Explainable Learning Outcomes Prediction: Information Fusion Based on Grades Time-Series and Student Behaviors

Yuan-Hao Jiang[*]
Shanghai Institute of Artificial Intelligence for Education
East China Normal University
Shanghai, China
Graduate School
Shanghai Jiao Tong University
Shanghai, China
yuanhao.cs.edu@gmail.com

Zi-Wei Chen[*†]
Jingdezhen Ceramic University
School of Information Engineering
Jingdezhen, China
2220042009@stu.jcu.edu.cn

Cong Zhao
School of Information Engineering
Jingdezhen Ceramic University
Jingdezhen, China
18409237546@163.com

Kezong Tang
Jingdezhen Ceramic University
School of Information Engineering
Jingdezhen, China
tangkezong@jci.edu.cn

Jicong Duan
School of Computer
Jiangsu University of Science and Technology
Zhenjiang, China
College of Automation
Jiangsu University of Science and Technology
Zhenjiang, China
jicong_duan@stu.just.edu.cn

Yizhou Zhou
Shanghai Institute of Artificial Intelligence for Education
East China Normal University
Shanghai, China
zhouyizhou25@gmail.com

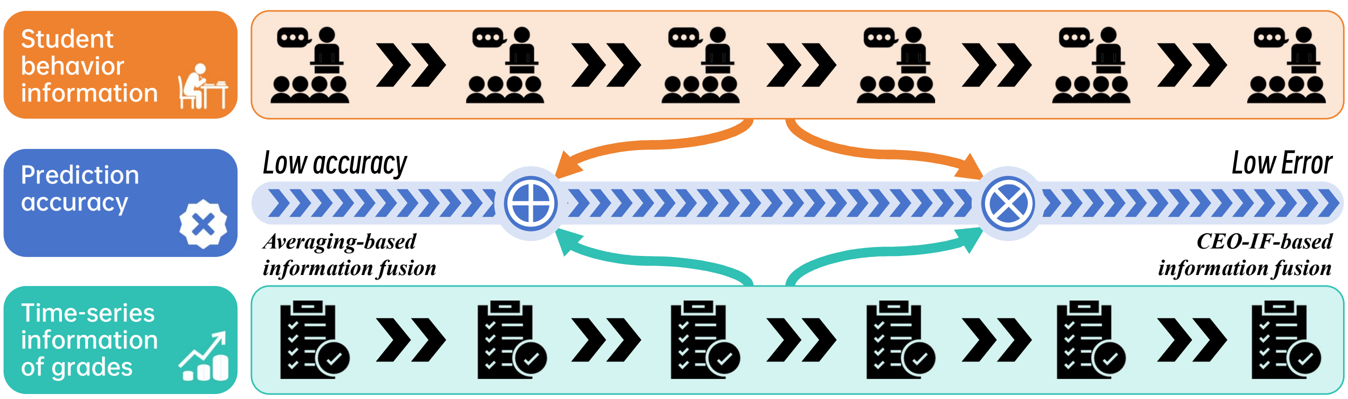


Figure 1: A schematic diagram of information fusion between student behavior information and grade time-series information. Experimental results from this study demonstrate that using a simple average weighting for information fusion decreases the accuracy of learning outcome predictions. However, the proposed Co-Evolutionary Optimization-based Information-fusion Framework (CEO-IF) effectively addresses this issue by implementing effective information fusion.

[*]Both authors contributed equally to this research.
[†]Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3721212

Abstract

Accurately and timely predicting learners' outcomes can assist educators in making instructional decisions or interventions. This helps prevent students from falling into a vicious cycle of decreased academic achievement and increased aversion to learning, potentially leading to dropout. Data-driven models often outperform eXplainable Artificial Intelligence (XAI) models in predicting learning outcomes, yet their lack of interpretability can hinder trust from educators. Therefore, this study developed an XAI information fusion framework that not only extracts potential trends from the time series of student grades to enhance predictive performance but also mines explicit relationships between classroom behaviors and learning outcomes. This reveals the behavioral causes behind changes in grades. Furthermore, we have made public the Dataset for Predicting Outcomes from Time sequences and Student behaviors (DPOTS), and validated the effectiveness of the developed XAI information fusion framework based on DPOTS. The results indicate that, the Mean Absolute Error (MAE) of CEO-IF was reduced by an average of 26.32% compared to the baseline algorithms, and it showed a 22.63% reduction compared to the averaging-based information fusion method. The homepage for the project can be accessed at https://doi.org/10.5281/zenodo.14958102.

CCS Concepts

• Human-centered computing → Human computer interaction (HCI); **• Applied computing → Education**.

Keywords

Large language model, AI agent, agent role play, educational dataset, sustainable development goals, AI for education

ACM Reference Format:

Yuan-Hao Jiang, Zi-Wei Chen, Cong Zhao, Kezong Tang, Jicong Duan, and Yizhou Zhou. 2025. Explainable Learning Outcomes Prediction: Information Fusion Based on Grades Time-Series and Student Behaviors. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3706599.3721212

1 Introduction

Ensuring inclusive and equitable quality education, as well as promoting lifelong learning opportunities for all [11], is one of the Sustainable Development Goals (SDGs) outlined in the United Nations 2030 Agenda for Sustainable Development [9, 17]. The emergence of Artificial Intelligence (AI) technologies has catalyzed rapid transformations in educational practices [22, 29], driving the accelerated development of AI for Education (AIED) [21, 27, 30] and offering viable pathways toward realizing this field [14, 16]. By predicting learning outcomes, educators can identify students experiencing academic difficulties and provide data-driven support for timely interventions and instructional decisions [1]. Furthermore, adjusting teaching strategies based on such predictions can enhance the quality of instruction [43]. This approach also enables teachers to diagnose students' cognitive states [35, 49, 50], facilitating early intervention and targeted support to help students overcome challenges.

To support teachers' decision-making with robust data, eXplainable Artificial Intelligence (XAI) [13] has gained widespread application in education, aiding educators in more accurately predicting learning outcomes and performing detailed learning analysis. Numerous XAI models have already been employed to identify factors influencing learners' performance [20, 48] and to assist teachers in making instructional decisions based on learning analytics [53]. However, some studies indicate that XAI models underperform deep learning algorithms in certain learning outcome tasks [4, 7, 48], particularly when dealing with sequential data such as students' performance over time [13]. Thus, balancing the prediction accuracy for learning outcomes with the transparency and interpretability of models remains an urgent challenge as these AI technologies move towards practical application in education.

This study seeks to accurately and partially explain the prediction of learning outcomes using both grade time-sequence data and student behavior data. We have developed a dataset that integrates both performance time-sequence and student behavior, along with corresponding information fusion methods, which are publicly available. The key contributions of this paper are as follows:

- This study developed an information fusion framework that balances the interpretability of the model with the accuracy of predicting learning outcomes, termed the Co-Evolutionary Optimization-based Information-fusion Framework (CEO-IF).
- We developed the Dataset for Predicting Outcomes from Time sequences and Student behaviors (DPOTS) using a large language model-driven Agent Role Play approach, comprising data from 100 students and four teachers totaling 4,720 entries. We have made the developed DPOTS dataset publicly available at https://doi.org/10.5281/zenodo.14958102 [24].
- We fused time-series grade information and student behavior data for predicting learning outcomes. We found that typical weighted average information fusion methods reduced prediction accuracy; however, our proposed fusion method significantly improved it, as shown in Fig. 1.

In the remainder of this study, related work is introduced in Appendix A. Then, in Section 2, the proposed methods and dataset details are detailed, including the development of the DPOTS dataset. Simulation experiments are conducted in Section 3 to validate the superiority of the proposed information fusion method. Finally, Section 4 summarizes the conclusions of this research.

2 Methodology

2.1 Development of the DPOTS Dataset Supported by Agent Role Play

This study aims to predict learning outcomes by integrating grade time-sequence data with student behavioral information. Specifically, we generate student behavioral data based on the CPS coding framework [45], which provides a structured method for representing students' behaviors in the classroom.The CPS coding framework provides a comprehensive overview of student behavior, capturing various aspects of their classroom conduct. Detailed information can be found in Appendix B.

For teachers, we introduce the Classroom Atmosphere, Teaching Level, Interactions with Students, and Frequency of Homework Assignments (CTIF) framework to implement ARP for teachers [52]. These dimensions include teacher-student interactions, teaching level, classroom atmosphere, and homework assignments [13]. Fig. 2 presents a case study of ARP for four teachers based on this framework. As shown in Fig. 2, the four teachers display distinct teaching levels, classroom atmospheres, and personal characteristics, all of which impact students' learning outcomes.

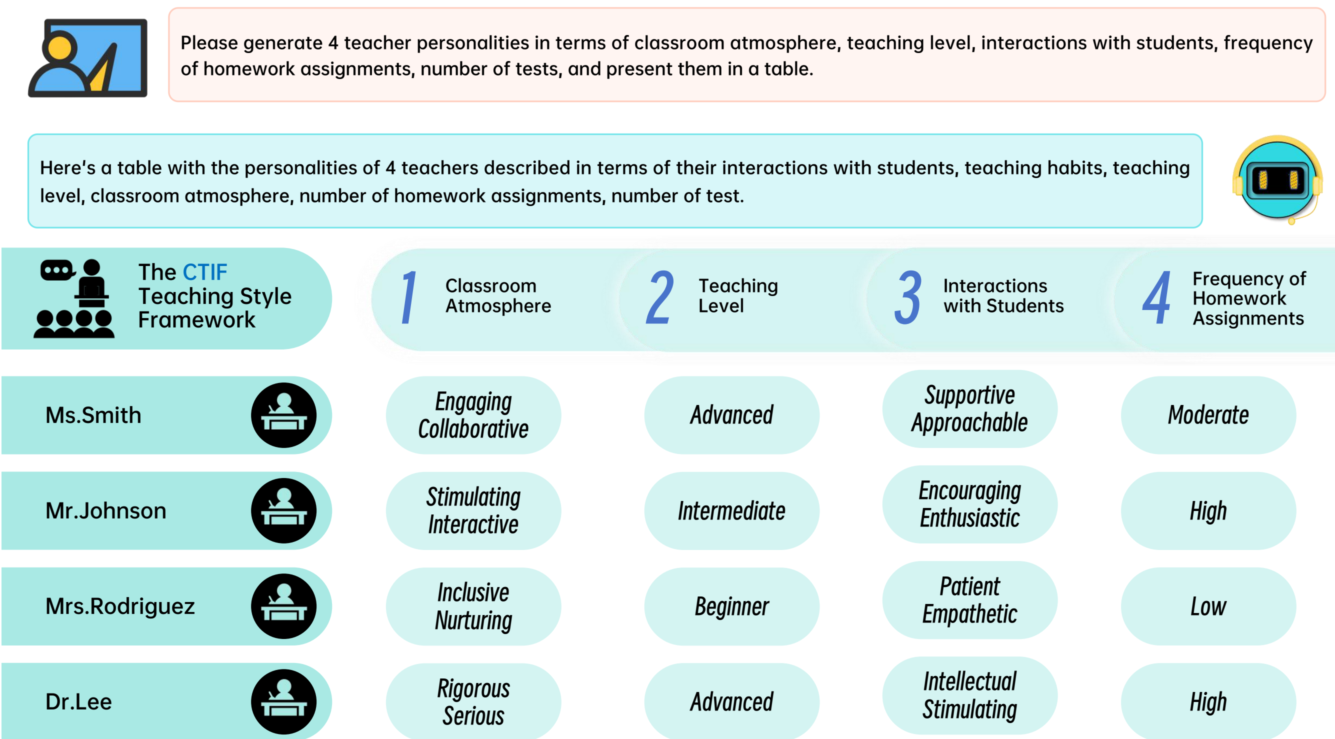


Figure 2: ARP for 4 teachers based on the CTIF framework. Based on this framework, AI agents can more realistically simulate teachers' behaviors and make differentiated actions during the teaching process.

2.2 Information Fusion Framework CEO-IF

In order to enhance the accuracy of learning outcome predictions while improving the transparency and interpretability of algorithms, we propose a novel information fusion framework, denoted as CEO-IF. This framework integrates performance time-sequence data with student behavioral data to predict learning outcomes, corresponding to the performance time-sequence model and the behavior representation model within the CEO-IF framework. The proposed framework represents an improvement over the Collaborative Structure Search Framework (CSSF) [36] and the Differential Evolution algorithm based on the Transdifferentiation Strategy (DE-TS) [25]. It combines the advantages of CSSF in handling multi-subgroup information fusion with the superior generalization capabilities of DE-TS in addressing complex tasks. This methodology enhances the interpretability of the model while maintaining high prediction accuracy, thereby offering valuable support for informed instructional decisions.

The overall framework for information fusion is depicted in Fig. 3. Initially, student behavioral data (encoded as CPS) are input into the DT model. Concurrently, historical academic performance data are fed into an LSTM model to extract features of student information across two dimensions: behavioral patterns and performance trajectories. Subsequently, the information fusion process begins. Following population initialization, the number of function evaluation counts, FE, is set to zero. The DE algorithm generates subpopulations which are then ranked according to fitness values and divided into three sub-populations: a superior sub-population, an exploration sub-population, and an eliminated sub-population. The developed CEO-IF algorithm selects the number of individuals from the superior sub-population to proceed to the next iteration based on the parameter AP. The exploration sub-population undergoes individual optimization of offspring, supported by the transdifferentiation strategy [25]. The eliminated sub-population is directly discarded. The algorithm merges the offspring with the parental population and sorts them, retaining only the individuals with higher fitness values. After completing an iteration, the evaluation count is incremented. This process repeats until the maximum number of function evaluations, $maxFE$, is reached, at which point the algorithm terminates and returns the optimal individual, I_{best}. Based on I_{best}, the predicted students' learning outcomes from the CEO-IF framework can be calculated and utilized to support instructional decision-making by teachers. Due to the proposed CEO-IF information fusion method being based on evolutionary computation rather than deep learning, it inherently possesses lower complexity and computational requirements.

3 Experiments and Results

3.1 Experimental Setup and Dataset Analysis

To facilitate subsequent information fusion operations, we first applied the standard DT and LSTM models to learn the learning behavior representation data and performance time-sequence data,

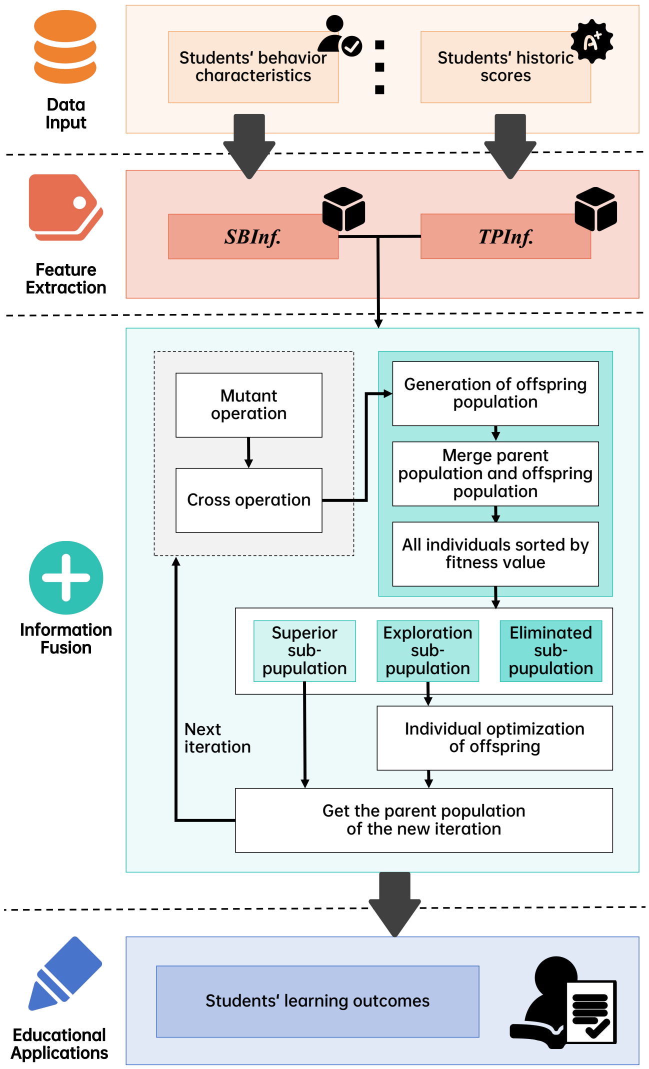


Figure 3: Overview of the CEO-IF information fusion framework. The developed information fusion framework encompasses four key components: data input, feature extraction, information fusion, and educational applications.

respectively, in preparation for the fusion operations and result comparisons. Initially, the DT model is used to train student behavioral features (i.e., CPS encoding). Subsequently, the historical performance data from 25 students in each class are organized, and predictions are made for each class. The LSTM model is then used to predict student performance by analyzing the historical performance time-sequence data of all students. The comparison of final test scores with predicted scores will be presented in the subsequent content of this chapter. In Table 1, we present an overview of the entire dataset. The detailed information about the experimental setup and parameter settings can be found in the Appendix D.

3.2 Comparison of Information Fusion Effects

In this section, we conduct a comparative analysis of the errors between the predicted and actual scores for the four types of information discussed previously. The procedures outlined earlier enable a more intuitive evaluation of the model's performance in prediction tasks. The comparison between the predicted and actual values is illustrated in Fig. 4. By calculating the differences between the predicted and actual values, we obtain the absolute errors.

Subfigures a-d in Fig. 4 present the prediction results for $SBInf$., $TPInf$.,CEO-IF-based information fusion and the Averaging-based information fusion method. Additionally, we introduce the Multi-Subgroup CEO algorithm to facilitate the information fusion process,CEO-IF-based information fusion. The mean errors for the four types of information across each class are indicated. The prediction results for student scores, derived from the absolute errors between the predicted and actual values, are presented in the chart. The absolute errors are arranged in ascending order, with most of the predicted values in Classes 1-4 falling within the range of [0, 5]. However, the rate of error growth accelerates in the later stages. When analyzing the various types of information, $SBInf$. demonstrates relatively strong performance in certain predictions, with some predictions in each class exhibiting an absolute error of zero. However, larger errors emerge later, with Class 4 displaying absolute errors exceeding 20. $TPInf$. performs poorly in Classes 1 and 4, with mean errors of 2.75 and 4.23, respectively. The figure illustrates that $TPInf$. has the highest absolute error values in each class. The performance of Averaging-based information fusion is moderate, with a moderate error growth rate and average error values. Overall, it performs adequately, but the maximum error values in Classes 1-3 are the lowest. In Classes 1 and 4, the mean errors for Averaging-based information fusion are slightly lower than those for $SBInf$. alone, while they are the lowest in Classes 2 and 3. In the early stages, the absolute error across Classes 1-4 are low, indicating good performance.

3.3 Interpretability Analysis

In this section, we focus on the decision tree model's ability to analyze students' behavioral characteristics and predict their academic performance. A case study involving Student 1 and Student 3 from Class 1.In a DT regression model, feature importance evaluates each feature's contribution to the model's predictive ability. It reflects the frequency with which each feature is used to split nodes or the magnitude of information gain during the building models. Thus, feature importance allows us to observe the varying influence of different features. The feature importance analysis quantifies the contribution of each feature to the model's predictive outcomes. Fig. 5 displays a partial decision tree structure for Class 1, where the two most important features are "Builds on others' ideas to improve solutions" and "Does not respond when spoken to by others", as indicated by the black boxes.

3.4 CEO-IF Information Fusion Analysis

In this section, we analyze the underlying reasons for the superior performance of the CEO-IF framework. The absolute error values for 100 students are visualized using a box plot, followed

Data Type	Dimension	Data Count	Data Type	Dimension	Data Count
CTIF	4	16	CPIPIP	6	600
Number of Tests	1	4	CPS	16	1600
Score Data	25	2500	Total	-	4720

Table 1: Overall framework of the DPOTS dataset.

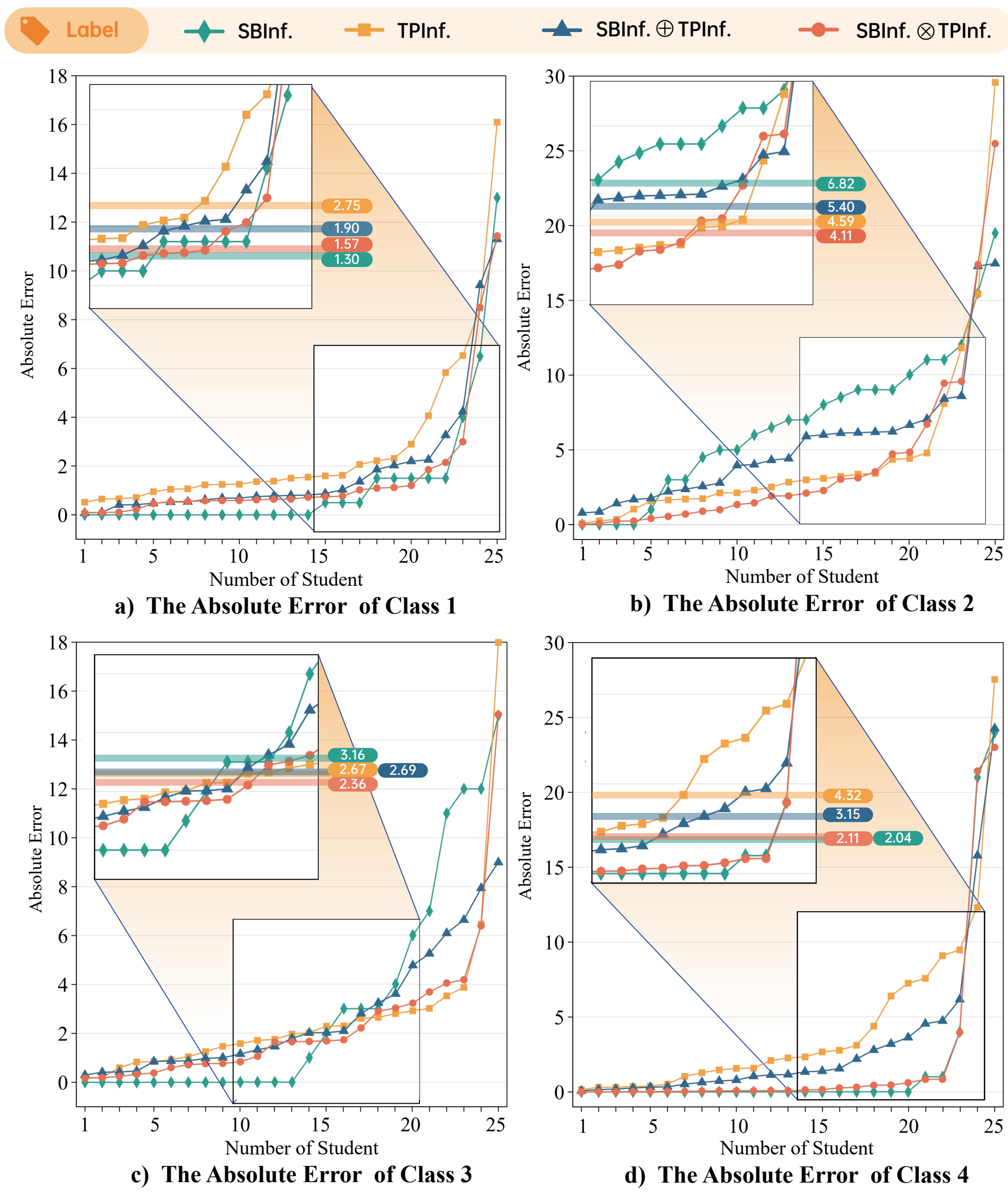


Figure 4: Line chart of absolute errors. This figure illustrates the error distribution of predictions for behavioral information, performance time-sequence information, and the information fusion framework across four classes. The errors are sorted in ascending order to display the growth rates of errors for different methods. Each subplot includes a horizontal line indicating the mean absolute error for each method, with the specific numerical value labeled to the right of the line.

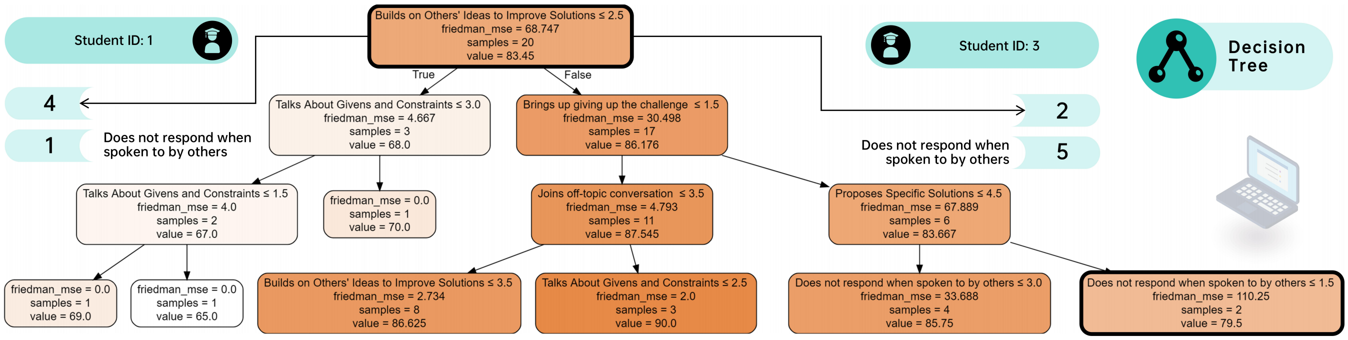


Figure 5: Case analysis of interpretability. The figure presents a partial decision tree structure for Class 1, where academic performance is predicted through regression based on students' behavioral characteristics.

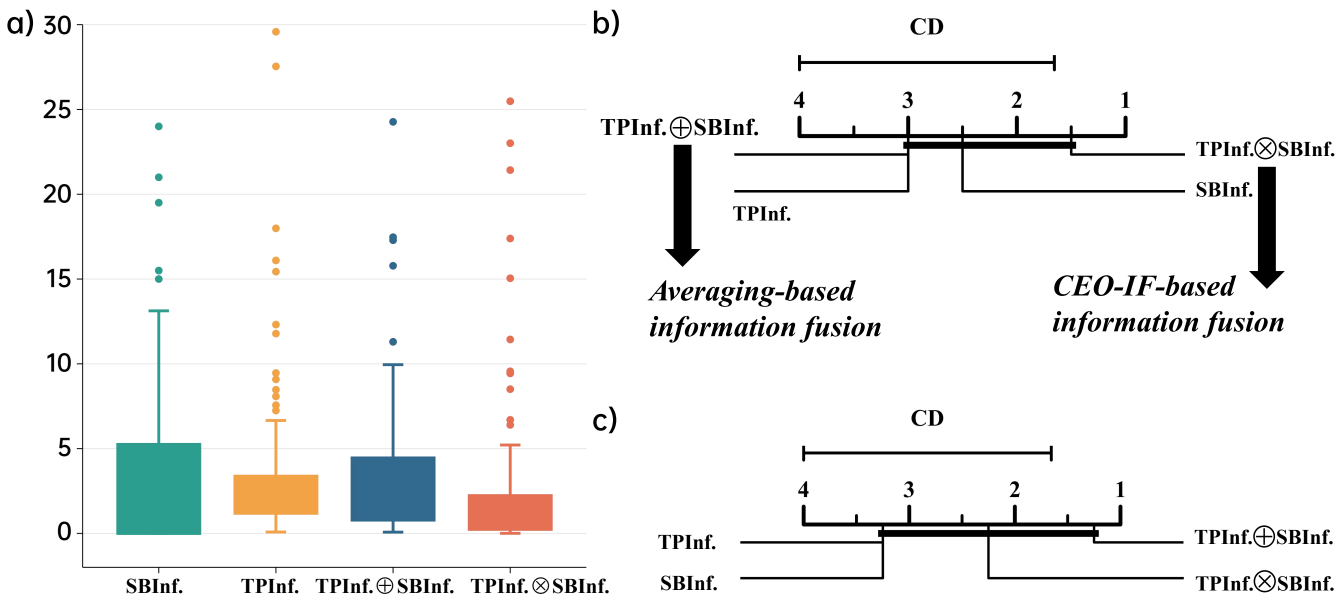


Figure 6: Comparative analysis of results for the CEO-IF information fusion framework. Subfigure (a) presents boxplots of the error distributions using different types of information fusion methods. Subfigure (b) shows the results of the Friedman test ranking for error distributions at the class level, while subfigure (c) displays the Friedman test ranking for the standard deviations at the class level.

by a detailed analysis. Additionally, Fig. 6 examines the prediction results for all students, and compares the outcomes of different prediction methods: CEO-IF-based information fusion, $SBInf$., Averaging-based information fusion, and $TPInf$. A Friedman test is conducted to compare and rank each prediction scenario, and the standard deviation of error distribution across the class dimension is also evaluated.

As depicted in Fig. 6 (a), the data distribution for CEO-IF-based information fusion is relatively concentrated. Compared to $TPInf$. and Averaging-based information fusion, its box position is closer to 0, with the data distribution focused in the [0,5] range. This indicates that, following weighted information fusion, the predictive performance has improved relative to traditional prediction methods. However, some outliers remain present. From the information in Fig. 6 (b), it is evident that CEO-IF-based information fusion ranks the highest, indicating that it provides more accurate predictions. The ranking of the remaining models is as follows: $SBInf$., Averaging-based information fusion, and $TPInf$., with the latter two ranked equally. As shown in Fig. 6 (c), although the standard deviation for Averaging-based information fusion is relatively low, it is essentially achieved by averaging the errors from DT and LSTM. This suggests that Averaging-based information fusion distributes errors more evenly across different classes. However, when compared to CEO-IF, Averaging-based information fusion exhibits higher overall error values. Therefore, CEO-IF outperforms both

LSTM and DT in terms of standard deviation, maintaining lower errors while balancing interpretability and predictive accuracy.

We further analyzed the error data. Specifically, the Mean Absolute Error (MAE) of the CEO-IF was compared against that of a baseline algorithm by dividing the MAE of CEO-IF by the MAE of the baseline and then subtracting the result from 100% to quantify the percentage reduction in MAE achieved by our proposed method. Subsequently, the average MAE of all comparison algorithms was calculated to determine the overall percentage reduction in MAE achieved by CEO-IF compared to the average of the baseline algorithms. Similarly, the MAE of the CEO-IF information fusion technology was divided by the MAE of a averaging-based information fusion approach, and the result was subtracted from 100% to calculate the percentage reduction in MAE of our proposed fusion method relative to the averaging-based fusion method. The results indicate that, in the task of predicting student learning outcomes, the MAE of CEO-IF was reduced by an average of 26.32% compared to the baseline algorithms, and it showed a 22.63% reduction compared to the averaging-based information fusion method.

4 Limitations

This study developed a multi-agent-based synthetic dataset, DPOTS [24], which encompasses structured data on student behaviors and performances. Although existing educational frameworks such as CTIF and CPS guided the simulation process of the multi-agents, it is undeniable that behavior simulations supported by LLMs might limit the generalizability of our findings due to potential deviations from real-world scenarios. On the one hand, collecting and constructing the corresponding real-world datasets could further validate the effectiveness of the proposed method, although it is more time consuming and challenging compared to generating data via LLMs. On the other hand, researching how to better align LLMs with human values could mitigate this issue to some extent, representing an intriguing research direction that we are currently pursuing. Furthermore, this study also engages in discussions and experiments surrounding the topic of interpretability. While the case study results presented in this research can assist educators in understanding the potential factors influencing student performance, it is undeniable that qualitative or quantitative research on interpretability would be beneficial to this study.

5 Discussion and Conclusion

This study introduces the DPOTS dataset, which is designed to predict student learning outcomes and identify influencing factors. The dataset is constructed using AI agents generated by large language models, which characterize teachers and students. Based on this dataset, this paper proposes the CEO-IF framework, which integrates student behavioral information and performance time-sequence data to balance interpretability and prediction accuracy. The experimental results demonstrate that the proposed information fusion framework outperforms methods that rely solely on either behavioral information or academic performance time-sequence data, leading to a reduction in prediction errors within a certain range. By examining the factors that influence student learning outcomes, educators can improve instructional design. Simultaneously, students can gain insight into areas of weakness within the subject matter through performance predictions. This research provides valuable guidance for identifying students at risk of poor academic performance, optimizing teaching methods, enhancing education quality, and fostering the development of personalized intelligent education. Future research will focus on further optimizing learning outcomes, with an emphasis on personalized teaching, to contribute to the high-quality development of education.

Acknowledgments

This work was partially supported by the Special Foundation for Interdisciplinary Talent Training in "AI Empowered Psychology / Education" of the School of Computer Science and Technology, East China Normal University, under the Grant 2024JCRC-03, and the Doctoral Research and Innovation Foundation of the School of Computer Science and Technology, East China Normal University, under the Grant 2023KYCX-03, and the National Innovative Training Program for College Students of Jingdezhen Ceramic University, under Grant 202410408028, and the Jiangxi Provincial Innovative Training Program for College Students of Jingdezhen Ceramic University, under Grant S202410408033. In this study, all the original data were collected by Zi-Wei Chen, and are now available on the project homepage. The authors would like to express their gratitude to Hui-Jun Chen from Shanghai Normal University and Zinuo Gu from East China Normal University for their valuable contributions to the documentation and data processing.

References

[1] Mohamed Abdel-Basset, Gunasekaran Manogaran, Mai Mohamed, and Ehab Rushdy. 2019. Internet of things in smart education environment: Supportive framework in the decision-making process. *Concurrency and Computation: Practice and Experience* 31, 10 (2019), e4515. doi:10.1002/cpe.4515 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.4515.

[2] Mohamed Abdel-Basset, Reda Mohamed, Mohammed Jameel, and Mohamed Abouhawwash. 2023. Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems. *Knowledge-Based Systems* 262 (2023), 110248. doi:10.1016/j.knosys.2022.110248

[3] Abtahi Ahmed, Farzana Akter Nipa, Wasi Uddin Bhuyian, Khaled Md Mushfique, Kamrul Islam Shahin, Huu-Hoa Nguyen, and Dewan Md. Farid. 2024. Students' performance prediction employing Decision Tree. *CTU Journal of Innovation and Sustainable Development* 16 (2024), 42–51. Issue Special issue: ISDS. doi:10.22144/ctujoisd.2024.321

[4] Mohammed Akour, Hiba Al Sghaier, and Osama Al Qasem. 2020. The effectiveness of using deep learning algorithms in predicting students achievements. *Indonesian Journal of Electrical Engineering and Computer Science* 19, 1 (2020), 388. doi:10.11591/ijeecs.v19.i1.pp388-394

[5] Nicholas Aksamit, Jinqiang Hou, Yifeng Li, and Beatrice Ombuki-Berman. 2024. Integrating Transformers and Many-Objective Optimization for Cancer Drug Design. doi:10.21203/rs.3.rs-4229436/v1

[6] Safwan Mahmood Al-Selwi, Mohd Fadzil Hassan, Said Jadid Abdulkadir, Amgad Muneer, Ebrahim Hamid Sumiea, Alawi Alqushaibi, and Mohammed Gamal Ragab. 2024. RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. *Journal of King Saud University - Computer and Information Sciences* 36, 5 (2024), 102068. doi:10.1016/j.jksuci.2024.102068

[7] Bayan Alnasyan, Mohammed Basheri, and Madini Alassafi. 2024. The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review. *Computers and Education: Artificial Intelligence* 6 (2024), 100231. doi:10.1016/j.caeai.2024.100231

[8] Navid Behmanesh-Fard, Hossein Yazdanjouei, Mohammad Shokouhifar, and Frank Werner. 2023. Mathematical Circuit Root Simplification Using an Ensemble Heuristic–Metaheuristic Algorithm. *Mathematics* 11, 6 (2023), 1498. doi:10.3390/math11061498

[9] Frank Biermann, Thomas Hickmann, Carole-Anne Sénit, Marianne Beisheim, Steven Bernstein, Pamela Chasek, Leonie Grob, Rakhyun E. Kim, Louis J. Kotzé, Måns Nilsson, Andrea Ordóñez Llanos, Chukwumerije Okereke, Prajal Pradhan, Rob Raven, Yixian Sun, Marjanneke J. Vijge, Detlef Van Vuuren, and Birka Wicke.

2022. Scientific evidence on the political impact of the Sustainable Development Goals. *Nat Sustain* 5, 9 (2022), 795–800. doi:10.1038/s41893-022-00909-5

[10] Hani Sami Brdesee, Wafaa Alsaggaf, Naif Aljohani, and Saeed-Ul Hassan. 2022. Predictive Model Using a Machine Learning Approach for Enhancing the Retention Rate of Students At-Risk:. *International Journal on Semantic Web and Information Systems* 18, 1 (2022), 1–21. doi:10.4018/IJSWIS.299859

[11] Maia Chankseliani and Tristan McCowan. 2021. Higher education and the Sustainable Development Goals. *High Educ* 81, 1 (2021), 1–8. doi:10.1007/s10734-020-00652-w

[12] Fu Chen and Ying Cui. 2020. Utilizing Student Time Series Behaviour in Learning Management Systems for Early Prediction of Course Performance. *JLA* 7, 2 (2020), 1–17. doi:10.18608/jla.2020.72.1

[13] Zi-Wei Chen, Kezong Tang, Yuan-Hao Jiang, Ling Dai, Hui-Jun Chen, and Xiao-Bin Chen. 2024. Assessing the Learning Outcomes of Students Portrayed by AI Agents Based on Grade Timelines and Student Behaviors. In *Enhancing Educational Practices: Strategies for Assessing and Improving Learning Outcomes*, Yuang Wei, Changyong Qi, Yuan-Hao Jiang, and Ling Dai (Eds.). Nova Science Publishers, New York, NY, USA, 133–154. https://doi.org/10.52305/RUIG5131

[14] Mutlu Cukurova. 2024. The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *Brit J Educational Tech* 55, 6 (2024), bjet.13514. doi:10.1111/bjet.13514

[15] Dr. Benaissa Brahim, Masakazu Kobayashi, Musaddiq Al Ali, Tawfiq Khatir, and Mohamed El Amine Elaissaoui Elmeliani. 2024. Metaheuristic Optimization Algorithms: an overview. *HCMCOUJS - Advances in Computational Structures* 14, 1 (2024), 47–62. doi:10.46223/HCMCOUJS.acs.en.14.1.47.2024

[16] Said Elbanna and Loreta Armstrong. 2024. Exploring the integration of ChatGPT in education: adapting for the future. *MSAR* 3, 1 (2024), 16–29. doi:10.1108/MSAR-03-2023-0016

[17] Luis Miguel Fonseca, José Pedro Domingues, and Alina Mihaela Dima. 2020. Mapping the Sustainable Development Goals Relationships. *Sustainability* 12, 8 (2020), 3359. doi:10.3390/su12083359

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press, Cambridge, MA, USA. Google-Books-ID: omivDQAAQBAJ.

[19] Harold Goss. 2022. Student Learning Outcomes Assessment in Higher Education and in Academic Libraries: A Review of the Literature. *The Journal of Academic Librarianship* 48, 2 (2022), 102485. doi:10.1016/j.acalib.2021.102485

[20] Thao-Trang Huynh-Cam, Long-Sheng Chen, and Huynh Le. 2021. Using Decision Trees and Random Forest Algorithms to Predict and Determine Factors Contributing to First-Year University Students' Learning Performance. *Algorithms* 14, 11 (2021), 318. doi:10.3390/a14110318

[21] Augustine Osamor Ifelebuegu, Peace Kulume, and Perpetua Cherukut. 2023. Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning and Teaching* 6, 2 (2023), 332–345. doi:10.37074/jalt.2023.6.2.29 Number: 2.

[22] Li Jiang. 2023. AI-Driven Educational Reform: The Impact and Prospects of ChatGPT/GPT. *Journal of East China Normal University(Educational Sciences)* 41, 7 (2023), 143. doi:10.16382/j.cnki.1000-5560.2023.07.013

[23] Yuan-Hao Jiang. 2023. *Research on Differential Evolutionary Integrated Control Algorithm for Rail-Guided Vehicle*. Master's thesis. Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, China. doi:10.13140/RG.2.2.28367.50087/1

[24] Yuan-Hao Jiang, Zi-Wei Chen, Cong Zhao, Kezong Tang, Jicong Duan, and Yizhou Zhou. 2025. *DPOTS: eXplainable Artificial Intelligence Dataset for Predicting Outcomes from Time Sequences and Student Behaviors*. East China Normal University. doi:10.5281/zenodo.14958102

[25] Yuan-Hao Jiang, Shang Gao, Yu-Hang Yin, Zi-Fan Xu, and Shao-Yong Wang. 2023. A control system of rail-guided vehicle assisted by transdifferentiation strategy of lower organisms. *Engineering Applications of Artificial Intelligence* 123 (2023), 106353. doi:10.1016/j.engappai.2023.106353

[26] Yuan-Hao Jiang, Ruijia Li, Yuang Wei, Rui Jia, Xiaobao Shao, Hanglei Hu, and Bo Jiang. 2024. Synchronizing Verbal Responses and Board Writing for Multimodal Math Instruction with LLMs. In *Proceedings of the 4th Workshop on Mathematical Reasoning and AI at the 38th Annual Conference on Neural Information Processing Systems (NeurIPS 2024)*. Proceedings of Machine Learning Research (PMLR), Vancouver Convention Centre, Vancouver, BC, Canada, 46–59. https://openreview.net/forum?id=esbIrV8N12

[27] Yuan-Hao Jiang, Ruijia Li, Yizhou Zhou, Changyong Qi, Hanglei Hu, Yuang Wei, Bo Jiang, and Yonghe Wu. 2024. AI Agent for Education: von Neumann Multi-Agent System Framework. doi:10.48550/arXiv.2501.00083 arXiv:2501.00083 [cs]

[28] Yuan-Hao Jiang, Jinxin Shi, Yukun Tu, Yizhou Zhou, Wenxuan Zhang, and Yuang Wei. 2024. For Learners: AI Agent is All You Need. In *Enhancing Educational Practices: Strategies for Assessing and Improving Learning Outcomes*, Yuang Wei, Changyong Qi, Yuan-Hao Jiang, and Ling Dai (Eds.). Nova Science Publishers, New York, NY, USA, 21–46. https://doi.org/10.52305/RUIG5131

[29] Johny Karam. 2023. Reforming Higher Education Through AI. In *Governance in Higher Education: Global Reform and Trends in the MENA Region*, Nehme Azoury and Georges Yahchouchi (Eds.). Springer Nature Switzerland, Cham, Switzerland, 275–306. doi:10.1007/978-3-031-40586-0_12

[30] Jinhee Kim. 2024. Leading teachers' perspective on teacher-AI collaboration in education. *Educ Inf Technol* 29, 7 (2024), 8693–8724. doi:10.1007/s10639-023-12109-5

[31] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. doi:10.48550/arXiv.1412.6980 arXiv:1412.6980 [cs]

[32] Yu-Ju Lan and Nian-Shing Chen. 2024. Teachers' agency in the era of LLM and generative AI: Designing pedagogical AI agents. *Educational Technology & Society* 27, 1 (2024), I–XVIII. https://www.jstor.org/stable/48754837 Publisher: International Forum of Educational Technology & Society, National Taiwan Normal University, Taiwan.

[33] Chen Li, Guo Chen, Gaoqi Liang, Fengji Luo, Junhua Zhao, and Zhao Yang Dong. 2022. Integrated optimization algorithm: A metaheuristic approach for complicated optimization. *Information Sciences* 586 (2022), 424–449. doi:10.1016/j.ins.2021.11.043

[34] Nian Li, Chen Gao, Yong Li, and Qingmin Liao. 2023. Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. doi:10.2139/ssrn.4606937

[35] Xiaoyu Li, Shaoyang Guo, Jin Wu, and Chanjin Zheng. 2025. An interpretable polytomous cognitive diagnosis framework for predicting examinee performance. *Information Processing & Management* 62, 1 (2025), 103913. doi:10.1016/j.ipm.2024.103913

[36] Tian-Yi Liu, Yuan-Hao Jiang, Yuang Wei, Xun Wang, Shucheng Huang, and Ling Dai. 2024. Educational Practices and Algorithmic Framework for Promoting Sustainable Development in Education by Identifying Real-World Learning Paths. *Sustainability* 16, 16 (2024), 6871. doi:10.3390/su16166871

[37] Amalesh Kumar Manna, Subhajit Das, Ali Akbar Shaikh, Asoke Kumar Bhunia, and Ilkyeong Moon. 2023. Carbon emission controlled investment and warranty policy based production inventory model via meta-heuristic algorithms. *Computers & Industrial Engineering* 177 (2023), 109001. doi:10.1016/j.cie.2023.109001

[38] Ethan Mollick, Lilach Mollick, Natalie Bach, LJ Ciccarelli, Ben Przystanski, and Daniel Ravipinto. 2024. AI Agents and Education: Simulated Practice at Scale. doi:10.48550/ARXIV.2407.12796 Version Number: 1.

[39] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–22. doi:10.1145/3586183.3606763

[40] Alejandro Rodríguez-Molina, Efrén Mezura-Montes, Miguel G. Villarreal-Cervantes, and Mario Aldape-Pérez. 2020. Multi-objective meta-heuristic optimization in intelligent control: A survey on the controller tuning problem. *Applied Soft Computing* 93 (2020), 106342. doi:10.1016/j.asoc.2020.106342

[41] M. A. Salahli, T. Gasimzade, V. Salahli, F. Alasgarova, and A. Guliyev. 2022. Use of Decision Tree and Fuzzy Logic Methods to Predict Academic Achievement of University Freshmen. In *11th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions and Artificial Intelligence - ICSCCW-2021*, Rafik A. Aliev, Janusz Kacprzyk, Witold Pedrycz, Mo Jamshidi, Mustafa Babanli, and Fahreddin M. Sadikoglu (Eds.). Springer International Publishing, Cham, 156–164. doi:10.1007/978-3-030-92127-9_24

[42] Gesina Schwalbe and Bettina Finzel. 2024. A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. *Data Min Knowl Disc* 38, 5 (2024), 3043–3101. doi:10.1007/s10618-022-00867-8

[43] Lei Shi and Xiaoling Wu. 2022. Generation and Optimization of Teaching Decision Generation Under a Smart Teaching Environment. *International Journal of Emerging Technologies in Learning (iJET)* 17, 5 (2022), 252–265. https://www.learntechlib.org/p/222853/ Publisher: International Journal of Emerging Technology in Learning.

[44] SPSSPRO. 2024. *SPSSPRO Online data analysis platform*. SPSSPRO Team. https://www.spsspro.com/

[45] Chen Sun, Valerie J. Shute, Angela Stewart, Jade Yonehiro, Nicholas Duran, and Sidney D'Mello. 2020. Towards a generalized competency model of collaborative problem solving. *Computers & Education* 143 (2020), 103672. doi:10.1016/j.compedu.2019.103672

[46] Laura Tensen and Klaus Fischer. 2024. Evaluating hybrid speciation and swamping in wild carnivores with a decision-tree approach. *Conservation Biology* 38, 1 (2024), e14197. doi:10.1111/cobi.14197

[47] Behnam Vahdani, S.T.A. Niaki, and S. Aslanzade. 2017. Production-inventory-routing coordination with capacity and time window constraints for perishable products: Heuristic and meta-heuristic algorithms. *Journal of Cleaner Production* 161 (2017), 598–618. doi:10.1016/j.jclepro.2017.05.113

[48] Hajra Waheed, Saeed-Ul Hassan, Naif Radi Aljohani, Julie Hardman, Salem Alelyani, and Raheel Nawaz. 2020. Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior* 104 (2020), 106189. doi:10.1016/j.chb.2019.106189

[49] Wenyi Wang, Yukun Tu, Lihong Song, Juanjuan Zheng, and Teng Wang. 2021. An Adaptive Design for Item Parameter Online Estimation and Q-Matrix Online Calibration in CD-CAT. *Front. Psychol.* 12 (2021), 710497. doi:10.3389/fpsyg.2021.710497

[50] Wenyi Wang, Juanjuan Zheng, Lihong Song, Yukun Tu, and Peng Gao. 2021. Test Assembly for Cognitive Diagnosis Using Mixed-Integer Linear Programming. *Front. Psychol.* 12 (2021), 623077. doi:10.3389/fpsyg.2021.623077

[51] Xusheng Wang. 2022. Designing digital circuits based on quantum-dots cellular automata using nature-inspired metaheuristic algorithms: A systematic literature review. *Optik* 262 (2022), 169251. doi:10.1016/j.ijleo.2022.169251

[52] Yuang Wei, Changyong Qi, Yuan-Hao Jiang, and Ling Dai. 2024. Preface: The Future of AI for Education. In *Enhancing Educational Practices: Strategies for Assessing and Improving Learning Outcomes*, Yuang Wei, Changyong Qi, Yuan-Hao Jiang, and Ling Dai (Eds.). Nova Science Publishers, New York, NY, USA, vii–x. https://doi.org/10.52305/RUIG5131

[53] Mustafa Yağcı. 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learn. Environ.* 9, 1 (2022), 11. doi:10.1186/s40561-022-00192-z

[54] Zheyuan Zhang, Daniel Zhang-Li, Jifan Yu, Linlu Gong, Jinchang Zhou, Zhanxin Hao, Jianxiao Jiang, Jie Cao, Huiqin Liu, Zhiyuan Liu, Lei Hou, and Juanzi Li. 2024. Simulating Classroom Education with LLM-Empowered Agents. doi:10.48550/ARXIV.2406.19226 Version Number: 2.

A Related Work

A.1 Prediction of Learning Outcomes

Student learning outcomes serve as critical indicators of both students' learning capabilities and the quality of education. As such, the assessment of learning outcomes is a fundamental task in educational research [19]. Learning outcomes are influenced by a range of factors, including non-academic elements [3], the educational environment [41]. Some studies have employed Decision Trees(DT) to analyze these factors, identifying key determinants that impact learning outcomes, thus providing a foundation for prediction. In recent years, deep learning approaches have been integrated into learning performance research, yielding improved prediction results. The LSTM model, a specialized variant of RNN, enhances the capacity to process long-term dependencies and has been widely adopted for predicting learning outcomes [7]. LSTM also has been utilized to assess students' time-series behavioral data [12] and predict those at risk of failure [10]. Therefore, achieving a balance between interpretability and grade time-sequence data represents a significant challenge in enhancing the practical applicability of these models.

A.2 Agent-based Simulation and Datasets

The increasing prominence of AI agents has led to the expanding application of agent-based models across various scenarios [27], relying on the autonomy, collaboration, and adaptability of these agents [28]. The emergence of multi-agent simulations and associated datasets is primarily driven by the need to simulate real-world scenarios to address the issues that arise within these contexts. Generative AI technologies, such as multi-agent systems, have enhanced interactive applications by simulating human behavior and generating credible individual and emergent social actions, as exemplified by Stanford's Town model [39]. Agents are employed to identify uncertainties and complexities across various domains. In macroeconomics, agents are used to overcome constraints by simulating human decision-making processes in economic environments, thus addressing pressing challenges [34]. The multi-agent systems across diverse fields, through scenario simulations, provides viable solutions to real-world problems.

A.3 Educational Applications of Large Language Models

In recent years, large language model (LLM) technology has advanced rapidly and become increasingly prevalent in the field of intelligent education [26]. AI agents, generated by LLMs, have the capacity to simulate real-world classroom scenarios, thereby transforming traditional educational models [38]. Recent studies have proposed the design of instructional AI agents, addressing key aspects such as functionality, programming, and structural design, while providing examples of their application in educational settings [32]. LLMs observe AI agents' interactions within the classroom, enhancing the student learning process [54]. This paper utilizes LLMs to generate AI agents for both teachers and students, simulating real classroom scenarios to address potential challenges and predict students' learning outcomes.

B Development of the DPOTS Dataset Supported by Agent Role Play

To simulate student behavior, we employ AI agents to capture the frequency of occurrence of each event within the CPS coding framework over specific time intervals. The developed DPOTS dataset includes 100 students, distributed evenly across four classes, each consisting of 25 students, as show in the Fig. 7. These four classes are taught by different instructors. By predicting student behavior based on real-world personality traits, the AI agents offer valuable insights for assessing learning outcomes.

Additionally,the generation of student grades is influenced by a variety of factors, including student behavioral traits, personal characteristics, teaching style, and the teaching level of the instructor. By providing detailed information about each student and their corresponding teacher to AI agents, these agents are able to analyze the student's behavior, learning habits, and the teacher's instructional style, which collectively determine the student's performance. Furthermore, the AI agents offer a detailed analysis process and provide an explanation for the reasons behind the generated grades. A portion of this explanation is depicted in Fig. 8 .

C Information Fusion Framework CEO-IF

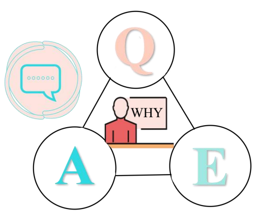
Metaheuristic algorithms are renowned for their powerful global search capabilities and adaptability [2, 15], making them well-suited for solving complex problems [33]. They have been applied in diverse fields, including production control [23, 37, 40, 47], circuit design [8, 51], and drug design [5]. The CEO-IF algorithm introduced in this paper is based on a multi-subgroup co-evolutionary strategy to fuse the predictive results of performance time-sequence and behavior representation models. The offspring population generated by this process is divided into three segments. Through multi-subgroup co-evolutionary search, each subgroup employs a different evolutionary strategy, thereby avoiding the limitations associated with a single search strategy and improving computational efficiency. The specific steps are as follows: the offspring population is divided into superior sub-pupulation, exploration sub-pupulation, and eliminated sub-pupulation based on individual fitness values. The number of superior sub-pupulation advancing to the next iteration is determined by the parameter AP, which is calculated as $AP \cdot N$, where N is the population size after iteration.

If you are a student, your condition is as follows: Pre-class study habits are diligent, in-class study habits are engaged, post-class study habits are reflective, conscientiousness in assignments is meticulous, interaction with the teacher is courteous, and personality is analytical. Your teacher is Ms. Smith. Her interactions with students are supportive and approachable, her teaching habits are innovative and dynamic, her teaching level is advanced, and her classroom atmosphere is engaging and collaborative. Please, according to the above data, generate a table of 25 students who may do the following: propose specific solutions, talk about the givens and constraints of a specific task, build on others' ideas to improve solutions, confirm understanding by asking questions or paraphrasing, repair misunderstandings, not respond when spoken to by others, make fun of, criticize, or be rude to others, provide reasons to support or refute a potential solution, talk about results, bring up giving up the challenge, not visibly focus on tasks and assigned roles, initiate off-topic conversation, join off-topic conversation, ask if others have suggestions, ask to take action before anyone on the team, and compliment or encourage others.

Sure, here is the behavior occurrences for the students:

Behavior occurrences	Student1	Student3	Behavior occurrences	Student1	Student3
Proposes specific solutions	4	2	Talks about results	4	2
Talks about givens and constraints of a specific task	4	1	Brings up giving up the challenge	1	5
Builds on others' ideas to improve solutions	4	2	Not visibly focused on tasks and assigned roles	2	4
Confirms understanding by asking questions/paraphrasing	4	1	Initiates off-topic conversation	3	5
Repairs misunderstandings	3	1	Join soff-topic conversation	3	5
Does not respond when spoken to by others	1	5	Asks if others have suggestions	3	1
Makes fun of, criticizes,or is rude to others	1	5	Asks to take action before anyone on the team	4	1
Provides reasons to support/refute a potential solution	4	1	Compliments or encourages others	4	1

Figure 7: Student behavior information based on the CPS coding framework. AI agents are used to simulate students, generating student behavior information for each individual based on the CPS coding framework and their personal characteristics.The case in the figure only lists some key dimensions of behavioral information.



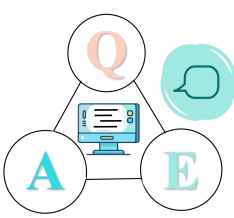
Q: Can you explain why the score is this?

A: Certainly! The student's high score can be attributed to a combination of strong study habits, effective learning style, and a supportive classroom environment. Here's a breakdown of the contributing factors:

1. Study Habits:
- Pre-class: The student is diligent in preparation, leading to better understanding.
- In-class: A serene attitude helps maintain focus and engagement during lessons.
- Post-class: Reflective habits reinforce learning and retention of material.

2. Conscientiousness:
The student is thorough in completing assignments, which contributes to higher quality work and grades.
3. Learning Style:
- Strong collaborative skills are evident, as the student excels in proposing solutions and confirming understanding with peers. This active engagement promotes deeper comprehension.
- While there are areas of lower engagement (e.g., not responding or initiating off-topic conversations), the overall strengths balance these out.

Overall, these elements work together, allowing the student to perform consistently well across multiple tests.



Q: Can you explain why the score is this?

A: Certainly! Here are the test score generated based on the provided student's personality and the teacher's attributes: The student's grades can be explained by several factors:

1. Study Habits:
- Pre-class: Lack of preparation likely affects performance.
- In-class: Calmness helps focus, but doesn't compensate for other issues.
- Post-class: Lackadaisical may not suffice without solid preparation.

2. Conscientiousness and Interaction:
- Approach: Casual attitude towards assignments leads to lower quality work.
- Interaction with Teacher: Carelessness in communication hinders support and clarification.

3. Learning Style:
- Low scores in engagement (e.g., proposing solutions, confirming understanding) show minimal participation.
- High scores in negative behaviors (e.g., not responding, making fun) indicate a lack of focus and collaboration.

Overall, his grades reflect a combination of inadequate preparation, disengagement, and a lack of constructive interaction, which prevent him from achieving higher scores.

Figure 8: Explanation of grade generation. Students and teachers engage in role-playing as agents based on the CPIPIP framework and the CTIF framework, respectively. At the same time, student agents generate student behavior data based on the CPS coding framework. The AI agents use these information to generate the students' grade time-sequence data and provide an explanation for the grade generation process.

Algorithms	Parameters
DT	$DTree_{MaxL}$ = 50, $DTree_{Max}$ = 10
LSTM	TE = 200, IL_{rate} = 0.001, BS = 32
CEO-IF	AP = 0.2, OD = 0

Table 2: Parameter settings.

A larger value of AP results in more superior sub-pupulation advancing to the next iteration, thereby accelerating the process of iteration. The number of exploration sub-pupulation is calculated as $(1-AP)\cdot N$. After selecting individuals with higher fitness values for the differentiation transfer strategy, they proceed to the next iteration, while the eliminated subgroup is discarded and cannot advance further.

D The Details of the Experimental Setting

The dataset used in this study is the DPOTS dataset, developed specifically and publicly available on the project's homepage. It includes both student behavioral data and performance time-sequence data [24]. All experiments are conducted on a device equipped with a dual-core Intel i7-7600U@2.80GHz processor. The hardware environment comprised 16GB of RAM and an Intel HD Graphics 620 GPU, while the software platform utilized for algorithm implementation is Matlab 2020b. Additionally, some of the visualizations in this paper are generated using tools available on chiplot.online.

To validate the effectiveness of the proposed information fusion method, we used a Decision Tree model [46] to learn the correlation between behavioral representation data and learning outcomes. This method, a classic approach in XAI [42], follows the parameter settings as described in [44]. Additionally, an LSTM model [6] is used to learn the correlation between performance time-sequence and learning outcomes, with parameters based on the settings from [18, 31]. Finally, we employed the proposed CEO-IF information fusion framework concurrently, and provided the parameter combinations used in the methods. These settings were established based on the literature [25, 36]. These parameter combinations were applied and validated in the contexts of learning path identification [36] and control system management [25], demonstrating the effectiveness of these configurations.

Specific hyperparameters are detailed in Table 2. In the DT model, '$DTree_{MaxL}$' set to the Maximum Number of Leaf Nodes, and '$DTree_{Max}$' set to the Maximum Depth of the Tree. In the LSTM model, 'TE' denotes the number of Training Epochs, 'IL_{rate}' is the initial learning rate, and 'BS' refers to the batch size.

The LSTM layer consists of 50 hidden units, and a batch size of 32 is selected to ensure smoother training. An initial learning rate of 0.001 is used, which is a widely accepted default and yields satisfactory results in most training scenarios. The CEO-IF framework is implemented on version 4.6 of the PlatEMO platform, where 'AP' represents ambient pressure, ranging from [0,1]. PlatEMO is a platform for the research and application of evolutionary optimization algorithms.

Received 01 23 2025; revised 03 03 2025; accepted 04 26 2025