Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 15, No. 2, 2026

A Time-Frequency Enhanced Attention Network for Student Performance Prediction in Online Learning

Zhiyang Li

School of Computer Science and Technology, Taiyuan Normal University, Jinzhong Shanxi, 030619, China

Abstract: Accurately predicting student performance remains challenging due to evolving trends, periodic fluctuations, and the inherent sparsity of behavioral sequences in online learning platforms. Existing approaches primarily focus on single-modality modeling and lack collaborative integration of temporal dependencies and frequency-domain characteristics, limiting their ability to capture correlations across critical time intervals. To address this limitation, we propose a Time-Frequency Enhanced Attention Network (TF-EAN), which jointly models the temporal dependency structure and spectral patterns of student learning behavior sequences. In the temporal branch, multi-head self-attention with sinusoidal positional encoding is employed to capture long-range dependencies. In the frequency branch, Fast Fourier Transform (FFT) is applied to extract amplitude and phase components for modeling periodic patterns. Experiments conducted on the Open University Learning Analytics Dataset (OULAD) demonstrate that TF-EAN outperforms traditional machine learning methods and several deep learning baselines, achieving an accuracy of 98.74% and 99.15% on two public datasets, respectively, thereby validating the effectiveness of time-frequency collaborative modeling.

Keywords: Online Student Performance Prediction; Time-frequency Modeling; Self-attention Mechanism; Fast Fourier Transform; Learning Behavior Sequence Analysis.

1. Introduction

The digital transformation of education has resulted in online learning platforms amassing substantial quantities of educational data. By mining and analyzing time-stamped learning behavior sequences and accurately forecasting students' final performance, it is feasible to identify learners at risk of attrition and deliver timely and effective instructional interventions, thereby enhancing educational quality and advancing the digital transformation of education.

The advancement of deep learning has yielded efficient techniques for predicting student performance, resulting in various time-series forecasting models. Researchers have utilized multidimensional time-series data to investigate the interrelationships among many characteristics that affect student performance (Shou et al. 2024). Comparisons between artificial neural networks and conventional time-series forecasting techniques underscore the significance of implementing artificial intelligence in educational settings (Etaat et al. 2025). Knowledge distillation (KD) has been utilized in recurrent neural network (RNN) models to enhance the early prediction of student performance (Ryusuke et al. 2023). Nonetheless, initial methodologies were constrained by inadequate parameters and low representational ability. The emergence of LSTM-based architectures has enabled effective modeling of temporal sequence characteristics. Optimization strategies have been implemented in CNN-LSTM architectures to enhance model parameters (Zhang et al. 2025). Bidirectional LSTM models have been employed for long-sequence time-series prediction, elucidating the correlations between students' knowledge levels and learning success (Gao et al. 2025). Ensembles integrating various machine learning methods, including support vector machines (SVM), naïve Bayes (NB), and decision trees (DT) with recurrent neural networks (RNN) and long short-term memory (LSTM) have been suggested to improve prediction performance (Kukkar et al. 2024).

Nonetheless, these methods encounter difficulties in modeling long-duration time-series characteristics and frequently concentrate on single-modality data, hence restricting their capacity to thoroughly capture intricate multi-scale structures and attain collaborative representation across several modalities. The advancement of frequency-domain prediction techniques has provided a novel perspective on modeling temporal periodicity and long-term structures. For instance, fractional Fourier transform convolutional neural networks have been employed to extract periodic time-series features from various angles by modifying rotations (Chen et al. 2025). A two-stage decomposition technique that integrates multivariate variational mode decomposition with discrete Fourier transform has been presented to effectively extract principal features and improve predictive performance (Wang et al. 2025). Despite encouraging outcomes across multiple domains, the majority of methodologies have failed to integrate temporal elements to improve global representation.

This paper presents the Time-Frequency Enhanced Attention Network (TF-EAN) to effectively leverage trends and periodic variations in long-term learning sequences for precise performance prediction. The contributions of this study are delineated as follows:

(1) TF-EAN incorporates Fast Fourier Transform (FFT) into the modeling of student behavior sequences, extracting amplitude and phase characteristics in the frequency domain to delineate periodic fluctuation patterns and integrating them with temporal variables. This time-frequency collaboration mechanism effectively addresses the limitations of single-domain temporal models in capturing periodic structures and implicit rhythms.

(2) In the temporal branch, multi-head self-attention utilizing sinusoidal positional encoding is used, facilitating the explicit modeling of the temporal

sequence of learning behaviors and enhancing the capture of long-range dependencies across weeks.

(3) TF-EAN markedly surpasses conventional machine learning techniques and prevalent deep learning benchmarks on two public test datasets, exhibiting robust generalization capabilities.

2. Methodology

2.1 Problem Definition

This study focuses on the weekly behavioral logs of students on an online platform, continuously recording the sequence of learning behaviors as a matrix $X \in R^{T\times F}$, where T represents the time steps and F denotes the feature dimensions. The label $y \in \{0,1\}$ indicates the final academic performance, with *y=0* corresponding to a Pass and *y=1* to Withdrawn. The objective of this research is to learn a mapping function f that integrates temporal and frequency features from the time-series data to achieve binary classification of students' final learning outcomes.

2.2 Proposed Model

The overall architecture of the proposed Time-Frequency Enhanced Attention Network (TF-EAN) is illustrated in Figure 1. The model consists of three core modules: the temporal branch, which captures sequential dependencies; the frequency branch, which extracts frequency patterns; and the feature fusion and prediction layer, which integrates features from both branches and outputs the classification results.

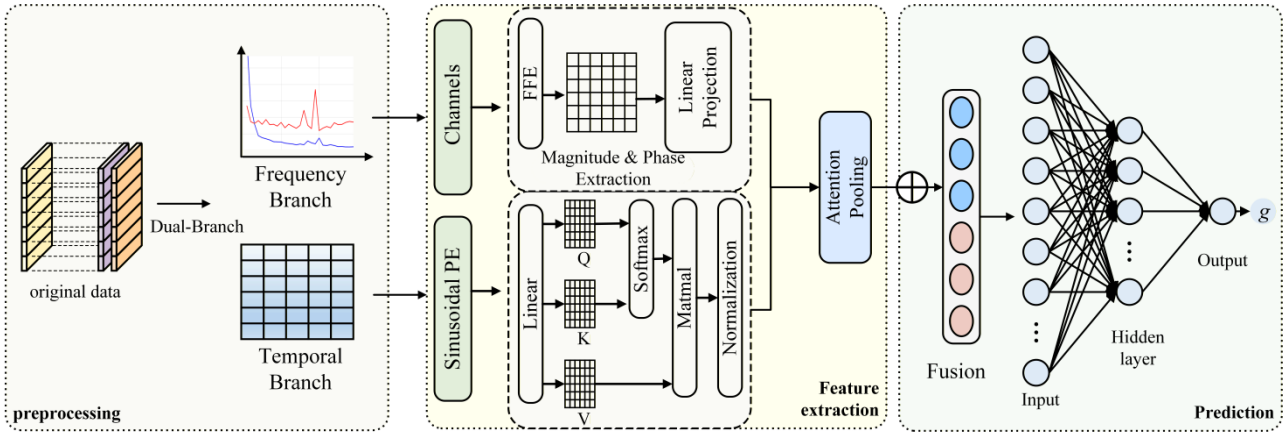


Fig 1. Architecture of the Time-Frequency Enhanced Attention Network

2.2.1 Temporal Branch

The core objective of the temporal branch is to capture dependencies in students' learning behaviors across weekly time intervals. It is composed of three key components: positional encoding, multi-head self-attention, and attention pooling.

(1) positional encoding

The self-attention mechanism lacks intrinsic temporal awareness. Consequently, positional encoding is employed to incorporate positional information into each time step. The computation of positional encoding is conducted as follows:

$$PE(pos,2i) = sin(\frac{pos}{10000^{2i/d_{model}}}) \quad (1)$$

$$PE(pos,2i+1) = cos(\frac{pos}{10000^{2i/d_{model}}}) \quad (2)$$

Here, *pos* denotes the time step, and d_{model} represents the input feature dimension. The resulting positional encoding is added element-wise to the input sequential data, producing $X_{pos} = X + PE$, the input enriched with positional information.

(2) multi-head self-attention

Multi-head self-attention projects the input into multiple subspaces and computes attention in parallel, enhancing the model's ability to capture temporal dependencies at different scales.The input X_{pos} is projected into queries (Q), keys (K), and values (V) through a shared linear layer $W_{qkv} \in R^{d_{in}\times 3d}$, where d_{in} denotes the input dimension and d represents the projection dimension. Then, Q, K, and V are split into *num_heads* separate heads, with each head having a dimension of $d_{head} = d/num_heads$.For each head, scaled dot-product attention is computed to capture dependencies while mitigating the vanishing gradient problem. Here, $Q_h, K_h, V_h \in R^{num_heads\times pos\times d_{head}}$ represents the query matrix for the head.

$$Attention(Q_h,K_h,V_h)=softmax(\frac{Q_h K_h^T}{\sqrt{d_{head}}})V_h \quad (3)$$

Finally, the outputs of all attention heads are concatenated and passed through a linear layer $W_o \in R^{d\times d}$, producing the final temporal feature representation X_{att}.

$$X_{att} = Concat(X_{att,1}, X_{att,2}, \cdots, X_{att,num_heads})W_o \quad (4)$$

(3) attention pooling

To compress variable-length temporal features into a fixed-dimensional temporal feature vector, attention pooling is applied to compute a weighted sum over the time steps:

$$\alpha_t = softmax(\frac{X_{att,t}\cdot q}{\sqrt{d}}) \quad (5)$$

$$X_{time} = \sum_{t=1}^{pos} \alpha_t \cdot X_{att,t} \quad (6)$$

Here, q denotes a learnable query vector, α represents the attention weight for the *t-th* time step, and X_{time} is the final output feature of the temporal branch.

2.2.2 Frequency Branch

The frequency branch transforms sequential learning behaviors into the frequency domain via the Fast Fourier Transform (FFT) to capture the periodicity, overall trends,

and stable patterns in students' learning behaviors.

(1) Real-valued Fourier Transform

The input sequence $X \in R^{T \times F}$ is transformed along the temporal dimension using a real-valued FFT, yielding a complex-valued frequency representation $F \in C^{T^i \times F}$, where $T^i = \lfloor T/2 \rfloor + 1$.

(2) Amplitude and Phase Features

The complex-valued frequency representation is decomposed into amplitude and phase components, which are then concatenated to form the frequency-domain feature vector:

$$F_{feat}=Concat(|F|,\angle F) \quad (7)$$

Here, $|F|$ represents the amplitude, reflecting the magnitude of each frequency component, and $\angle F$ represents the phase, indicating the temporal offset of each frequency component. The concatenated feature vector therefore has a dimensionality of $2F$.

(3) Frequency Feature Projection and pooling

The $2F$-dimensional frequency features are projected into a d-dimensional latent space through a combination of linear mapping, layer normalization, activation, and dropout, yielding the frequency-domain hidden representation.

$$F_{proj} = Dropout(ReLU(LayerNorm(F_{feat}W_f + b_f))) \quad (8)$$

Finally, consistent with the temporal branch, attention pooling is applied to compress the sequence into a fixed-dimensional vector, producing the frequency-domain feature X_{freq}.

$$X_{freq}=AttentionPooling(F_{proj}) \quad (9)$$

2.2.3 Feature Fusion and Prediction

To jointly leverage temporal and frequency-domain features, X_{time} and X_{freq} are directly concatenated to form the combined feature vector g, which serves as the final input to the classification layer. The fused feature g is fed into a fully connected network to produce the final classification output. This can be expressed as:

$$g=[X_{time}:X_{freq}] \quad (10)$$

3. Experiments

3.1 Dataset

This study uses the Open University Learning Analytics Dataset (OULAD) from the UK Open University, selecting data from the FFF course for the 2013B (first semester) and 2013J (second semester) academic terms. The dataset contains four possible outcomes—distinction, pass, fail, and withdrawn—which are grouped into two classes for binary prediction: pass (students who passed the exam, including distinction and pass) and withdrawn (students who dropped out). The students' learning behavior data are organized on a weekly basis for the experiments. The total number of students and the frequency of learning behaviors for the two datasets are summarized in Table 1.

Table 1. Statistics on Learning Behavior Clicks

Dataset	Student	Pass-clicks	withdrawn-clicks
2013B	1193	3296931	269178
2013J	1770	4174919	384769

3.2 Experimental Settings

All experiments were conducted on an AMD Ryzen 9 9740H mobile platform equipped with 16 GB DDR5 memory and an Radeon 780M integrated GPU. GPU acceleration was not used during model training or inference. The proposed TF-EAN (Time-Frequency Enhanced Attention Network) adopts an embedding dimension of $d=64$ with $h = 4$ attention heads. In the temporal branch, sinusoidal positional encoding is introduced, with the positional encoding dimension aligned with the input feature dimension (20). The frequency branch employs a full Fourier frequency encoder (Full FFT), where real-valued rFFT is applied to perform time-frequency transformation and extract amplitude and phase features.For feature fusion, temporal and frequency features (both 64-dimensional) are concatenated via a concatenation operation. The fused representation is subsequently transformed through fully connected layers with dimensions $128 \rightarrow 128 \rightarrow 2$, with GELU activation functions applied between layers to introduce nonlinearity. To mitigate overfitting, the dropout rate is set to 0.1 for both the temporal and frequency branches, and 0.3 for the fusion layer.The model is optimized using the AdamW optimizer with a learning rate of 1×10^{-3} and a weight decay coefficient of 1×10^{-4}. The model is trained for 50 epochs.

All experiments adopt an 8:2 train-test split. The training set is used for model training and parameter optimization, while the test set remains strictly isolated throughout model development and is only used for final performance evaluation.

3.3 Evaluation Metrics

This study employs Accuracy, Recall, and F1-score as evaluation metrics to comprehensively assess model performance.

Accuracy represents the proportion of correctly predicted samples among all samples, reflecting the overall predictive performance of the model.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (11)$$

Recall measures the proportion of correctly identified dropout students among all actual dropout students, indicating the model's ability to detect at-risk learners.

$$Recall = TP/(TP + FN) \quad (12)$$

The F1-score, defined as the harmonic mean of precision and recall, provides a balanced evaluation of the model's overall dropout prediction capability.

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall} \quad (13)$$

Here, TP (True Positives) denotes the number of dropout students correctly identified as dropout, FN (False Negatives) denotes the number of dropout students incorrectly classified as passing, FP (False Positives) denotes the number of passing students incorrectly classified as dropout, and TN (True Negatives) denotes the number of passing students correctly identified as passing.

3.4 Results and Analysis

3.4.1 Comparative Experiments

The proposed TF-EAN model is compared with traditional machine learning models, including Random Forest (RF) and Logistic Regression (LR) (Gul et al. 2025; Mohammad et al. 2023), as well as classical time-series feature extraction models, namely Bi-LSTM and the Time Transformer(Yan et al. 2025). All models were evaluated on both the 2013B and 2013J datasets to validate the effectiveness of the proposed approach. The experimental results are presented in Table 2.

Table 2. Model comparison experiment results

Dataset	Model	Accuracy	Recall	F1-score
2013B	TF-EAN	**98.74**	97.18	**97.87**
	Random Forest	98.33	94.37	97.10
	LR	97.91	97.18	96.50
	Bi-LSTM	98.74	97.18	97.87
	Time Transformer	98.33	**98.59**	97.22
2013J	TF-EAN	**99.15**	**97.90**	**98.94**
	Random Forest	98.02	95.10	97.49
	LR	98.31	96.50	97.87
	Bi-LSTM	99.15	97.90	98.94
	Time Transformer	98.59	97.20	98.23

Note: Bold text indicates the best experimental results.

Overall, the proposed TF-EAN model achieves the best overall performance on both datasets. On the 2013B dataset, TF-EAN attains an Accuracy of 98.74% and an F1-score of 97.87%, while on the 2013J dataset, the performance further improves to 99.15% and 98.94%, respectively. Its performance is comparable to that of Bi-LSTM and significantly outperforms the other baseline models.

Compared with traditional machine learning models such as Random Forest and Logistic Regression, the superior performance of TF-EAN demonstrates the effectiveness of the time-frequency fusion architecture in accurately capturing temporal patterns in learning behavior sequences. In comparison with the Time Transformer, TF-EAN benefits from the complementary periodic and trend features extracted by the frequency branch, compensating for the limitations of purely temporal modeling and achieving improved overall performance. The performance comparable to Bi-LSTM further confirms the effectiveness of the proposed temporal modeling mechanism, while the time-frequency fusion framework provides enhanced interpretability.

Moreover, the stable performance of TF-EAN across the two datasets verifies its strong generalization capability, indicating its potential to provide reliable support for practical dropout early-warning systems. These results collectively demonstrate the rationality and superiority of the proposed time-frequency fusion design.

3.4.2 Positional Encoding Experiments

To evaluate the impact of positional encoding on the model's ability to extract temporal features, ablation experiments were conducted on the 2013B dataset. Specifically, three variants were compared: sinusoidal positional encoding (Sinusoidal PE), learnable positional encoding (Learned PE), and without positional encoding (w/o PE). The experimental results are presented in Table 3.

Table 3. Positional encoding experimental results

Encoding	Accuracy	Recall	F1-score
Sinusoidal PE	98.74	97.18	97.87
Learned PE	97.49(↓1.25)	95.77(↓1.41)	95.77(↓2.10)
w/o PE	97.49(↓1.25)	94.37(↓2.81)	95.71(↓2.16)

The experimental results indicate that the model employing sinusoidal positional encoding achieves the best performance across all evaluation metrics. When sinusoidal encoding is replaced with learnable positional encoding or removed entirely, the model's Accuracy, Recall, and F1-score all exhibit varying degrees of decline. Notably, the largest decrease is observed in Recall when positional encoding is removed, with a drop of 2.81 percentage points, while the F1-score correspondingly decreases by 2.16 percentage points.

These findings demonstrate that positional encoding provides essential temporal positional information for weekly learning behavior sequences and plays a critical role in enabling the model to accurately capture sequential dependencies. The superior performance of sinusoidal positional encoding suggests that its stable and deterministic representation is better suited to the characteristics of the temporal data used in this study. In contrast, the absence of positional encoding prevents the model from effectively perceiving the temporal order of learning behaviors, leading to a noticeable degradation in overall predictive performance.

3.4.3 Ablation Studies

To evaluate the respective contributions of the temporal

branch and the frequency branch, as well as the necessity of their integration within the TF-EAN framework, branch ablation experiments were conducted. Specifically, two variant models were constructed: TF-EAN-T, which retains only the temporal branch, and TF-EAN-F, which retains only the frequency branch. These variants were compared with the complete model (TF-EAN, Full model) on both the 2013B and 2013J datasets. Model performance was evaluated using Accuracy, Recall, and F1-score. The experimental results are presented in Table 4.

Table 4. Two-branch ablation experiment results

Dataset	Model	Accuracy	Recall	F1-score
2013B	TF-EAN (Full model)	98.74	97.18	97.87
	TF-EAN-T(temporal)	98.33(↓0.41)	98.59(↑1.41)	97.22(↓0.65)
	TF-EAN-F(frequency)	94.56(↓4.18)	90.14(↓7.04)	90.78(↓7.09)
2013J	TF-EAN (Full model)	99.15	97.90	98.94
	TF-EAN-T(temporal)	98.59(↓0.56)	97.20(↓0.70)	97.23(↓1.71)
	TF-EAN-F(frequency)	98.87(↓0.28)	97.90	98.59(↓0.35)

The experimental results demonstrate that the full TF-EAN model achieves the best overall performance on both datasets, significantly outperforming the single-branch variants, thereby validating the necessity of integrating the temporal and frequency branches.

On the 2013B dataset, the temporal-only model (TF-EAN-T) shows a decrease of 0.41 percentage points in Accuracy and 0.65 percentage points in F1-score compared with the full model. Although Recall slightly improves, the overall predictive performance declines. In contrast, the frequency-only model (TF-EAN-F) exhibits a substantial performance degradation, with all metrics decreasing by more than 4 percentage points. Notably, Recall and F1-score drop by 7.04 and 7.09 percentage points, respectively. These results indicate that, on this dataset, the frequency branch plays a critical complementary role to the temporal features, while frequency-only modeling is insufficient to effectively capture the sequential patterns of students' learning behaviors.

On the 2013J dataset, the performance decline of the single-branch models is relatively moderate; however, the full model still maintains a clear advantage. Compared with the full model, TF-EAN-T shows decreases of 0.56 percentage points in Accuracy and 1.71 percentage points in F1-score, while TF-EAN-F exhibits slight reductions across most metrics, with Recall remaining comparable to that of the full model. This discrepancy can be attributed to differences in feature distributions between the two datasets, leading to variations in the relative contributions of the temporal and frequency branches. Nevertheless, their integration consistently improves overall performance.

In summary, the temporal and frequency branches complement each other and provide synergistic benefits. By fusing time-domain and frequency-domain features, the full model effectively overcomes the limitations of single-branch modeling. It preserves the temporal branch's ability to capture sequential dependencies while leveraging the frequency branch to extract periodic and trend-related characteristics of learning behaviors. These findings collectively demonstrate the rationality and superiority of the proposed time-frequency fusion architecture.

4. Conclusion

This study proposes a Time-Frequency Enhanced Attention Network (TF-EAN) for student performance prediction in online learning environments. The model establishes a dual-branch collaborative framework that jointly captures temporal dependency structures and spectral characteristics of student learning behavior sequences. In the temporal branch, multi-head self-attention with sinusoidal positional encoding is employed to model long-range dependencies across weeks. In the frequency branch, Fast Fourier Transform (FFT) is applied to transform behavioral sequences into the frequency domain, extracting amplitude and phase components to characterize periodic fluctuations and trend patterns in learning behaviors.

During feature fusion, TF-EAN integrates temporal and frequency representations through feature-level concatenation and nonlinear transformation, enabling frequency-domain information to serve as an enhancement mechanism rather than a replacement for temporal modeling. This design improves the model's ability to capture complex behavioral patterns. Ablation studies further confirm the effectiveness of frequency-domain features in strengthening sequential modeling. Experimental results on the Open University Learning Analytics Dataset (OULAD) demonstrate that TF-EAN achieves stable and superior predictive performance, thereby validating the effectiveness of time-frequency collaborative modeling.

Future work will explore multi-scale frequency-domain modeling strategies, adaptive frequency selection mechanisms, and cross-course transfer learning approaches to enhance the model's adaptability and interpretability in more complex educational scenarios.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Shou Z, Xie M, Mo J, et al. Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach[J]. Applied Sciences,2024,14(6): DOI:10. 3390/ APP14062522.

[2] Etaat F. AI-based assessment of students' writing skill progress: Implementing "Artificial neural network modeling" vs. "Time series prediction"[J]. International Journal of Educational Research,2025,131102575-102575. DOI:10. 1016/J. IJER. 2025. 102575.

[3] Ryusuke M, Fumiya O, Tsubasa M, et al.Recurrent Neural Network-FitNets: Improving Early Prediction of Student Performanceby Time-Series Knowledge Distillation[J].Journal

of Educational Computing Research,2023,61(3):639-670. DOI: 10. 1177/07356331221129765.

[4] Zhang Q Q, Li M, Xu T, et al. A Single-Feature Financial Time Series Forecasting Model Based on CNN-LSTM with SE-Attention Mechanism and Grey Wolf Optimization Algorithm [J]. Computational Economics,2025,(prepublish):1-27. DOI: 10. 1007/S10614-025-11223-5.

[5] Gao A ,Liu Z .Long sequence temporal knowledge tracing for student performance prediction via integrating LSTM and informer.[J].PloS one,2025,20(9): e0330433. DOI:10.1371/ JOURNAL. PONE.0330433.

[6] Kukkar A ,Mohana R ,Sharma A, et al. A novel methodology using RNN + LSTM + ML for predicting student's academic performance[J].Education and Information Technologies, 2024, 29 (11):14365-14401.DOI:10.1007/S10639-023-12394-0.

[7] Chen W, Ye J ,Zhao C , et al.MFFCNN: multi-scale fractional Fourier transform convolutional neural network for multivariate time series forecasting[J]. The Journal of Supercomputing, 2025,81(2):416-416.DOI:10.1007/S11227-024-06888-Y.

[8] Wang M ,Meng Y ,Sun L , et al. Multivariate variational mode decomposition combined with discrete Fourier transform and lightweight Mixture-of-Experts models for predicting multivariate time series with strong volatility[J]. Expert Systems With Applications,2026, 303130731-130731.DOI: 10. 1016/ J.ESWA.2025.130731.

[9] Gul N M ,Abbasi W ,Babar Z M , et al. Data driven decisions in education using a comprehensive machine learning framework for student performance prediction[J].Discover Computing, 2025, 28(1): 153-153.DOI:10.1007/S10791-025-09585-3.

[10] Mohammad S A, Kaltakchi A S T M ,Ani A A J , et al. Comprehensive Evaluations of Student Performance Estimation via Machine Learning[J]. Mathematics, 2023, 11 (14): DOI:10.3390/MATH11143153.

[11] Yan Z ,Wang Y ,Chen T .Time series prediction of college student satisfaction based on BiLSTM in big data scenarios[J].Journal of Computational Methods in Sciences and Engineering,2025,25(3):2396-2410.DOI:10.1177/ 147279 78241313259.