Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.28970

AI-Driven Personalized Learning Effect Prediction Based on Machine Learning Algorithms

Guangze Wang

Faculty of Business, Economics & Law, University of Queensland, Brisbane, Australia
williamwang479@gmail.com

Abstract. Generative and learning analytics artificial intelligence (AI) is rapidly entering middle school classrooms, expected to enhance personalized learning, improve formative assessment, and reduce the burden on teachers. Australia has formed a relatively clear ecosystem at the policy and curriculum levels: The federal education department has released the "Generative AI Framework for Australian Schools", putting forward principles such as safety, ethics, transparency and capacity building; In the V9 version of the ACARA course, resources for the connection between AI and disciplinary capabilities are provided. NAPLAN has fully shifted to online adaptive evaluation, providing the institutional and technical foundation for personalized assessment. Meanwhile, UNESCO advocates "people-oriented" generative AI education governance, while the OECD emphasizes the value of learning analytics and data-driven adoption. To address the limitations of existing algorithms, this paper proposes a regression prediction algorithm based on multi-head attention mechanism to optimize the bidirectional long short-term memory network (BiLSTM). The study first conducted a correlation analysis, and at the same time selected Decision tree, Random Forest, AdaBoost, CatBoost and LightGBM as comparison models to carry out experiments. From the perspective of performance indicators, the proposed model performs the best in all dimensions and comprehensively outperforms the comparison models. Among the error-related indicators, its MSE (31.688) is approximately 21.7% and 22.0% lower than LightGBM (40.491) and AdaBoost (40.604), respectively, providing better control over large errors. RMSE (5.629) and MAE (4.616) were also the smallest, decreasing by 11.5%-11.7% and 8.1%-11.6% respectively compared with the two models, and the overall deviation was smaller. The MAPE (6.436) is 8.0%-12.5% lower than that of the two models, with small relative errors and stable accuracy. In terms of goodness of fit, its R^2 (0.765) is 6.7%-11.8% higher than that of the two models, which can explain 76.5% of the variation of the dependent variable and has a stronger ability to capture data patterns. This model provides reliable technical support for AI education platforms to accurately analyze students' learning status and optimize personalized teaching strategies, which has significant practical significance for promoting the efficient implementation of personalized education.

Keywords: AI education platform, multi-head attention mechanism, bidirectional long short-term memory network.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.28970

1. Introduction

To clearly present the application value of AI in the field of education from usability to usefulness, we first integrate the core content around the student learning level: AI can dynamically recommend resources based on students' homework results, assessment curves, and platform behavior data, achieving differentiated teaching, which is in line with NAPLAN's adaptive evaluation concept. In English and the humanities, generative AI can provide language and structural feedback, and teaching design should follow a closed loop of student initial draft - AI review - student revision - teacher review to avoid the problems of gwriting and weakened thinking. In STEM subjects, AI can provide thought prompts and mark the causes of mistakes during the math problem-solving process, and at the same time generate personalized consolidation questions, emphasizing process guidance rather than directly providing answers.

Under the dual impetus of the deepening of the "Double Reduction" policy and the digital transformation of education, junior high school mathematics education is facing a crucial transformation from large-scale teaching to personalized and precise teaching [1]. The junior high school stage, as a core period for the construction of mathematical thinking, has problems such as the differentiation of basic levels, fragmented learning behaviors, and uneven knowledge absorption efficiency among students. The traditional teaching model that relies on post-class tests for lagging assessment is difficult to identify weak links in learning in advance, resulting in a lack of precise basis for personalized intervention [2]. With the popularization of AI education platforms, although they can collect multi-dimensional data such as students' pre-class diagnosis, interactive answering, and resource preferences, how to mine key features and predict learning outcomes from massive heterogeneous data has become the core demand to break through the bottleneck of personalized teaching - it is urgent to build efficient prediction models to transform data advantages into teaching decision support This also provides a practical scenario and application value for AI-driven learning effect prediction research [3].

Machine learning algorithms are the core technical support for achieving personalized learning effect prediction in junior high school mathematics. Their core role lies in converting educational data into interpretable prediction results [4]. The prediction accuracy of traditional statistical models is limited because it is difficult to capture the nonlinear relationship between learning duration and the mastery degree of knowledge points. Although tree models can handle multi-feature interactions, they are insufficient in capturing the temporal nature of learning behaviors. The basic time series model can mine the dynamic laws of the learning process, but it lacks targeted weight distribution for key features, resulting in insufficient attention to features with high impact factors. Overall, machine learning algorithms have achieved a transformation from empirical judgment to data-driven prediction models through feature engineering, model training and optimization, providing quantitative tools for personalized learning effect evaluation. However, existing algorithms still have room for improvement in the collaborative optimization of timing and key feature focusing [5].

To address the limitations of existing algorithms, this paper proposes a regression prediction algorithm based on multi-head attention mechanism to optimize the bidirectional long short-term memory network (BiLSTM) [6]. Bidirectional LSTM can simultaneously capture the historical dependencies and future correlations of learning behaviors, effectively mining the bidirectional information of time series data. The multi-head attention mechanism, through parallel computing of multiple sets of attention weights, can specifically enhance the influence of key features and solve the problem of equal attention to important information in traditional BiLSTM [7]. The regression model formed by the integration of the two not only retains the ability of time series modeling but also achieves precise focus on key features, which can more accurately predict the score of students'

DOI: 10.54254/2755-2721/2025.28970

mastery of knowledge points in unit tests. This algorithm provides a better technical path for predicting the personalized learning effect of junior high school mathematics, and can offer more reliable quantitative basis for the dynamic adjustment of learning paths and intelligent resource recommendation of AI education platforms, facilitating the upgrade of personalized teaching from passive response to active predictive intervention.

2. Data sources

This dataset is AI-driven personalized learning effect prediction data for junior high school mathematics, containing a total of 762 samples. The data contains 7 characteristic variables and 1 predictor variable: Among the characteristic variables, grade and pre-class diagnostic score are the basic attributes of students; The daily AI learning duration, the number of interactive questions answered, and the preference for resource types are the AI learning behavior data. The type of home learning equipment and the fixed learning frequency per week are external environmental variables. The predictor variable is the score of mastery of unit test knowledge points (0-100 points, directly measuring the learning effect). It can be used to train regression models, providing model support for the customization of learning paths and intelligent resource recommendation in AI education systems.

Correlation analysis was conducted on each variable, and the correlation heat map is shown in Figure 1. In the correlation analysis of the retention probability of online education users, there are significant differences in the degree of association between each variable and it. Among them, the positive correlation between the number of logins per week and the retention probability is the strongest, with a correlation coefficient of 0.477, indicating that the higher the login frequency of users, the significantly higher the retention probability. The number of assignment submissions, the average session duration, and the number of badges obtained are also moderately positively correlated with the retention probability, with correlation coefficients of 0.408, 0.328, and 0.288 respectively. This reflects that when users actively participate in learning tasks, stay on the platform for a longer time, and receive learning rewards, they are more likely to maintain their willingness to use the platform. However, the difficulty of the course is slightly negatively correlated with the retention probability, with a correlation coefficient of -0.078.

DOI: 10.54254/2755-2721/2025.28970

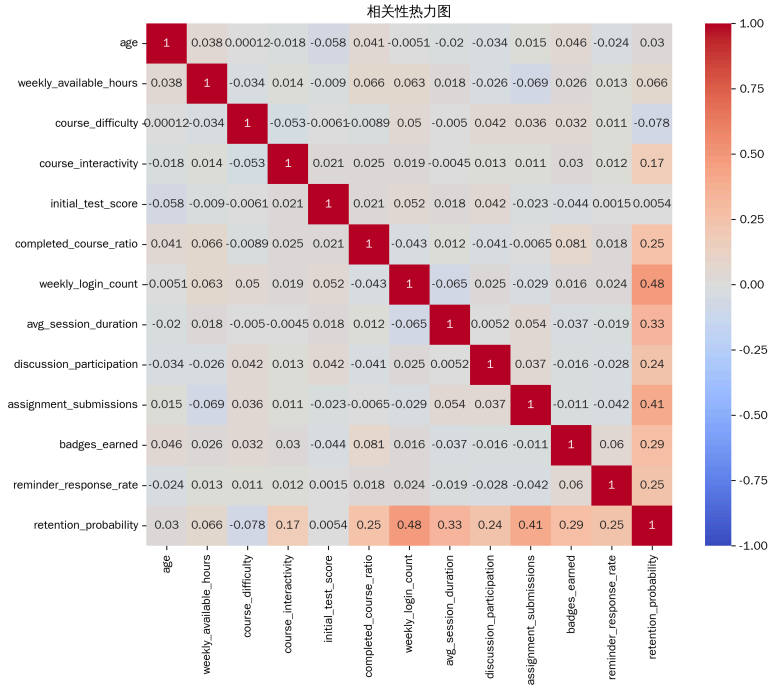


Figure 1. Correlation heat map

3. Method

3.1. Multi-head attention mechanism

The multi-head attention mechanism is a core component of the Transformer model. Its core idea is divide-and-conquer, which captures the correlation information of different dimensions among variables by computing multiple attention heads in parallel. During the calculation, the input query (Q), key (K), and value (V) are first mapped to multiple low-dimensional subspaces respectively through three independent linear projection layers. Then, an attention head is assigned to each subspace, and each head independently performs self-attention calculation - the similarity is calculated through the dot product of Q and K, normalized by Softmax, and then weighted and summed with V to obtain the local attention output corresponding to that head. Each head will focus on different types of dependencies. The schematic diagram of the multi-head attention mechanism is shown in Figure 2.

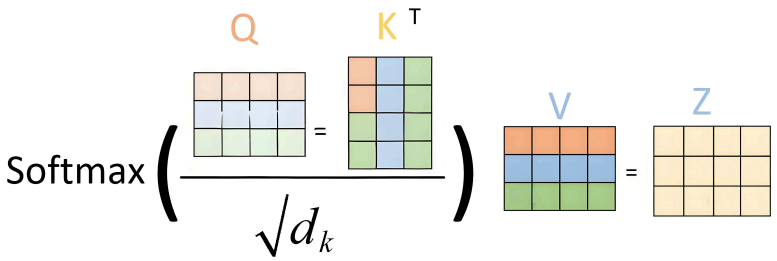


Figure 2. The schematic diagram of the multi-head attention mechanism

The key to multi-head attention lies in integrating the outputs of multiple heads: concatenating the local attention outputs of all heads by dimension, and then mapping the dimensions back to the

DOI: 10.54254/2755-2721/2025.28970

original input dimension through a linear projection layer to obtain the final global attention output. This design not only retains the ability of single-head attention to capture dependencies, but also, through multi-head space parallel computing, enables the model to simultaneously focus on feature associations at different levels, avoiding key dependencies that might be overlooked by single-head attention. Meanwhile, the independent parameter learning of each head enables the model to adaptively adjust the attention weights for different associations, enhancing the representation ability of complex data [8]. In addition, the parallelized computing approach is also compatible with hardware acceleration, providing support for the efficient training and inference of the Transformer.

3.2. BiLSTM

The Bidirectional Long Short-Term Memory Network (BiLSTM) is an improved structure proposed on the basis of the traditional Long Short-Term Memory Network (LSTM) to address the limitation that "unidirectional time series modeling cannot utilize future information". The core idea is to enable the model to simultaneously capture the bidirectional dependencies from the past to the present and from the future to the present in time series data through "bidirectional parallel modeling". Traditional unidirectional LSTM can only transfer information from the starting point to the ending point of the time series and cannot utilize future information at subsequent moments. However, BiLSTM fills this gap by adding a reverse modeling path [9].

Structurally speaking, BiLSTM is composed of two independent LSTM units running in parallel. Both share the input data at the same time, but the direction of information transmission is opposite. Forward LSTM processes data in chronological order. At time t, it receives the current input and the forward hidden state at time t-1, and outputs the forward hidden state containing the accumulated past information. Reverse LSTM processes data in reverse chronological order. At time t, it receives the reverse hidden state of the current input and at time t+1, and outputs the reverse hidden state containing the accumulated future information. At the final output stage of each moment t, BiLSTM concatenates the forward and reverse hidden states to form a comprehensive feature vector that simultaneously contains the bidirectional dependencies of past, present, future, and present [10]. In addition, both LSTM units retain the core gating mechanism of LSTM, which can effectively alleviate the vanishing gradient problem in long sequences, ensure the stable transmission of long-distance information in both directions, and ultimately enable the model to represent the features of time series data more comprehensively and accurately.

3.3. BiLSTM with multi-head attention mechanism optimization

The core of the multi-head attention mechanism optimization bidirectional long short-term memory network (BiLSTM) regression algorithm lies in dynamic weighted temporal features and enhanced global dependency modeling. Traditional BiLSTM captures the context information of a sequence through bidirectional hidden state, but it lacks explicit weights for key time steps or variables in long sequences. The introduction of the multi-head attention mechanism first projects the bidirectional hidden state output by BiLSTM onto multiple subspaces. Each attention head independently calculates the correlation between time steps to generate differentiated weight distributions. After these weights are normalized by Softmax, they are weighted and summed on the hidden states of BiLSTM to form a context vector that integrates local temporal features and global dependencies. Compared with single-attention, multi-head design allows the model to learn multiple dependency patterns in parallel, avoiding information attenuation caused by unidirectional transmission in BiLSTM. Ultimately, the weighted features are input into the regression layer, significantly

DOI: 10.54254/2755-2721/2025.28970

enhancing the sensitivity to nonlinear relationships, long-distance dependencies, and key variables, and strengthening the features that contribute greatly to the prediction target, thereby optimizing the accuracy of continuous value prediction.

3.4. Risk boundary and cost-benefit analysis

In terms of risk boundary management, it is necessary to transform various concerns into controllable key points. Regarding academic integrity, it is essential to clearly define the available and prohibited scenarios of AI. Generative AI should be prohibited in examinations and score-based evaluations, while it is allowed to be marked and used in regular assignments. Privacy and data sovereignty should follow the principle of "data minimization + localization priority" to avoid cross-border storage or improper use of student data. When it comes to content related to specific ethnic groups, it is necessary to follow the relevant cultural sensitivity guidelines to prevent stereotyped or false information. The promotion of AI should also take net burden reduction as a hard target to avoid a rebound in teachers' burden due to unreasonable processes. From the perspective of the CBA, the costs cover platform licensing, hardware equipment, teacher training, and system maintenance and security audits. The benefits are reflected in academic improvement, time savings for teachers, early detection of student risks, and fairness improvement. It is necessary to adopt the CBA+TCO framework to review each academic year and make investment and expansion decisions based on evidence.

This pilot project is a small sample verification lasting 8 to 10 weeks. The subjects selected are English writing or math problem-solving areas for grades 9 to 10, and the experimental group and the control group are divided according to the matching of academic levels. In terms of the intervention process, the writing field adopts the model of "generative writing assistant → student revision → teacher review → reflection log", while the mathematics field provides process prompts and adaptive consolidation questions. At the same time, it clearly stipulates the integrity requirements that AI is prohibited in exams and the use of AI must be marked on homework. The indicator system consists of four major dimensions: academic increment, learning behavior, net reduction of teachers' burden, and compliance. The evaluation uses ANCOVA or the difference method to estimate the effect size, and combines the questionnaires of teachers and students to assess satisfaction and self-efficacy.

4. Result

For the comparative models, this paper conducts comparative experiments using Decision tree, Random Forest, AdaBoost, CatBoost and LightGBM. For the model evaluation parameters, quantitative evaluations are conducted using MSE, RMSE, MAE, MAPE and R^2.

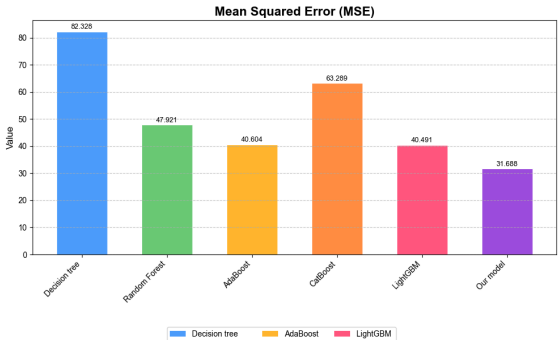
Table 1 presents the comparative experimental results of each model, and Figure 3 shows the comparative bar chart results of each model, as shown in Figure 3. From the perspective of model performance evaluation indicators, Our model demonstrates the best regression prediction performance in all dimensions, comprehensively outperforming decision tree, random forest, AdaBoost, CatBoost and LightGBM models. In terms of error indicators, the MSE of Our model (31.688) is significantly lower than that of other models. Compared with the second-best LightGBM (40.491) and AdaBoost (40.604), it is reduced by approximately 21.7% and 22.0% respectively, indicating that it has better control over the square penalty of prediction bias. Effectively reduced the impact of large errors; RMSE (5.629) and MAE (4.616) were also the smallest, respectively decreasing by approximately 11.5% and 11.6% compared to LightGBM's 6.363 and 5.22, and by

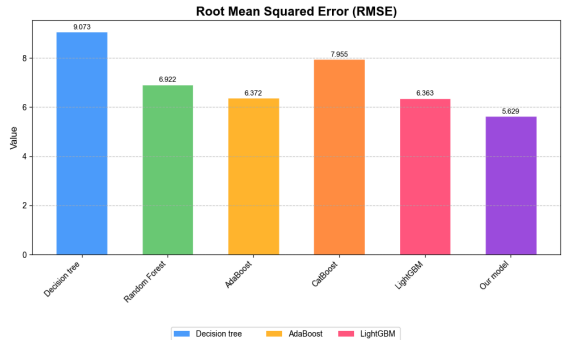
DOI: 10.54254/2755-2721/2025.28970

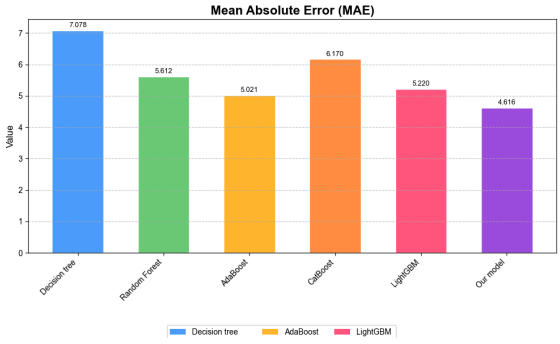
approximately 11.7% and 8.1% compared to AdaBoost's 6.372 and 5.021. It reflects a smaller overall prediction deviation and average absolute deviation degree; MAPE (6.436) was also the lowest, decreasing by approximately 8.0% and 12.5% respectively compared to AdaBoost's 6.993 and LightGBM's 7.352, indicating that its relative error was smaller and its prediction accuracy for data of different magnitudes was more stable. In terms of the goodness of fit index, the R^2 of Our model (0.765) is the highest, which is approximately 6.7% and 11.8% higher than that of LightGBM (0.717) and AdaBoost (0.684), respectively.

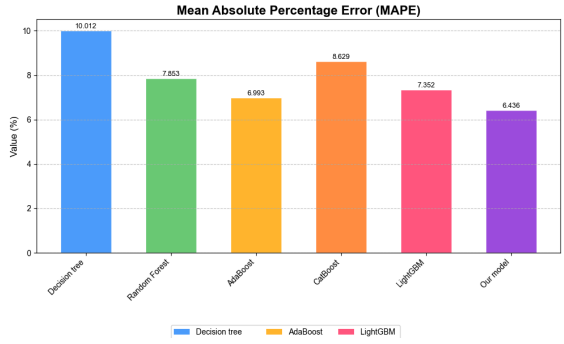
Table 2. The experimental results

Model	MSE	RMSE	MAE	MAPE	R^2
Decision tree	82.328	9.073	7.078	10.012	0.446
Random Forest	47.921	6.922	5.612	7.853	0.653
AdaBoost	40.604	6.372	5.021	6.993	0.684
CatBoost	63.289	7.955	6.17	8.629	0.534
LightGBM	40.491	6.363	5.22	7.352	0.717
Our model	31.688	5.629	4.616	6.436	0.765









DOI: 10.54254/2755-2721/2025.28970

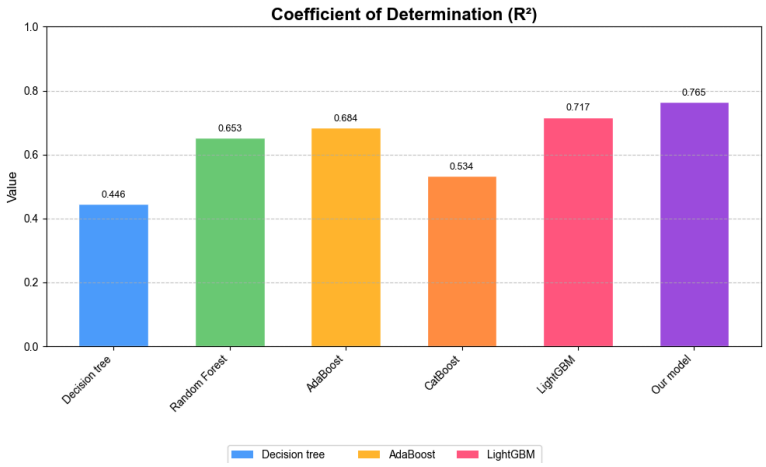


Figure 3. The comparison bar chart of the output experimental results

5. Conclusion

The value of AI in Australian secondary education does not lie in replacing teachers, but in amplifying teachers' professional judgment and supporting differentiated teaching. Through a strategy of small-scale pilot projects, evidence-based evaluation, and gradual expansion, combined with cost-benefit analysis and policy alignment, AI can enhance academic performance and reduce the burden on teachers while effectively managing risks related to academic integrity, privacy, and cultural sensitivity. The data analysis results of existing studies (such as the significant advantages of Attention-BiLSTM in learning behavior prediction) indicate that this direction has practical feasibility and academic value; However, in practice, local data must be used for verification to avoid blind application. The key to the future lies in truly transforming "usable" into "useful".

From the perspective of specific indicators, at the level of error indicators, the mean square error (MSE) of Our model is 31.688, significantly lower than that of other models. Compared with the subperforming LightGBM (40.491) and AdaBoost (40.604), it is reduced by approximately 21.7% and 22.0% respectively. This indicates that it has better control over the square penalty of prediction deviation and can effectively reduce the impact of large errors on the results. The root mean square error (RMSE) of 5.629 and the mean absolute error (MAE) of 4.616 are also the smallest. They are approximately 11.5% and 11.6% lower than LightGBM's 6.363 and 5.22 respectively, and approximately 11.7% and 8.1% lower than AdaBoost's 6.372 and 5.021 respectively. It reflects a smaller overall prediction deviation and average absolute deviation degree; The mean absolute percentage error (MAPE) of 6.436 is also the lowest, which is approximately 8.0% and 12.5% lower than AdaBoost's 6.993 and LightGBM's 7.352 respectively. This indicates that its relative error is smaller and its prediction accuracy for data of different scales is more stable. In terms of the goodness-of-fit index, the coefficient of determination (R^2) of Our model reached 0.765, which was the highest value. Compared with 0.717 of LightGBM and 0.684 of AdaBoost, it increased by approximately 6.7% and 11.8% respectively, indicating that this algorithm can explain 76.5% of the variation of the dependent variable.

References

[1] Sumathy, V., and G. Navamani. "AI-Driven Personalized Learning: Enhancing Student Success through Adaptive Technologies." Library of Progress-Library Science, Information Technology & Computer 44.3 (2024).

DOI: 10.54254/2755-2721/2025.28970

[2] Yekollu, Roop Kumar, et al. "AI-driven personalized learning paths: Enhancing education through adaptive systems." International Conference on Smart data intelligence. Singapore: Springer Nature Singapore, 2024.
[3] Tariq, Muhammad Usman. "AI-Driven Personalized Learning-Enhancing Student Engagement and Outcomes." AI and Learning Analytics in Distance Learning. IGI Global Scientific Publishing, 2025. 151-178.
[4] Yu, Haixia, et al. "Artificial intelligence-based quality management and detection system for personalized learning." Journal of Interconnection Networks 22.Supp02 (2022): 2143004.
[5] Maghsudi, Setareh, et al. "Personalized education in the artificial intelligence era: what to expect next." IEEE Signal Processing Magazine 38.3 (2021): 37-50.
[6] Castro, Gina Paola Barrera, et al. "Harnessing AI for Education 4.0: Drivers of Personalized Learning." Electronic Journal of e-Learning 22.5 (2024): 1-14.
[7] Tavakoli, Mohammadreza. "Hybrid human-AI driven open personalized education." (2023).
[8] Vashishth, Tarun Kumar, et al. "AI-driven learning analytics for personalized feedback and assessment in higher education." Using traditional design methods to enhance AI-driven decision making. IGI Global Scientific Publishing, 2024. 206-230.
[9] Lannelongue, Loïc, and Michael Inouye. "Pitfalls of machine learning models for protein–protein interaction networks." Bioinformatics 40.2 (2024): btae012.
[10] Gm, Dhananjaya, et al. "A digital recommendation system for personalized learning to enhance online education: A review." IEEE Access 12 (2024): 34019-34041.