REVIEW ARTICLE

A Review of Recent Advances in Academic Performance Prediction within Educational Data Mining

Xiaowen Zhu[1,*]

[1] College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China

Abstract

With the rapid advancement of artificial intelligence and smart education, academic performance prediction has emerged as a critical research direction within educational data mining. This paper provides a systematic review of recent progress in this field, focusing on technological evolution, application scenarios, model adaptation, emerging trends, and persistent challenges. The development of prediction technologies has progressed through three stages: statistical methods, machine learning, and deep learning. These approaches have been widely applied to grade prediction, academic risk and dropout prediction, knowledge tracing, and personalized learning recommendation. Different models vary significantly in their data scale requirements, interpretability, and suitability for specific tasks. Current research is shifting toward interpretability, multimodal data fusion, privacy preservation, and lightweight deployment. Despite these advances, challenges remain, including data imbalance, cold-start problems, and limited model interpretability. To better support precision teaching and personalized education, future research should prioritize interpretability, actionability, and deployability.

Keywords: educational data mining, academic performance prediction, machine learning, deep learning, model adaptation.

1 Introduction

In the context of digital transformation in education, large amounts of data on student learning behaviors, grades, personal backgrounds, and online interactions are being collected and stored [1]. Leveraging these data to accurately predict student academic performance has become a key strategy for improving teaching quality and enabling targeted interventions [2]. Reducing dropout rates is among the most critical goals of such efforts, as research has demonstrated that student engagement behaviors—such as social interaction in online courses—are closely associated with attrition [3]. Academic performance prediction typically takes multi-dimensional student characteristics as input and employs intelligent algorithms to forecast outcomes such as grades, risk levels, and dropout probability, thereby providing decision support for teachers, educational institutions, and online learning platforms. In recent years, rapid advancements in related technologies have driven continued in-depth research. This paper reviews the latest progress in this field, aiming to serve as a reference for future research

Submitted: 27 July 2026
Revised: 19 August 2026
Accepted: 15 September 2026
Published: 28 September 2026

Vol. 2, **No.** 2, 2026.
10.62762/TEDM.2026.918140

***Corresponding author:**
Xiaowen Zhu
2425191037@stu.hqu.edu.cn

Citation
Zhu, X. (2026). A Review of Recent Advances in Academic Performance Prediction within Educational Data Mining. *ICCK Transactions on Educational Data Mining*, 2(2), 56–62.

and practice [4, 5].

Table 1 outlines the standard six-step workflow for academic performance prediction in educational data mining. The process begins with data collection, which gathers multi-source student information such as grades, learning behaviors, personal backgrounds, and online interactions. This is followed by data preprocessing, including missing value handling, outlier removal, and data standardization [4]. The third step, feature engineering, involves feature selection, construction, and dimensionality reduction (e.g., PCA) [6]. Next, model training applies machine learning, deep learning, or ensemble methods with hyperparameter tuning. The model is then evaluated using metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. Finally, the prediction application stage supports tasks such as grade warning, dropout risk identification, and personalized learning recommendation [7]. This workflow provides a structured framework for researchers and practitioners to implement and compare prediction systems.

Table 1. General workflow of academic performance prediction.

Step	Name	Specific Content
1	Data Collection	Student grades, learning behavior logs, personal background information, online interaction data
2	Data Preprocessing	Missing value processing, outlier removal, data standardization, category encoding
3	Feature Engineering	Feature selection, feature construction, dimensionality reduction (PCA; t-SNE for visualization)
4	Model Training	Machine learning / deep learning / ensemble learning model training and tuning
5	Model Evaluation	Accuracy, precision, recall, F1 score, AUC-ROC
6	Prediction Application	Grade warning, dropout risk identification, personalized learning recommendation

2 Review Methodology

To ensure the transparency and reproducibility of this review, we adopted a systematic literature search strategy. We searched four major academic databases: Web of Science, Scopus, IEEE Xplore, and ACM Digital Library. The search keywords included combinations of the following terms: "educational data mining", "academic performance prediction", "student dropout prediction", "student grade prediction", and "knowledge tracing". We limited our search to peer-reviewed journal articles and conference papers published between January 2015 and December 2025 (with six additional 2026 studies included at the manuscript preparation stage, as noted below). Studies were included if they: (1) focused on predicting student academic outcomes using statistical, machine learning, or deep learning methods; (2) were empirical studies with quantitative evaluation; and (3) were written in English. We excluded studies that were not in the educational domain or that did not involve predictive modeling. After title and abstract screening and full-text review, 21 empirical studies met all inclusion criteria and were analyzed in this review.

To provide the necessary context, 20 additional works were cited outside the systematic search. These comprise four foundational works published no later than 2015 that defined key concepts such as knowledge tracing and learning analytics; ten surveys, handbook chapters, methodological papers (e.g., the original SHAP and LIME publications), and dataset descriptors; and six closely related studies published in 2026, which were identified during manuscript preparation. In total, 41 works are cited in this review.

3 Evolution of Core Technologies

3.1 Statistical Methods

Statistical methods are the foundational techniques for academic performance prediction [4]. They offer strong interpretability, low computational cost, and suitability for small samples, and are still widely used for baseline experiments and administrative decisions [2]. Main methods include linear regression, logistic regression, the Cox proportional hazards model [8], and structural equation modeling. In recent years, statistical methods have been combined with causal inference, shifting from correlation-based prediction to causal analysis. This enables the identification of the true effects of factors such as study time and tutoring frequency on academic performance, thereby enhancing the scientific rigor and operability of teaching interventions [9]. Although machine learning and deep learning are developing rapidly, statistical methods remain an important benchmark for academic performance prediction and are especially valuable in educational decisions involving small samples and high interpretability requirements [2, 10].

3.2 Machine Learning Methods

Machine learning is currently the mainstream technology for academic performance prediction, significantly improving prediction accuracy through feature engineering and ensemble learning. Core models include ensemble learning models such as Random Forest, XGBoost, and LightGBM, as well as traditional classifiers such as SVM, KNN, and Naive Bayes [6]. Recent research is mainly reflected in three aspects. First, explainable AI techniques such as SHAP [11] and LIME [12] have been introduced, enabling models to output justifications for their predictions [13–15]. Second, multimodal feature fusion has been realized, integrating learning behaviors, family background, psychological states, and other information [15]. Third, methods such as SMOTE and Focal Loss have been used to address class imbalance, improving the identification of minority samples such as dropout and course failure [16]. Ensemble models, owing to their high stability and good prediction accuracy, have become the most commonly used academic performance prediction solutions in primary and secondary schools, university academic affairs systems, and online education platforms [5].

It is important to distinguish between two types of interpretability approaches. Inherently interpretable models, such as logistic regression and decision trees, provide direct transparency because their decision rules are human-readable [10]; however, they may sacrifice some predictive performance. In contrast, post-hoc explanation methods such as SHAP and LIME can be applied to complex black-box models to explain individual predictions, but sampling-based explainers (e.g., LIME and KernelSHAP) may produce explanations that vary across runs [11, 12]. In educational practice, this distinction has real implications: teachers may prefer inherently interpretable models for immediate classroom use, while administrators may adopt post-hoc methods to audit complex systems [10]. For example, SHAP values can reveal that low submission rates of weekly assignments are the strongest predictor of poor academic performance, enabling instructors to send timely alerts to at-risk students [14].

3.3 Deep Learning Methods

Deep learning represents the cutting-edge direction in academic performance prediction in recent years, automatically extracting deep features from complex data such as time series, graphs, and text [17]. Time series modeling mainly uses LSTM, GRU, and Transformer, and is suitable for data such as learning behavior sequences and homework submission records [17–19]. Graph structure modeling mainly uses GCN, GAT, and hypergraph neural networks, and is suitable for scenarios such as student-course relationships and knowledge point associations [20–22]. Multimodal fusion combines CNN, large language models, and graph neural networks to jointly model text, behavior, grade, and other information [20, 23]. Recent advances include the integration of large models with educational data, hybrid combination of traditional machine learning and deep learning, and lightweight model deployment, significantly improving prediction performance in complex scenarios [5, 17, 24].

4 Main Practical Applications

4.1 Student Grade Prediction

Grade prediction is the most classic task in academic performance prediction, aiming to predict final exam scores or the probability of course failure [2, 6, 18]. The technology has gradually shifted from single Random Forest models to a combination of ensemble learning and explainable AI [6]. XGBoost combined with SHAP is now widely used to output both prediction results and key influencing factors, although the reported performance varies considerably across datasets and tasks [13–15]. Grade prediction can help teachers identify at-risk students in advance and enable early intervention [25]; subsequent research has further demonstrated how structured pedagogical support scenarios, designed around academic risk profiles, can be implemented to promote subject learning success [26].

4.2 Academic Risk and Dropout Prediction

Dropout prediction is an important requirement in educational management and is a typical imbalanced classification task [27]. Research has gradually shifted from traditional logistic regression to deep learning and imbalanced data optimization methods. Previous studies on the Open University Learning Analytics Dataset (OULAD) [28] have shown that deep learning models outperform logistic regression and SVM baselines, with reported accuracies of 84%-93% [29], whereas early-stage multi-class prediction remains considerably harder [30]. Combined with federated learning, cross-institutional collaborative modeling can be achieved while protecting privacy [31–34]. Dropout early warning can effectively reduce attrition rates in universities and online courses, improving the utilization rate of educational resources [3].

4.3 Knowledge Tracing

Knowledge tracing is used to dynamically track students' mastery of knowledge points and is a core technology for personalized teaching [35, 36]. The technology has evolved from Bayesian knowledge tracing [35] to deep knowledge tracing [36] and graph-based knowledge tracing [22]. In recent years, it has been further enriched with exercise text semantics [37] and combined with large language models to enhance semantic understanding and knowledge point association capabilities [24], and it is widely used in intelligent tutoring systems and adaptive learning systems. Knowledge tracing can reflect students' mastery levels in real time, providing a basis for adaptive learning.

4.4 Personalized Learning and Path Recommendation

The combination of academic performance prediction and recommendation systems can customize learning plans and push learning resources based on student performance [38]. The research trend is to incorporate causal reasoning to achieve explainable recommendations, making the recommendation results more trustworthy and more easily accepted by teachers and students [9]. Personalized learning recommendations can dynamically adjust learning paths based on prediction results, improving learning efficiency [38].

5 Model-Scenario Adaptability Analysis

Statistical methods are suitable for small-scale data, strong interpretability requirements, and baseline experiments. Among machine learning methods, Random Forest is suitable for course experiments and small-to-medium-scale data, while XGBoost is suitable for high-precision early warning systems. Deep learning is suitable for sequential behaviors, large-scale interaction data, and complex relationship modeling. As summarized in Table 2, logistic regression and Random Forest are preferred for small samples; boosting ensembles such as XGBoost are preferred for medium-to-large samples (more than about 5000 records); and models such as LSTM, Transformer, and GNN can be used for large samples (more than about 10000 records). When selecting a model, data scale, interpretability requirements, computational resources, and deployment conditions should be comprehensively considered to achieve the best prediction results [2, 5].

6 Research Timeline

Figure 1 presents a chronological timeline of key milestones in academic performance prediction research within educational data mining. The timeline begins in 2007 with foundational reviews of educational data mining (EDM) [39], followed by the establishment of learning analytics (LA) as a discipline in 2013 [40]. A major breakthrough occurred in 2015 with the introduction of deep knowledge tracing (DKT) [36], which leveraged recurrent neural networks to dynamically model student knowledge states. By 2019, explainable AI techniques such as SHAP [11] and LIME [12] had gained prominence,

Table 2. Comparison of different models for academic performance prediction.

Model Category	Representative Models	Data Requirement	Interpretability	Predictive Accuracy	Typical Application Scenarios
Statistical Methods	Logistic Regression, Cox	Small (<500)	★★★★★	★★☆☆☆	Baseline, causal inference
Traditional ML	SVM, KNN, Naive Bayes	Small-Medium (500-5000)	★★★★☆	★★★☆☆	Course experiments
Ensemble Learning	RF, XGBoost, LightGBM	Medium-Large (>5000)[a]	★★★☆☆	★★★★☆	Early warning systems
Deep Learning	LSTM, GRU, Transformer	Large (>10000)	★★☆☆☆	★★★★★	Sequence modeling
Graph Neural Networks	GCN, GAT	Large graph	★★☆☆☆	★★★★☆	Knowledge tracing

Ratings are qualitative (★ = higher). Sample-size thresholds are indicative only. [a]Random Forest also performs robustly on small-to-medium data sets, whereas boosting models (XGBoost, LightGBM) benefit more from larger samples.

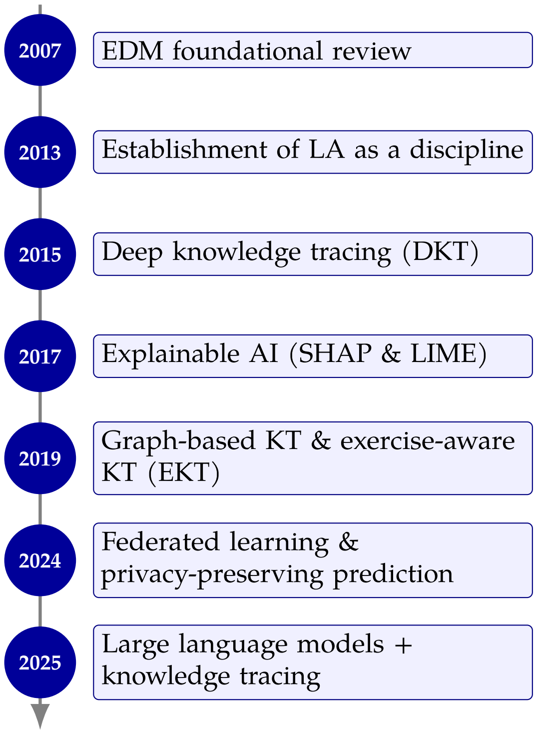

Figure 1. Timeline of key milestones in academic performance prediction research.

and the debate on interpretability in high-stakes decisions [10] addressed the black-box nature of complex models. Around 2019–2021, graph-based knowledge tracing [22] emerged and exercise-aware knowledge tracing (EKT) [37] was subsequently developed to model student-course and knowledge point relationships. In 2024, federated learning was applied to privacy-preserving, cross-institutional prediction [31, 32, 34], and most recently large language models have been integrated into student modeling [24]. This timeline illustrates the field's rapid evolution from statistical and rule-based methods toward deep learning, interpretability, and privacy-aware architectures.

7 Conclusion

With the fast development of artificial intelligence, academic performance prediction has evolved from statistical methods to machine learning and now to deep learning, and has become a vital research area within educational data mining [2, 4, 5]. This review has highlighted several key trends that define the current and future direction of the field. First, model interpretability has become essential, with tools such as SHAP and LIME being widely adopted to make predictions more transparent and actionable for educators [14, 15, 17]. Second, the integration of large models and multimodal data fusion is emerging as a cutting-edge approach to capture the complexity of student learning [17, 24]. Third, privacy-preserving technologies are advancing rapidly [32, 33]. Fourth, the incorporation of causal reasoning and intervention optimization is enhancing the practical value of predictions, moving beyond correlation to actionable insight. Finally, the development of lightweight and deployable models represents a promising direction for enabling real-time early warning systems in classroom settings.

In the future, academic performance prediction will continue to advance toward lightweight architectures, edge computing, multimodal large models, and causal interpretability. Greater emphasis will be placed on practical application and closed-loop teaching, ensuring that predictive technologies genuinely serve classroom instruction and administrative decision-making. Beyond the persistent challenges of data imbalance, cold-start problems, and insufficient interpretability, several emerging issues require urgent attention:

- **Fairness and Bias.** Predictive models may inadvertently introduce biases against certain demographic groups due to imbalanced training data, and their performance can vary across student subgroups and cohorts [41]. Future research should incorporate fairness-aware algorithms and conduct bias audits before deploying models in real educational settings.

- **Privacy-Preserving Learning.** As educational data often contain sensitive student information, cross-institutional data sharing faces legal and ethical barriers. Federated learning and differential privacy are promising techniques that enable model training without exposing raw student records [33, 34], and these warrant further exploration.

- **Cross-Institutional Generalization.** Most existing models are trained on a single institutional dataset, leading to poor transferability. Future work should test models across multiple institutions with diverse student populations and investigate domain adaptation methods.

- **From Prediction to Intervention.** A critical gap remains between achieving high prediction accuracy and implementing effective educational interventions. Future research should design actionable intervention strategies and rigorously evaluate their impact on student retention outcomes.

Data Availability Statement

Not applicable.

Funding

Data will be made available on request.

Conflicts of Interest

The author declares no conflicts of interest.

AI Use Statement

The author declares that DeepSeek was used for language editing, proofreading, and translation of the manuscript. The author has carefully reviewed, revised, and verified the AI-assisted output and takes full responsibility for the content of the manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Baker, R. S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. *Emergence and innovation in digital learning: Foundations and applications* (pp. 83-98). Athabasca University Press. [CrossRef]

[2] Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. *Technology, Knowledge and Learning, 24*(4), 567-598. [CrossRef]

[3] Wang, W., Guo, L., He, L., & Wu, Y. J. (2019). Effects of social-interactive engagement on the dropout ratio in online learning: Insights from MOOC. *Behaviour & Information Technology, 38*(6), 621-636. [CrossRef]

[4] Sarker, S., Paul, M. K., Thasin, S. T. H., & Hasan, M. A. M. (2024). Analyzing students' academic performance using educational data mining. *Computers and Education: Artificial Intelligence, 7*, 100263. [CrossRef]

[5] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355. [CrossRef]

[6] Galiatsatos, D., & Galiatsatou, P. (2026). Leveraging feature selection and ensemble learning to predict secondary school achievement: A comparative study of three grade granularities. *Information, 17*(6), 517. [CrossRef]

[7] Baker, R. S., & Siemens, G. (2022). Learning analytics and educational data mining. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed., pp. 259-278). Cambridge University Press. [CrossRef]

[8] Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016, October). Survival analysis based framework for early prediction of student dropouts. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 903-912). [CrossRef]

[9] Weidlich, J., Gašević, D., & Drachsler, H. (2022). Causal inference and bias in learning analytics: A primer on pitfalls using directed acyclic graphs. *Journal of Learning Analytics, 9*(3), 183-199. [CrossRef]

[10] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206-215. [CrossRef]

[11] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems, 30*.

[12] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). [CrossRef]

[13] Ren, Y., & Yu, X. (2024). Long-term student performance prediction using learning ability self-adaptive algorithm. *Complex & Intelligent Systems, 10*(5), 6379-6408. [CrossRef]

[14] Esin, R. V., & Kustitskaya, T. A. (2026). Two-level monitoring system for preventing academic failure, based on predictive models and SHAP analysis. *Education Sciences, 16*(6), 842. [CrossRef]

[15] Lau, W. J., & Abdul Rahman, H. (2026). Predicting academic performance through machine learning: integrating demographic, psychological, and behavioral predictors using explainable AI. *Asia Pacific Education Review*, 1-18. [CrossRef]

[16] Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications, 41*(2), 321-330. [CrossRef]

[17] Yang, M., Li, Z., & Liu, S. (2026). A machine learning based framework for predictive school management using student and faculty analytics. *Scientific Reports, 16*, 16224. [CrossRef]

[18] Lakshmi, S., & Maheswaran, C. P. (2024). Effective deep learning based grade prediction system using gated recurrent unit (GRU) with feature optimization using analysis of variance (ANOVA). *Automatika, 65*(2), 425-440. [CrossRef]

[19] Li, X., Zhu, X., Zhu, X., Ji, Y., & Tang, X. (2020, May). Student academic performance prediction using deep multi-source behavior sequential network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 567-579). Cham: Springer International Publishing. [CrossRef]

[20] Zhang, X., Zhang, Y., Chen, A. L., Yu, M., & Zhang, L. (2025). Optimizing multi label student performance prediction with GNN-TINet: A contextual multidimensional deep learning framework. *PloS one, 20*(1), e0314823. [CrossRef]

[21] Li, M., Wang, X., Wang, Y., Chen, Y., & Chen, Y. (2022). Study-GNN: a novel pipeline for student performance prediction based on multi-topology graph neural networks. *Sustainability, 14*(13), 7965. [CrossRef]

[22] Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019, October). Graph-based knowledge tracing: modeling student proficiency using graph neural network. In *IEEE/WIC/aCM international conference on web intelligence* (pp. 156-163). [CrossRef]

[23] El Aouifi, H., El Hajji, M., Es-Saady, Y., & Douzi, H. (2021). Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining. *Education and Information Technologies, 26*(5), 5799-5814. [CrossRef]

[24] Yang, L., Sun, X., Li, H., Xu, R., & Wei, X. (2025). Difficulty aware programming knowledge tracing via large language models. *Scientific Reports, 15*(1), 11475. [CrossRef]

[25] Bouallegue, S., Omri, A., & Al-Naemi, S. (2026). Machine learning approaches for early student performance prediction in programming education. *Information, 17*(1), 60. [CrossRef]

[26] Somova, M. V., Vainshtein, Y. V., Noskov, M. V., & Vonog, V. V. (2026). Scenarios of Pedagogical Support to Promote Subject Learning Success Among University Students Based on Academic Risk Analysis. *Education Sciences, 16*(6), 882. [CrossRef]

[27] Pan, F., Zhang, H., Li, X., Zhang, M., & Ji, Y. (2024). Achieving optimal trade-off for student dropout prediction with multi-objective reinforcement learning. *PeerJ Computer Science, 10*, e2034. [CrossRef]

[28] Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. *Scientific data, 4*(1), 170171. [CrossRef]

[29] Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human behavior, 104*, 106189. [CrossRef]

[30] Al-azazi, F. A., & Ghurab, M. (2023). ANN-LSTM: A deep learning model for early student performance prediction in MOOC. *Heliyon, 9*(4), e15382. [CrossRef]

[31] Farooq, U., Naseem, S., Mahmood, T., Li, J., Rehman, A., Saba, T., & Mustafa, L. (2024). Transforming educational insights: strategic integration of federated learning for enhanced prediction of student learning outcomes. *The Journal of Supercomputing, 80*(11), 16334-16367. [CrossRef]

[32] Zhang, Y., Li, Y., Wang, Y., Wei, S., Xu, Y., & Shang, X. (2024). Federated learning-outcome prediction with multi-layer privacy protection. *Frontiers of Computer Science, 18*(6), 186604. [CrossRef]

[33] Chen, S., & Qi, X. (2025). Entropy-adaptive differential privacy federated learning for student performance prediction and privacy protection: A case study in Python programming. *Frontiers in Artificial Intelligence, 8*, 1653437. [CrossRef]

[34] Fachola, C., Tornaría, A., Bermolen, P., Capdehourat, G., Etcheverry, L., & Fariello, M. I. (2023). Federated learning for data analytics in education. *Data, 8*(2), 43. [CrossRef]

[35] Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction, 4*(4), 253-278. [CrossRef]

[36] Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems 28* (pp. 505-513).

[37] Liu, Q., Huang, Z., Yin, Y., Chen, E., Xiong, H., Su, Y., & Hu, G. (2021). EKT: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering*, 33(1), 100-115. [CrossRef]

[38] Nabizadeh, A. H., Leal, J. P., Rafsanjani, H. N., & Shah, R. R. (2020). Learning path personalization and recommendation methods: A survey of the state-of-the-art. *Expert Systems with Applications, 159*, 113596. [CrossRef]

[39] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135-146. [CrossRef]

[40] Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American behavioral scientist, 57*(10), 1380-1400. [CrossRef]

[41] Glandorf, D., Lee, H. R., Orona, G. A., Pumptow, M., Yu, R., & Fischer, C. (2024, March). Temporal and between-group variability in college dropout prediction. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 486-497). [CrossRef]

Xiaowen Zhu is currently an undergraduate student majoring in Artificial Intelligence at the College of Computer Science and Technology, Huaqiao University, Xiamen, China. Her research interests include educational data mining, machine learning, and academic performance prediction. (Email: 2425191037@stu.hqu.edu.cn)