PERSPECTIVE

Advancing Educational Data Mining through Multi-Source Data Fusion and Explainable Knowledge

Xin Gu[1,2], **Fariza Sabrina**[3], **Shaleeza Sohail**[4] and **Zongwen Fan**[5,*]

[1] King's Own Institute, Sydney 2000, Australia
[2] Kent Institute of Higher Education, Sydney 2000, Australia
[3] School of Engineering and Technology, Central Queensland University, Rockhampton 4701, QLD, Australia
[4] College of Engineering, Science and Environment, The University of Newcastle, Callaghan 2308, NSW, Australia
[5] College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China

Abstract

The rapid digitalization of education has led to an explosive growth of multi-source and multi-modal learning data, providing new opportunities for advancing Educational Data Mining (EDM). By learning diverse data such as learning behaviors, assessment records, and interaction logs, EDM enables deeper insights into student learning processes and supports the development of personalized and intelligent education. However, several critical challenges remain, including the heterogeneity and fragmentation of multi-source data, the difficulty of extracting meaningful knowledge through effective data fusion, and the limited interpretability of high-performance predictive models. To address these challenges, this paper emphasizes the importance of integrating multi-source data fusion with explainable knowledge modeling. Multi-modal data fusion provides the basis for constructing comprehensive and coherent learner representations. The explainable approaches are essential for ensuring transparency, trust, and practical usability in educational settings. In particular, fuzzy knowledge representation, characterized by interpretable rules and membership functions, is important for bridging data-driven analytics and pedagogical reasoning. Fuzzy systems are the ideal candidates for explainable knowledge modeling as they combine human-like linguistic reasoning with mathematically precise computation by allowing symbolic rules and numerical data in a transparent and interpretable way. By analyzing the key challenges and emerging directions in multi-source educational data mining, this editorial highlights the necessity of combining data integration, personalized learning, and interpretability. Such an approach enables more precise learning path design, targeted intervention, and trustworthy decision support, ultimately advancing the development of intelligent, explainable, and learner-centered educational systems.

Keywords: educational data mining, multi-source data fusion, explainable knowledge, multi-modal learning data, personalized learning.

Submitted: 22 February 2026
Accepted: 23 March 2026
Published: 25 March 2026

Vol. 2, **No.** 1, 2026.
10.62762/TEDM.2026.793467

***Corresponding author:**
Zongwen Fan
zongwen.fan@hqu.edu.cn

Citation
Gu, X., Sabrina, F., Sohail, S., & Fan, Z. (2026). Advancing Educational Data Mining through Multi-Source Data Fusion and Explainable Knowledge. *ICCK Transactions on Educational Data Mining*, 2(1), 29–32.

1 Introduction

The rapid advancement of information technology is driving a profound digital transformation in educational data mining [1]. Traditional pedagogical models are confronted with several challenges, including limited instructional diversity and inability to accurately capture individual learner differences. These limitations hinder their ability to accommodate increasingly heterogeneous student populations and align with the complex demands of modern education. To address these issues, the emergence of Educational Data Mining (EDM) provides new opportunities [2]. By analyzing large-scale educational data, EDM enables in-depth analysis of students' learning processes, behavioral patterns, and performance outcomes, thereby supporting the development of personalized learning and intelligent education systems.

In modern education, the volume of data has grown explosively [3], driven largely by the widespread adoption of online learning platforms and digital learning systems. Learning management systems, intelligent tutoring platforms, and other online educational environments continuously generate diverse forms of data, ranging from learning behaviors and assignment records to assessment results, interaction logs, and knowledge mastery indicators. Such data is inherently heterogeneous, encompassing both structured formats such as final exam scores and test results, and unstructured formats like discussion texts.

Consequently, the key challenge has shifted from data collection to the effective integration and utilization of multi-source, multi-modal data [4]. In particular, extracting meaningful and actionable knowledge through cross-modal data fusion has emerged as a critical bottleneck in advancing intelligent education.

2 Challenges in Multi-Source Educational Data Mining

2.1 Heterogeneity and Data Fusion Challenge

Educational data is inherently multi-source, multi-modal, and heterogeneous [5]. While different data types capture complementary aspects of student learning, they are often stored and processed in isolation. Such fragmentation hinders the ability to uncover deeper semantic relationships across modalities.

Existing research on data fusion in education is still in early stages and often rely on simple feature concatenation or shallow integration strategies. In this case, they often fail to fully exploit the intrinsic relationships among heterogeneous data sources. Therefore, developing effective multi-source data fusion mechanisms remains a fundamental challenge in EDM.

2.2 Personalized Learning and Intervention Challenge

Every learner is unique, they vary widely in learning styles, cognitive abilities, and educational needs, making traditional "one-size-fits-all" instructional approaches inadequate for addressing such diversity [6]. In this case, EDM provides the potential to design personalized learning pathways through the analysis of student behavior and performance data [7]. However, achieving fine-grained and precise personalized intervention remains challenging.

In the context of multi-source data, achieving accuracy, adaptivity, and real-time responsiveness in personalized recommendations demands deeper integration of heterogeneous information and more sophisticated modeling techniques. Bridging the gap between coarse-grained personalization and precise, dynamic intervention remains an open challenge.

2.3 Interpretability Challenge in EDM Models

Beyond personalization, model interpretability poses another critical challenge in EDM [8]. While machine learning and deep learning methods have achieved strong predictive performance, their "black-box" nature limits their applicability in educational settings.

In practice, to make informed instructional decisions, educators need to understand the reasoning behind model predictions. While for students, they require transparency to build trust in intelligent learning systems. The lack of interpretability not only reduces model usability but also hinders the adoption of model prediction-driven educational technologies.

3 Explainable Knowledge Modeling

To address the challenge of interpretability, explainable knowledge modeling has emerged as a promising direction [9]. Fuzzy systems stand out because they combine human-like linguistic reasoning with mathematically precise handling of uncertainty, giving them a natural advantage in transparency and interpretability. Other explainable approaches such as decision trees, expert systems, and probabilistic models each offer strengths, but they either struggle with ambiguity, lose clarity as complexity grows, or rely on explanations that are less intuitive for non-experts. Compared with these methods, fuzzy systems provide a more flexible, human-aligned way to represent knowledge while still maintaining a fully traceable decision process [12].

Based on fuzzy rules and membership functions, fuzzy knowledge representation provides a natural and interpretable approach [10]. Fuzzy systems represent knowledge in a linguistic and human-understandable form, enabling transparent reasoning processes [11]. This characteristic makes them particularly suitable for educational scenarios, where interpretability and trust are essential.

Fuzzy knowledge has been successfully applied in various domains, such as the Internet of Things, healthcare, transportation, energy, economics, multi-criteria decision-making [13], clinical diagnosis, and special needs education [14]. However, its application in multi-source educational data mining remains relatively underexplored.

Given its ability to handle uncertainty and model gradual learning states, fuzzy knowledge offers a promising solution for integrating heterogeneous educational data while maintaining interpretability. It provides a bridge between data-driven analytics and pedagogical understanding, enabling both accurate and explainable decision-making.

4 Towards Intelligent and Explainable Educational Data Mining

In summary, advancing EDM requires addressing three key challenges, namely, (1) Efficient multi-source and multi-modal data fusion; (2) Fine-grained and adaptive personalized learning; and (3) Interpretable and trustworthy modeling frameworks.

By integrating multi-source educational data and leveraging explainable knowledge representations such as fuzzy systems, it becomes possible to:

Construct personalized learning pathways Enable precise tutoring and intervention Provide transparent decision support for educators and learners

By integrating multi-source educational data and leveraging explainable knowledge representations (e.g., fuzzy systems), it becomes possible to construct personalized learning pathways, enable precise tutoring and intervention, and provide transparent decision support for both educators and learners.

5 Conclusion

Educational data mining plays a vital role in advancing personalized, intelligent, and data-driven education. Effectively fusing multi-source, multi-modal educational data with explainable knowledge modeling is essential to overcoming the limitations in the field of EDM.

This editorial highlights the importance of integrating data fusion and interpretability to build intelligent and trustworthy educational systems. Such integration not only improves the effectiveness of personalized learning and teaching interventions but also strengthen user trust and practical applicability.

Ultimately, advancing EDM through multi-source data fusion and explainable knowledge contributes to the broader goal of education: enabling every learner to achieve their full potential through intelligent and reliable support.

Data Availability Statement

Not applicable.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Baker, R. S., Martin, T., & Rossi, L. M. (2016). Educational data mining and learning analytics. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, 379-396. [CrossRef]

[2] 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]

[3] Quadir, B., Chen, N. S., & Isaias, P. (2022). Analyzing the educational goals, problems and techniques used in educational big data research from 2010 to 2018.

Interactive Learning Environments, *30*(8), 1539–1555. [CrossRef]

[4] Ji, X., Sun, L., & Huang, K. (2025). The construction and implementation direction of personalized learning model based on multimodal data fusion in the context of intelligent education. *Cognitive Systems Research*, *92*, 101379. [CrossRef]

[5] Zhou, Y., Zou, S., Liwang, M., Sun, Y., & Ni, W. (2025). A teaching quality evaluation framework for blended classroom modes with multi-domain heterogeneous data integration. *Expert Systems with Applications*, *289*, 127884. [CrossRef]

[6] Poole, A., & Li, X. (2025). Beyond a one-size-fits-all approach: Considering English language teachers' differential capacities for engaging in and implementing professional learning in China. *Professional Development in Education*, *51*(4), 735-747. [CrossRef]

[7] Hariyanto, Kristianingsih, F. X. D., & Maharani, R. (2025). Artificial intelligence in adaptive education: a systematic review of techniques for personalized learning. *Discover Education*, *4*(1), 458. [CrossRef]

[8] Kalita, E., Oyelere, S. S., Gaftandzhieva, S., Rajesh, K. N., Jagatheesaperumal, S. K., Mohamed, A., ... & Ali, T. (2025). Educational data mining: a 10-year review. *Discover Computing*, *28*(1), 81. [CrossRef]

[9] Wang, Z., Ma, S., Wang, K., & Zhuang, Z. (2025). Rule-guided graph neural networks for explainable knowledge graph reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, pp. 12784–12791). [CrossRef]

[10] Zhang, W., Deng, Z., Wang, G., & Choi, K.-S. (2026). Fuzzy Rule-Based Differentiable Representation Learning. *IEEE Transactions on Neural Networks and Learning Systems*, *37*(2), 603–616. [CrossRef]

[11] Cao, J., Zhou, T., Zhi, S., Lam, S., Ren, G., Zhang, Y., ... & Cai, J. (2024). Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. *Information sciences*, *662*, 120212. [CrossRef]

[12] Capitoli, G., Nobile, M. S., Ambags, E. L., L'Imperio, V., Provenzano, M., & Liò, P. (2025). Assisting clinical diagnosis with interpretable fuzzy probabilistic modelling. *BMC Medical Informatics and Decision Making*, *25*(Suppl 3), 330. [CrossRef]

[13] Nwokoro, C. O., & Ejegwa, P. A. (2025). A decade of picture fuzzy sets in multi-criteria decision-making: A comprehensive review of trends, gaps, and future directions. *Knowledge and Decision Systems with Applications*, *1*, 145–164. [CrossRef]

[14] Elfakki, A. O., Sghaier, S., & Alotaibi, A. A. (2023). An intelligent tool based on fuzzy logic and a 3D virtual learning environment for disabled student academic performance assessment. *Applied Sciences*, *13*(8), 4865. [CrossRef]

Xin Gu completed her Bachelor degree in Information Science with Honours at Massey University, New Zealand and her PhD study at the University of Newcastle, Australia. Her research interests include machine learning, agent-based modeling and simulation, and artificial intelligence. (Email: xin.gu@koi.edu.au)

Fariza Sabrina received her Ph.D. in Computer Science and Engineering from the University of New South Wales, Australia and Master of Engineering (by research) in Electrical and Information Engineering from the University of Sydney, Australia. She has many years of research and development, teaching and industrial experience in information and communication technologies. Her experience includes working as a postdoctoral research fellow and a research scientist in the Networking Research Lab at Commonwealth Scientific and Research Organisation (CSIRO), Australia. Currently she is working as a Associate Professor in the College of ICT at Central Queensland University, Australia. Her current research interest includes Networking, Internet of Things (IoT), Cybersecurity, Blockchain, and Machine Learning. (Email: f.sabrina@cqu.edu.au)

Shaleeza Sohail received her Ph.D. and Masters in Computer Science and Engineering from the University of New South Wales, Australia. Currently she is the Program Convenor (Sydney) for Master of IT and Graduate Certificate of IT in the discipline of Computing and IT at The University of Newcastle. She has been teaching undergraduate and postgraduate courses in the field of IT for more than fifteen years and has extensive experience in teaching and learning design and curriculum development. Dr. Sohail has been actively involved in research in various fields of computing and published a number of papers. Her current research interest includes Networking, Quality of Service, Internet of Things (IoT), Cybersecurity, Blockchain, and Machine Learning. (Email: shaleeza.sohail@newcastle.edu.au)

Zongwen Fan received both his B.S. and M.Sc. degrees from Huaqiao University, China, in 2014 and 2017, respectively. He completed his Ph.D. degree from the University of Newcastle, Australia, in 2021. He is currently a lecturer with the College of Computer Science and Technology, Huaqiao University, China. His main research interests include machine learning, data mining, and fuzzy computing. (Email: zongwen.fan@hqu.edu.cn)