due to student privacy.

References

Allen, I. E., & Seaman, J. (2017). Digital Learning Compass: Distance Education Enrollment Report 2017. *Babson Survey Group Report, 2017*.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., & Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of empirical literature. *Review of Educational Research*, *74*(3), 379–439. https://doi.org/10.3102/00346543074003379

Bosch, N. (2021). Identifying supportive student factors for mindset interventions: A two-model machine learning approach. *Computers & Education*, *167*, 104190. https://doi.org/10.1016/j.compedu.2021.104190

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies and performance in learning online. *Computers in Human Behavior*, *50*, 41–47.

Chen, H., Li, M., & Zhao, J. (2022). Prediction of online learning persistence based on attention-LSTM network. *Journal of Educational Technology & Society*, *25*, 189–203.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Damez, M., Bouchon-Meunier, B., Ha, T., & Marsala, C. (2005). Fuzzy decision tree for user modeling from human-computer interactions. *Proc. of the 5th International Conference on Human System Learning* (pp. 287–302).

Fu, E. Y., Ngai, G., Leong, H. V., Chan, S. C. F., & Shek, D. T. L. (2023). Using attention-based neural networks for predicting student learning outcomes in service-learning. *Education and Information Technologies*, *28*(10), 13763–13789. https://doi.org/10.1007/s10639-023-11592-0

Huang, H., Han, G., Xiao, F., & Wang, R. (2022). An online teaching video evaluation scheme based on EEG signals and machine learning. *Wireless Communications and Mobile Computing*, *2022*, 1399202. https://doi.org/10.1155/2022/1399202

Jin, X., & Zhuo, Y. (2020). Prediction of college entrance examination scores based on genetic algorithm and support vector machine model. *Journal of Henan Institute of Engineering (Natural Science Edition)*, *32*, 62–65.

Jordan, K. (2019). MOOC completion rates: Analysis and trends. *Journal of Asynchronous Learning Networks*, *23*, 5–21.

Li, M., Ge, M., Zhao, H., & An, Z. (2022). Modeling and analysis of learners' emotions and behaviors based on online forum texts. *Computational Intelligence and Neuroscience*, *2022*, 9696422. https://doi.org/10.1155/2022/9696422

Li, Q., Wang, P., & Zhang, R. (2024). Preprocessing method for high-dimensional online learning data based on PCA-ICA hybrid algorithm. *Journal of Data Analysis and Information Processing*, *12*, 34–45.

Li, Y., Jiang, Q., Fang, H., & Zhao, W. (2023). Analysis and modeling of online learning behavior patterns based on human dynamics. *E-Education Research*, *44*, 77–83.

Li, Z., & Zhan, Z. (2020). Integrated infrared imaging techniques and multi-model information via convolution neural network for learning engagement evaluation. *Infrared Physics & Technology*, *109*, 103430. https://doi.org/10.1016/j.infrared.2020.103430

Li, Z., You, F., & Sun, D. (2020). Research on the evaluation of learning behavior on MOOCs based on cluster analysis. In *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering* (pp. 1055–1059).

Liu, J., Zhang, Y., & Wang, L. (2024). Online learning behavior analysis based on MLP-Bagging ensemble classification model. *Journal of Yunnan University*, *46*, 987–996.

Luo, Y., Han, X., & Zhang, C. (2024). Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses. *Asia Pacific Education Review*, *25*(2), 267–285. https://doi.org/10.1007/s12564-022-09749-6

Stojanov, A., & Daniel, B. K. (2024). A decade of research into the application of big data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*, *29*(5), 5807–5831. https://doi.org/10.1007/s10639-023-12033-8

Sun, W., & Li, M. (2022). Research on the influence of college students' online learning behavior investment on learning effect. *Journal of Bingtuan Education Institute*, *32*, 54–60.

Sun, Y., Ye, L., Zheng, S., & Li, T. (2021). Educational big data: Predictions, applications and challenges. *Big Data Research*, *25*, 100196.

Touya, K., & Fakir, M. (2014). Mining students' learning behavior in moodle system. *Journal of Information Technology Research*, *7*(4), 12–26. https://doi.org/10.4018/jitr.2014100102

Zhao, G. Y., & Tu, C. Z. (2025). Comparative Analysis of Applying SMOTE Techniques to Imbalanced Data for Online Instructors' Performance Prediction. *Proceedings of the Twentieth Midwest Association for Information Systems (MWAIS 2025), 2025*.

Wang, T. (2022). Analysis and research on online learning behavior path based on K-means clustering algorithm under the background of big data. *Network Security Technology and Application*, *5*, 43–45.

Yu, T., Gan, C., Li, X., & Yang, C. (2020). Application research of SVR regression in grade prediction warning. *Chinese Journal of ICT in Education*, *11*, 76–80.

Yuan, F., & Nie, Y. (2021). Online classroom teaching quality evaluation system based on facial feature recognition. *Scientific Programming*, *2021*, 1–10. https://doi.org/10.1155/2021/7374846

Zhang, L., Lu, X., Lu, C., Wang, B., & Li, F. (2017). Application of support vector machine in analysis of college entrance examination score prediction. *Journal of University of Science and Technology of China*, *47*, 1–9.

Zhang, Q., Liu, H., & Yang, Z. (2021). Weight assignment of online learning behavior indicators based on AHP-Entropy method. *Journal of Computer Applications in Education*, *27*, 76–82.

Zhao, L., Sun, X., & Wu, Y. (2023). Multimodal data fusion for online learning outcome prediction: A deep learning approach. *IEEE Transactions on Learning Technologies*, *16*, 245–258.