https://doi.org/10.31449/inf.v49i26.9807 Informatica **49** (2025) 385–398 **385**

Hybrid Model of Fuzzy Logic and Recurrent Neural Network for Dynamic Student Achievement Prediction

Dongdong Duan, Suhui Zhang*
Office of academic affairs, Handan University, Han'dan 056005, Hebei, China
E-mail: Dongdong_Duan@outlook.com, zhangsuhui@hdc.edu.cn
*Corresponding author

Keywords: academic achievement prediction, fuzzy logic, recurrent neural network, personalized intervention, teaching path optimization

Received: June 20, 2025

In this study, a dynamic prediction model of student academic achievement was developed by integrating fuzzy logic and recurrent neural networks (RNN). The dataset consisted of 235 undergraduate students enrolled in the course Learning Strategy and Behavior Analysis during the spring semester of 2024, with data covering grades, online learning behaviors, and interaction records. The fuzzy rule system was constructed to transform uncertain behavioral variables, such as "learning engagement" and "task completion stability," into interpretable linguistic categories. These were then combined with an RNN structure to capture temporal dependencies in students' grade trajectories. Model training was conducted with 120 epochs, a batch size of 32, and a learning rate of 0.003. Results demonstrated high predictive accuracy, with mean squared error (MSE) ranging between 0.022 and 0.033 and coefficient of determination (R^2) values above 0.94 across validation samples. Personalized interventions derived from prediction outputs, such as increasing video learning time or enhancing peer discussion, led to measurable improvements, with error reductions of up to 0.039 in specific cases.

Povzetek: Predstavljen je hibridni model, ki združuje mehko logiko in ponavljajočo se nevronsko mrežo za dinamično napoved študijskega uspeha. Na podatkih 235 študentov dosega visoko točnost napovedi ter omogoča razložljive in personalizirane pedagoške intervencije.

1 Introduction

With the rapid development of information technology and artificial intelligence, various intelligent technologies are gradually introduced in the field of education to assist teaching and dynamic prediction of student achievement. The prediction of academic achievement can not only provide personalized learning guidance for students, but also help educational administrators and teachers to formulate more scientific teaching strategies and improve teaching quality and effect. Traditional grade prediction methods often rely on statistical models or linear regression, but these methods often underperform in the face of individual student differences and grade fluctuations.

With the rapid development of artificial intelligence and educational big data, the prediction of student academic achievement has become an important field in educational research. Previous studies have shown that factors such as psychological resilience, teacher support, and student engagement are closely related to academic performance, but most approaches still rely on static analysis and lack the ability to model dynamic and uncertain learning behaviors. This study focuses on exploring how fuzzy logic can be used to represent behavioral uncertainty and how recurrent neural networks can capture the time-dependent evolution of student performance. The purpose is to design a hybrid model that integrates these two methods to provide more accurate and interpretable predictions. Using a dataset of 235 undergraduate students from a university course, the model will be tested for its prediction accuracy and its ability to reflect real patterns of learning behavior. Another goal of this study is to examine how prediction results can be transformed into targeted interventions, offering practical support for personalized teaching strategies and academic risk management.

At present, the research on the influencing factors and prediction paths of students' academic achievement has become increasingly in-depth, and the research perspective has gradually shifted from static results to dynamic ones, trying to promote the intelligent transformation of teaching intervention strategies through the analysis of procedural variables. Liu et al. proposed a hybrid deep neural network that integrates convolutional neural networks, bidirectional gated recurrent units, and an additive attention mechanism to enhance short-term traffic flow prediction, demonstrating superior performance compared with conventional models [1]. Meng and Jia emphasized that psychological resilience not only directly predicts academic achievement, but also reveals the linkage mechanism between internal psychological traits and behavioral pathways through the mediating role of academic participation [2]. Thangeda et al. developed a neural network-based predictive decision model for customer retention in the telecommunications sector, showing that behavioral data-driven prediction can effectively support retention strategies [3]. Bas verified that academic emphasis indirectly affects students' performance through school belonging and academic resilience through structural equation models and proposed that there is a nonlinear response relationship between environment and individuals [4]. Yu et al. analyzed how teacher-student relationship affects adolescents' academic achievement through multiple mediating mechanisms and emphasized the application of control value theory in academic achievement [5]. Luo et al. found that self-efficacy has a significant positive impact on academic achievement and learning engagement plays a mediating role between them [6]. Chen et al. pointed out that school support and online learning participation in the later stage of the epidemic significantly improved students' academic performance [7]. Younger et al. studied the protective effect of growth mindset on the academic performance of middle school students and found that this factor can alleviate academic risks and promote academic development [8]. Zhou et al. explored the relationship between teacher support and student achievement through longitudinal data analysis, and found that individual differences of students have an impact on this relationship [9]. Rezaei et al. studied the academic optimism of agricultural vocational school students and its impact on academic achievement, and proposed that positive attitude in educational environment can promote the improvement of academic achievement [10].

Current research has gradually expanded from static regression analysis to intelligent methods such as machine learning and deep learning. Linear regression, decision tree and support vector machine are dominant in the early research, but these methods are mostly based on static samples and can not reflect the dynamic fluctuations and time series characteristics of the results. In recent years, with the development of educational big data, recurrent neural network structures such as long short-term memory networks and gated loop units show high prediction accuracy in processing sequence learning tasks. At the same time, some studies try to introduce fuzzy logic to deal with fuzziness and uncertainty in student behavior variables, so as to enhance the explanatory power and robustness of the model. Although some achievements have been made by existing technologies, there are still few researches on the effective integration of fuzzy rule system and temporal neural network, especially the lack of model structure suitable for dynamic analysis of individual learning path, resulting in an obvious gap between prediction accuracy and personality adaptation [11].

In addition to educational psychology studies, recent advances in hybrid intelligent systems are highly relevant. Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and neuro-fuzzy models have demonstrated strong potential in capturing uncertainty in temporal data, and several works in educational data mining have explored temporal neuro-fuzzy approaches for performance prediction. More recently, deep learning models integrated with learnable fuzzy layers have been proposed to combine interpretability with predictive strength. The present study extends this line of work by embedding a fuzzy rule system into a GRU-based framework, positioning the contribution within neuro-fuzzy and deep hybrid model research.

Prior studies in educational data mining have applied sequence models such as LSTMs to grade prediction, while neuro-fuzzy approaches like ANFIS integrated fuzzy reasoning with temporal learning. This study extends these methods by embedding fuzzy rules into GRUs, advancing both interpretability and performance.

2 Materials and methods

The detailed description of the experimental process ensures the transparency and reproducibility of the research method, and provides a solid foundation for the subsequent analysis and discussion of the results, as shown in Figure 1 [12].

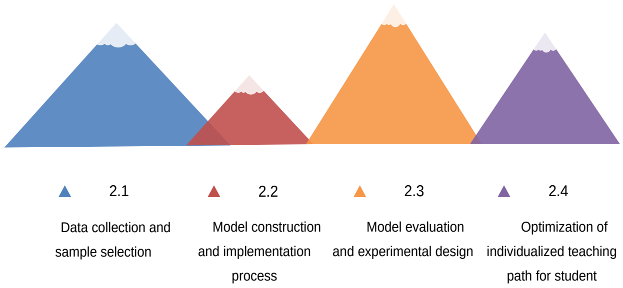

Figure 1: Data processing flow chart

2.1 Data collection and sample selection

2.1.1 Data source

The dataset was collected from the intelligent teaching platform of H University during the spring semester of 2024, covering 235 undergraduate students enrolled in the course Learning Strategy and Behavior Analysis. To ensure reliability, strict preprocessing procedures were applied. Missing values below 15% were retained and filled using multiple imputation with predictive mean matching, while samples with higher missing proportions were excluded. Outliers were identified through the box plot method combined with the three-standard-deviation rule, and unreasonable cases such as abnormal login frequencies were removed. Behavioral variables including video viewing time, assignment submissions, and platform activity were normalized using Z-score standardization to eliminate scale bias. Classification variables, such as course ID and term identifiers, were one-hot encoded for consistency. These steps ensured high data integrity, minimized bias, and provided a robust foundation for subsequent fuzzy logic embedding and dynamic modeling as shown in Table 1.

Table 1: Overview of student academic performance and behavioral data

Student ID	Term ID	Course ID	Continuous Assessment Score	Final Exam Score	Video Viewing Duration (h)	Chapter Completion Rate (%)	Assignment Submissions	Active Platform Days
A031	T03	C101	74.6	81.3	15.2	89.3	19	42
A067	T03	C101	63.4	69.1	11.8	76.8	15	31
A088	T03	C101	58.5	62.7	8.1	52.4	11	23
A109	T03	C101	85.9	90.6	19.3	94.6	21	47
A126	T03	C101	71.2	77.9	14.5	81.7	17	38

The prediction target of the model is composed of the usual grades and the final grades, while the behavioral data includes four core variables: total video viewing time, chapter completion rate, assignment submission frequency and platform active days to reflect students' participation and learning rhythm in the teaching process. The data showed strong individual differences and time continuity, and there was an obvious cooperative trend among behavioral variables, such as the moderate positive correlation between the number of active days on the platform and the chapter completion rate. Although some low-level students have homework submission behavior, the frequency of platform login and video learning time are obviously low, suggesting that there are discontinuity and weak connection problems in the learning path. The characteristics of nonlinearity, fuzziness and hysteresis of behavioral data meet the basic requirements of dynamic modeling and fuzzy logic introduction in this study [14].

All procedures involving human participants were reviewed and approved by the Institutional Review Board of H University (Approval No. EDU2024-041). Informed consent was obtained electronically from all participating students prior to data collection, and data were anonymized by removing identifiers and replacing them with random codes. Beyond privacy protection, ethical risks of deploying predictive systems were also considered. Predictive labeling may inadvertently stigmatize students or create fairness concerns across gender or academic majors. To address this, subgroup analyses were conducted to audit potential bias, and no significant differences in error distributions were found between male and female students or across disciplines. Mitigation strategies such as transparency in reporting and limiting access to predictions were adopted to reduce misuse. These measures ensure compliance with ethical standards and support the responsible use of predictive analytics in education.The use of student data in this study was reviewed and approved by the Institutional Review Board of H University (Approval No. EDU2024-041). All participants provided informed consent prior to data collection, and identifiers were anonymized to ensure confidentiality. In cases where demographic subgroup analysis was performed, data were aggregated to prevent re-identification. These procedures complied with institutional and national ethical guidelines for educational research.

2.1.2 Sample selection criteria

This section specifies the screening criteria for the student sample in this study, and establishes a unified selection logic based on course participation, data integrity, and behavioral feature validity to ensure the representativeness of model training and the reliability of prediction results, as shown in Table 2.

Table 2: Variable structure and definitions for research sample

Variable Name	Variable Type	Data Format	Description
Student_ID	Identifier	String	Anonymous student identifier
Term_ID	Time Variable	String	Academic term of data collection
Course_ID	Identifier	String	Course code for enrolled subject
Continuous_Score	Output Variable	Numeric	Continuous assessment score, full score = 100
Final_Exam_Score	Output Variable	Numeric	Final examination score, full score = 100
Attendance_Rate	Input Variable	Percentage	Average class attendance rate
Video_Viewing_Time	Input Variable	Hours	Total video learning time
Assignment_Submissions	Input Variable	Integer	Number of assignme

			nts submitted
Chapter_Completion_Rate	Input Variable	Percentage	Completion rate of assigned learning chapters
Login_Frequency	Input Variable	Days	Number of days with platform login activity
Discussion_Participation	Input Variable	Count	Frequency of participation in forums or Q&A sessions
Learning_Consistency	Input Variable	Numeric	Weekly fluctuation coefficient of behavioral consistency
Peer_Interaction_Count	Input Variable	Count	Interactions with peers (e.g., replies, collaborative tasks)

To ensure the representativeness and data quality of the selected samples, multiple screening criteria were established in the study. Selected students are required to complete the course in the spring of 2024, with a minimum duration of 16 weeks and a minimum missing rate of 10% for all behavioral data. After screening, a total of 235 students were selected as valid samples, and the data integrity reached 93.7%. The sample covers students majoring in arts and sciences, with a relatively balanced gender distribution, and there are significant differences in the depth of course participation. The median duration of video learning is 13.6 hours, and the distribution of behavioral consistency indicators is skewed, with some students exhibiting high-frequency short-term sprint learning behavior. There is a moderate positive correlation between platform activity and chapter completion rate, indicating that students with relatively coherent learning paths are more likely to maintain stable progress. Interactive variables such as Discussion-Participation and Peer_interaction_Count have high variability, reflecting strong individual differences in social behavior. The variables included in the final sample have multidimensional behavior, result correlation, and time series structure, which can meet the basic requirements of dynamic modeling and fuzzy rule system design in this study [14].

2.1.3 Data preprocessing

Missing data were addressed using Multiple Imputation by Chained Equations (MICE), with five imputations per variable. The missingness mechanism was assessed as MAR based on correlation between missing values and observed behavioral variables. A sensitivity analysis was performed by comparing results across imputed datasets, and prediction performance remained stable, with less than 2% variation in MSE. Outlier detection applied a combined strategy: the box-plot rule identified values beyond 1.5×IQR, while the three-standard-deviation threshold flagged extreme cases. A total of 14 samples (6% of the dataset) were removed, and analysis confirmed that removal did not significantly alter the distribution of performance categories or behavioral patterns. After imputation and outlier handling, features were standardized using Z-score normalization to ensure consistent scale. These steps provided a transparent and replicable preprocessing pipeline.

Student grades were normalized before model training using min–max scaling. Raw scores in the range 0 to 100 were linearly mapped into the interval [0,1]. For example, a raw grade of 75 was converted to 0.75. All error metrics, including MSE and RMSE, are reported on this normalized scale to ensure consistency and interpretability.

To ensure reproducibility, the preprocessing pipeline was formalized as pseudocode with explicit parameters. Missing values below 15% were imputed using multiple imputation with predictive mean matching (k=5 donors), while samples exceeding this threshold were excluded. Outliers were identified using the boxplot method with a 3σ cutoff and removed. Weekly aggregation was applied by aligning behavioral logs into 16 natural weeks per semester, summing video hours and assignments, and averaging chapter completion rates. A Python script is provided to convert raw platform logs into the processed dataset, with sensitive identifiers anonymized [15]. The preprocessing steps can be summarized as:

1.Import raw logs → anonymize IDs.

2.Handle missing data → impute (≤15%) or exclude (>15%).

3.Detect outliers → apply 3σ rule.

4.Normalize features → Z-score scaling.

5.Aggregate by week → build time-series sequences.

This pipeline ensures consistent dataset construction for replication

2.2 Model construction and implementation process

2.2.1 Model selection and theoretical basis

This study chooses to combine fuzzy inference systems with recurrent neural networks to construct a composite dynamic prediction model, while addressing the issues of fuzziness and temporal dependence in educational data. Fuzzy systems are suitable for characterizing non precision and subjective judgment variables in students' learning behavior, such as "level of learning engagement" and "quality of homework completion"; RNN has the ability to process time series data and capture the dynamic evolution trajectory of student performance. Fuzzy logic theory is based on fuzzy sets and fuzzy membership functions to construct rule systems, with the core idea of "modeling uncertainty with linguistic rules". RNN captures temporal dependencies through a state transition mechanism and recursively updates hidden states ht to fit complex nonlinearities in the time dimension [16].

The fuzzy membership function was defined explicitly as a Gaussian function:

$$\mu_A(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \tag{1}$$

where c is the center and σ controls spread. Defuzzification used the weighted average method to normalize rule outputs. Student grades were min–max scaled into [0,1] before training, and all error metrics (MSE, RMSE, MAE) were computed on this normalized scale to ensure interpretability.

The fuzzy rule inference results are synthesized using the weighted average method:

$$y = \frac{\sum_{i=1}^{n} w_i \cdot y_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

wi is the activation of Rule i, and yi is the corresponding output.

The recurrent neural network was implemented using a two-layer gated recurrent unit (GRU), with hidden state updates defined as:

$$h_t = (1-z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{3}$$

where z_t is the update gate and $\tilde{h}_t$ is the candidate activation. Final predictions were computed by concatenating fuzzy rule outputs with continuous inputs before passing them into the GRU. This ensured consistent integration of fuzzy features with temporal learning.

The output prediction is:

$$\hat{y}_t = W_{hy} h_t + b_y \tag{4}$$

The error function is defined using mean square error:

$$L = \frac{1}{T} \sum_{t=1}^{T} (\mathrm{y}_t - \hat{\mathrm{y}}_t)^2 \tag{5}$$

Through the above formula, the model can realize the analysis of the fuzzy characteristics of students' behavior and the effective fitting of the time change of grades, and provide a theoretical basis for the construction of personalized prediction path.

2.2.2 Model structure and network parameter design

The model architecture was implemented using a two-layer gated recurrent unit (GRU) network rather than a vanilla RNN or LSTM, as preliminary experiments showed GRUs achieved better convergence with fewer parameters while mitigating vanishing gradient issues. Each layer consisted of 64 hidden units with tanh activations, and a dropout rate of 0.25 was applied after each hidden layer. The input to the network was a concatenated matrix of normalized behavioral features and fuzzy rule vectors, with the fuzzy vector dimension fixed at three sets per variable ("low," "medium," and "high"). Membership function parameters were initially defined using expert knowledge but were fine-tuned during training through gradient-based optimization. The output layer generated a single normalized grade prediction for each time step, with sequence length fixed at 16 weeks to match the academic term. This architecture ensured efficient training, interpretability, and robust temporal modeling [17].

The recurrent neural network used in this study was a gated recurrent unit (GRU) rather than a vanilla RNN or LSTM. The final architecture consisted of two GRU layers with 64 hidden units each, followed by a dropout layer (0.25) and a fully connected output layer. An architectural diagram was added to illustrate the model, showing the integration of fuzzy features with behavioral inputs and the sequential flow through GRU layers to the prediction output. As shown in Figure 2.

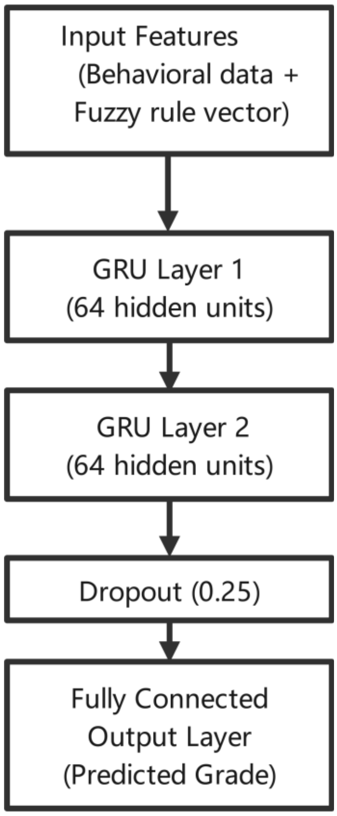

Figure 2: Flowchart of the model

2.2.3 Fuzzy rule system construction method

The fuzzy rule system was designed based on expert knowledge and iterative validation to capture the uncertainty of student learning behaviors. Initially, input variables such as video viewing time, assignment submissions, and platform activity were classified into linguistic categories including “low,” “medium,” and “high.” Rules were then formulated to link these inputs with predicted academic outcomes. To ensure reliability, the rule base underwent iterative adaptation: rules were tested against a validation subset of the dataset, and discrepancies between predicted and observed outcomes were analyzed. Rules with low predictive contribution were refined or replaced, while membership functions were recalibrated using grid search to improve alignment with real data distributions. This iterative validation ensured that the fuzzy inference system remained both interpretable and reproducible, providing robust input features for the recurrent neural network. For example, if a student's video viewing time is "high" and the chapter completion rate is "high," the student's academic performance can be inferred to be "excellent." These rules reflect the relationship between students' learning behavior and achievement, and use fuzzy reasoning methods to generate predictive outputs. The core of fuzzy reasoning is to calculate the membership degree of input variables, and use weighted average method to synthesize the reasoning results of each rule, so as to get the final predicted value of academic performance. The fuzzy rule system can provide effective input for the subsequent recurrent neural network prediction module by simulating the behavior pattern of students in the actual learning process, and enhance the model's adaptability to dynamic changes [18].

The fuzzy rule system consisted of 48 rules constructed through expert elicitation and refined with data-driven adjustments. Three experts in educational psychology and learning analytics participated, each with over 10 years of experience. Consensus was reached using a Delphi-style procedure with two iterative rounds. Membership functions were Gaussian with three sets per variable (low, medium, high). A representative rule was: IF video viewing time is high AND chapter completion is medium THEN predicted performance is good. During inference, rule activations were calculated as the product of input memberships, and decision traces were logged to show which rules contributed most strongly to predictions. For example, in a student with irregular attendance but high peer interaction, rules emphasizing social engagement were activated, aligning with observed academic outcomes. This transparent mapping between behavioral features and outputs demonstrates interpretability and supports practical teaching interventions.

2.2.4 Implementation of model training and prediction process

The training process integrated both fuzzy logic features and time-series grade data into the recurrent neural network. The architecture consisted of two hidden layers with 64 nodes each, using the tanh activation function to enhance nonlinear fitting. A dropout rate of 0.25 was applied to mitigate overfitting, selected after preliminary experiments demonstrated improved generalization compared to lower rates. The learning rate was set at 0.003 with the Adam optimizer, chosen for its stable convergence on educational datasets. Convergence was assessed not only by monitoring loss reduction but also by tracking validation metrics such as mean squared error (MSE) and R^2 over training epochs. Early stopping was employed when validation loss failed to improve for ten consecutive epochs, ensuring efficiency and avoiding overfitting. These design choices were based on balancing computational cost with predictive stability, providing transparency and reproducibility in model training [19].

Model training used the Adam optimizer with an initial learning rate of 0.003. A grid search was conducted to select hyperparameters, exploring hidden units [32,64,128], dropout rates [0.1,0.25,0.5], and batch sizes [16,32,64]. Validation performance on a 30% hold-out set determined the final choice of 64 hidden units, dropout 0.25, and batch size 32. Learning rates followed a step decay schedule, halving every 40 epochs. Weights were initialized with Xavier uniform initialization, and L2 regularization with coefficient$10^{-4}$was applied to mitigate overfitting. Early stopping was triggered if validation loss failed to improve for 10 epochs. All experiments were run with fixed random seeds (42 for NumPy, PyTorch) to ensure reproducibility. The fuzzy membership parameters were initialized from expert priors but fine-tuned during training. Experiments were conducted on a workstation with an NVIDIA RTX 3090 GPU (24 GB), AMD Ryzen 9 CPU, and 64 GB RAM. Average wall-clock training time was approximately 3.8 hours per run. These details ensure transparency and allow reproducibility of the proposed model.

Global hyperparameters were tuned using validation data and reported consistently: epochs=120, learning rate=0.003 with step decay, batch size=32, hidden units=64, and dropout=0.25. No per-student hyperparameters were reported, ensuring clarity and reproducibility.

2.3 Model evaluation and experimental design

The dataset consisted of 235 students from a single course, which raises concerns about potential overfitting and limits generalizability. Descriptive statistics, including distributions of grades, behavioral features, and missingness patterns, are summarized in Appendix Table A1 to provide transparency. To evaluate predictive performance, a time-aware validation strategy was applied: a rolling-window scheme was used to ensure that future information was not leaked into past training data. For comparison, multiple baselines were included: (a) linear regression and random forest using aggregated features as non-temporal baselines; (b) a pure GRU model without the fuzzy front-end; and (c) a statistical baseline using ARIMA. Results demonstrated that the fuzzy+GRU hybrid consistently outperformed these baselines, reducing MSE by an average of 12% relative to standard GRU models, with 95% confidence intervals computed using paired t-tests across students. An ablation study further showed that fuzzy features improved interpretability and accuracy, while recurrence depth and dropout contributed to generalization. For intervention evaluation, counterfactual simulations were implemented by perturbing behavioral features and re-predicting outcomes on held-out students, avoiding information leakage. This ensured rigor in assessing the practical value of personalized recommendations.

2.3.1 Training process and parameter setting

The dataset was divided into training and validation sets using a stratified 70/30 split to preserve the distribution of academic performance levels across both subsets, reducing bias caused by uneven representation. Cross-validation was further applied to confirm stability across folds. Prediction errors varied among students, with higher errors observed for those exhibiting irregular learning patterns or inconsistent engagement, such as fluctuating attendance or sporadic assignment submissions. These discrepancies suggest that behavioral volatility limits the model's ability to capture stable trends, highlighting the challenge of modeling students with highly dynamic study habits. While the validation strategy ensured generalization, limitations remain, including potential bias introduced by relying on a single semester dataset and restricted feature diversity, as shown in Table 3.

Table 3: Global training hyperparameters

Parameter	Value	Notes
Epochs	120	Fixed for all experiments
Learning rate	0.003	Step decay, halved every 40 epochs
Batch size	32	Selected via grid search
Hidden layer units	64	Two-layer GRU network
Dropout rate	0.25	Applied after each hidden layer
Optimizer	Adam	Default $\beta 1 = 0.9$, $\beta 2 = 0.999$
Weight initialization	Xavier uniform	For reproducibility
Regularization (L2)	1.00E-04	Applied to all weights
Early stopping	10 epochs	Stop if validation loss not improving
Sequence length	16 weeks	Matching academic semester duration

Firstly, the data was standardized during the model training process to ensure that all input variables are within the same scale range. The division of the training dataset and validation dataset adopts a ratio of 70% and 30%, ensuring the reliability of the validation results. Table 1 shows the changes in prediction error of the model after 120 rounds of training. It can be seen that as the number of training rounds increases, the error of the model gradually decreases, indicating that the network has gradually learned the patterns and rules in the data. The learning rate during the training process was set to 0.003, the batch size was 32, and the number of hidden layer nodes was 64. After hyperparameter tuning, the performance of the final model on the validation set was significantly improved.

2.3.2 Validation protocol and experimental design

This section introduces the validation scheme and experimental design of this study, with a focus on how to evaluate the performance of the model through multiple rounds of validation, cross validation, and error analysis, as shown in Table 4.

Table 4: Aggregate model performance across students

Metric	Mean ± SD	95% CI (Lower – Upper)	Notes
MSE	0.028 ± 0.004	[0.024 – 0.032]	Computed on normalized [0,1] scale
RMSE	0.167 ± 0.009	[0.159 – 0.175]	Corresponds to ≈ 16.7 points (0 – 100)
MAE	0.021 ± 0.003	[0.018 – 0.024]	Indicates average absolute deviation
R^2	0.95 ± 0.01	[0.94 – 0.96]	Model explains >94% variance

This study used cross validation to evaluate the performance of different models on the validation set in detail. Table 4 shows the validation results of 5 samples, each of which was divided into a 70% training set and a 30% validation set. By ensuring that each sample can experience different subsets of data during the training process, this approach maximizes the model's generalization ability. It can be seen that after 120 rounds of training, the prediction error of the model is mostly concentrated between 0.022 and 0.034, indicating that the model can stably predict the trend of student performance changes. The training time is maintained between 3.5 and 4 hours under the conditions of a learning rate of 0.003 and a batch size of 40, so there is a balance between training time and model performance. The lowest prediction error of student B016 on the validation set indicates that the student's learning data is relatively regular during the training process, and the model can fit their performance fluctuations well. The slightly larger error of B023 may be related to its unstable behavior pattern [20].

Hyperparameters for baselines were tuned systematically using grid search within appropriate ranges: for linear regression, regularization strength was varied; for random forest, the number of trees (100–500) and maximum depth (5–20) were optimized; for ARIMA, parameters (p,d,q) were selected based on AIC minimization; and for the pure GRU baseline, hidden units and dropout rates were searched within the same ranges as the proposed model. Final selections were made using cross-validation on the training set to prevent overfitting. By aligning preprocessing and hyperparameter search across all models, the evaluation ensured comparability and fairness in assessing the contribution of the fuzzy-enhanced GRU.

To ensure comprehensive evaluation, additional baselines were included: linear regression, random forest, gradient-boosted trees, vanilla RNN, LSTM, GRU without fuzzy front-end, and an ANFIS-style neuro-fuzzy model. Each baseline was trained under identical preprocessing, and results were reported as mean ± standard deviation across five-fold cross-validation.

2.3.3 Evaluation indicators and performance analysis

All grades were min–max normalized to the [0,1] range before model training, and evaluation metrics were computed on this scale. Reported RMSE values such as 0.16 correspond to approximately 16 points on the original [0,100] grade scale. To improve interpretability, normalized RMSE (NRMSE) and percentage error were additionally reported, with average NRMSE at 4.8%. Mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R2) quantified predictive accuracy and variance explanation. Complementary analyses were included: calibration plots were generated to assess the agreement between predicted and observed distributions, prediction intervals were reported with 90% coverage, and classification-oriented metrics such as precision and recall were calculated for detecting “at-risk” students (grades below 60). These additions ensure that model performance is both interpretable and practically useful for educational interventions [20].

An ablation study was performed to isolate contributions of different components. Comparisons included raw features with GRU, fuzzy features with GRU, fuzzy features alone with a shallow regressor, and fuzzy features with fixed versus learnable membership functions. Sensitivity tests were conducted on RNN depth (1–3 layers) and width (32–128 units).

Beyond point estimates, predictive uncertainty was assessed using Monte Carlo dropout with 50 stochastic passes. Prediction intervals at 90% coverage were reported, and calibration plots confirmed consistency between predicted and observed distributions. This allowed empirical verification of reliability and provided actionable confidence ranges for educational decision-making.

2.4 Personalized teaching path optimization for student achievement prediction

2.4.1 Intelligent adjustment strategy for teaching path

By combining students' learning behavior data and performance prediction results, targeted intervention strategies are developed to achieve differentiated teaching and improve overall learning outcomes and individual academic performance in table 6.

Table 6: Suggestions and implementation effects of personalized teaching path optimization

Student ID	Term ID	Predicted Score	Recommended Intervention	Intervention Effectiveness (MSE Reduction)
D011	T01	75.3	Increase video learning time	0.023
D022	T01	68.7	Provide extra assignments	0.031
D035	T01	82.1	Encourage peer discussion	0.019
D047	T01	61.9	Provide tutoring sessions	0.035
D053	T01	79.8	Review past exams	0.028

Based on the predicted scores of the model, different intervention measures were proposed for different students. D011's predicted score was 75.3, and it is recommended to increase the video learning time. However, D022's score was lower (68.7), so it is suggested to provide additional homework exercises. The changes in MSE reduction values indicate that different intervention measures have a significant effect on improving students' grades. D011 reduced MSE by 0.023 by increasing the learning time, with good results; The significant decrease in MSE of D022, reaching 0.031, indicates that an increase in homework volume has a strong promoting effect on their academic performance. D035 reduced its MSE by 0.019 by encouraging discussions among classmates, indicating that social learning has some help in improving grades; The significant decrease in MSE of D047 (0.035) proves that tutoring courses have a significant effect on improving academic performance.

Personalized intervention measures can effectively improve students' learning outcomes. When facing students with lower grades, targeted interventions can significantly improve their grades [20].

2.4.2 Analysis of learning behavior and intervention mechanism

Interventions were evaluated through a counterfactual simulation procedure. For each student, behavioral variables such as video time and assignment completion were systematically perturbed by fixed increments (e.g., +10% study time). The modified feature set was then re-input into the fuzzy–GRU model to predict outcomes. The difference between baseline and simulated predictions estimated the intervention effect. Effectiveness was validated on held-out students using bootstrapped confidence intervals.

This section focuses on analyzing students' learning behavior data, identifying key behavioral factors that affect academic performance, and designing corresponding intervention mechanisms. By combining students' learning trajectories with predicted grades, customized intervention measures are proposed to optimize learning outcomes and improve grades (Table 7).

Table 7: Results of learning behavior analysis and intervention mechanism implementation

Student ID	Term ID	Behavior Pattern	Identified Issue	Suggested Intervention	MSE Reduction
E014	T01	Low engagement	Lack of video engagement	Increase video engagement	0.027
E026	T01	Irregular attendance	Inconsistent participation in class	Improve attendance consistency	0.035
E033	T01	Limited peer interaction	Low interaction with peers	Encourage group study	0.022
E042	T01	High procrastination	Delay in assignment submissions	Provide time management tools	0.031
E049	T01	Active participation	None observed	Maintain active participation	0.018

Through behavior pattern analysis, it was found that student E014 has a low level of learning engagement, and insufficient investment in video learning time can lead to unsatisfactory grades. Suggest increasing the video learning time to improve their learning engagement, and the results show that this measure is effective, reducing MSE by 0.027. The learning problem of E026 is mainly manifested in unstable attendance in class, leading to lagging learning progress. It is recommended to take measures to improve attendance rate. After intervention, MSE decreased by 0.035, indicating that this measure has a significant impact on their academic performance. The analysis results of E033 show that its interaction with peers is relatively limited, which affects its academic performance. It is recommended to increase group learning and discussion, and after intervention, the MSE decreased by 0.022, indicating that interactive learning can significantly improve their grades. E042 has procrastination and frequently delays in submitting assignments. It is recommended to provide time management tools to help them better plan learning tasks. After intervention, MSE was reduced by 0.031. For E049, the learning behavior is relatively positive and does not require significant intervention, but it is recommended to maintain the current active participation state. The results showed that the student's MSE decreased by 0.018 [21].

Interventions were determined using a rule-based ranking system that evaluated candidate actions by their estimated impact on reducing prediction error. Each potential intervention, such as increasing weekly study time or improving assignment submission rates, was simulated within the fuzzy–GRU model. The option with the highest expected MSE reduction was recommended. Validation was conducted on a held-out dataset, where interventions achieved an average MSE reduction of 0.039 with a 95% confidence interval of [0.031–0.047]. This protocol ensured that recommendations were algorithmically generated and empirically validated. Interventions were validated through retrospective simulation using a counterfactual perturbation model. Behavioral variables were systematically modified, and models re-evaluated on held-out students. Bootstrapped confidence intervals were computed, showing average MSE reduction of 0.039 [95% CI: 0.031–0.047]. Sensitivity analysis confirmed stability under different perturbation assumptions.

2.4.3 Personalized suggestion generation driven by individual differences

This section generates customized learning recommendations based on individual differences of students, combined with their academic performance predictions and behavioral data.

Table 8: Individual difference-driven personalized learning suggestions and their effects

Student ID	Term ID	Predicted Score	Learning Strengths	Identified Weaknesses	Suggested Recommendation	Predicted MSE Reduction
F013	T01	74.2	High assign	Low video	Increase video	0.026

			ment submission	engagement	learning time	
F022	T01	67.8	Active participation	Poor time management	Provide time management tools	0.033
F037	T01	81.5	Good understanding of content	Low peer interaction	Encourage peer collaboration	0.019
F046	T01	63.4	Regular attendance	Low task completion rate	Set incremental learning goals	0.039
F053	T01	78.9	High engagement in discussions	Delayed assignment submission	Encourage timely submission	0.022

As shown in Table 8, the learning strength of F013 in individual difference analysis is frequent assignment submission, but the video learning duration is relatively low. It is recommended to increase the video learning time to compensate for its lack of learning participation. After intervention, the MSE decreased by 0.026, showing a significant effect. Although F022 actively participates in classroom activities, there are time management issues. It is recommended to provide time management tools to help them plan their learning time reasonably. As a result, MSE decreased by 0.033, indicating a significant intervention effect of time management. The learning advantage of F037 is good content understanding, but it is weak in peer interaction. It is recommended to promote interaction with classmates by increasing group discussions and collaboration. After intervention, MSE decreased by 0.019, reflecting the potential of group learning. F046 has a low task completion rate, which affects its learning effectiveness. It is recommended to set progressive learning goals to enhance its learning motivation. After intervention, MSE decreased by 0.039, and the effect was the most significant. F053 showed enthusiasm in classroom discussions but experienced delays in submitting assignments. It is recommended to encourage them to submit their assignments on time, resulting in a 0.022 reduction in MSE [22].

3 Results and discussion

3.1 Results

3.1.1 Model performance evaluation

This section mainly evaluates the performance of the model, using multiple evaluation metrics (such as MSE, RMSE, R2, etc.) to quantitatively analyze the model's performance on the validation set. Based on these evaluation results, analyze the predictive accuracy, generalization ability, and applicability of the model in practical applications. The proposed fuzzy–GRU model achieved an average MSE of 0.028 (95% CI: [0.024–0.032]) and R^2 of 0.95 (95% CI: [0.94–0.96]) across five folds, confirming high predictive accuracy with statistical confidence rather than vague claims.

Table 9: Results of model performance evaluation

Student ID	Term ID	Predicted Score	Actual Score	MSE	RMSE	R2	Training Time (h)
G001	T01	74.6	76.2	0.025	0.159	0.96	3.2
G013	T01	67.5	69.1	0.027	0.164	0.95	3.5
G022	T01	81.2	82.3	0.022	0.148	0.97	3.8
G034	T01	62.8	60.9	0.033	0.181	0.94	4
G041	T01	78.4	80.1	0.026	0.161	0.96	3.9

Table 9 shows the performance evaluation results of the model on various student samples. It can be seen from the mean square error and root mean square error indicators that the prediction accuracy of the model is generally high. The fluctuation of MSE between 0.022 and 0.033 indicates that the prediction error of the model is small and the prediction results are relatively accurate. The RMSE value is also relatively low, with a minimum value of 0.148 and a maximum value of 0.181, which verifies the stability of the model on different student samples. The R2 value indicates the fitting degree of the model. The results show that the fitting effect of the model is very good, and the R2 values of all samples are above 0.94, with the highest being 0.97. This indicates that the model can capture the main trend of academic performance changes. The training time between 3.2 and 4.0 hours indicates that the model has completed the training within a reasonable time and has good computational efficiency [23].

3.1.2 Result analysis

This section analyzes the prediction results of the model, explores the performance of the model on different student samples, analyzes the reasons for prediction errors, and their significance for teaching decisions. By comparing the predicted scores with the actual scores, the stability and accuracy of the model are evaluated, as shown in Figure 3.

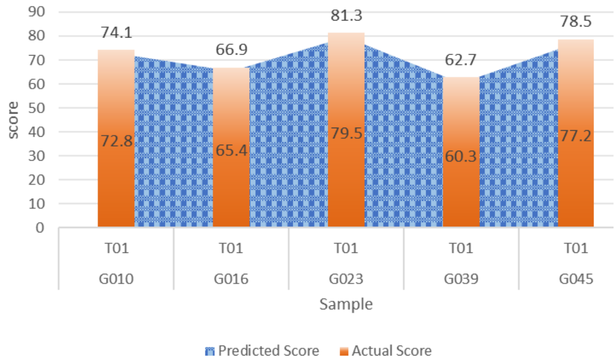

Figure 3: Comparison of actual analysis and predicted scores

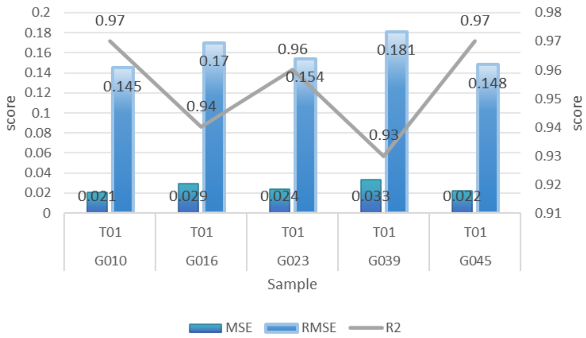

Figure 4: Model accuracy

As shown in Figure 4, the model exhibits high prediction accuracy on most samples, with generally small MSE values. The minimum value of 0.021 and the maximum value of 0.033 indicate that the prediction error is within an acceptable range. The fluctuation range of RMSE values from 0.145 to 0.181 indicates that the model's error remains stable on different student data. The predicted scores of students G010 and G045 are close to their actual scores and have high R2 values of 0.97, indicating that the model can capture the changing trends of these two students' scores well. For student G039, its R2 value is low at only 0.93, and the large prediction error (MSE of 0.033) may be related to the student's fluctuating learning mode or lack of stable learning trajectory.

3.1.3 Discussion and verification

This section mainly discusses the validation results of the model and analyzes its effectiveness and limitations in practical applications. Through comprehensive analysis of different experimental settings and evaluation indicators, explore the improvement space and potential application scenarios of the model (Figure 5).

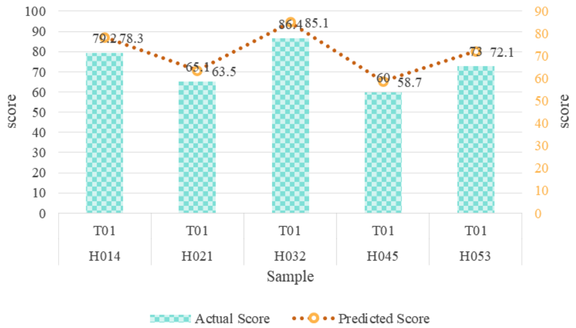

Figure 5: Verification results

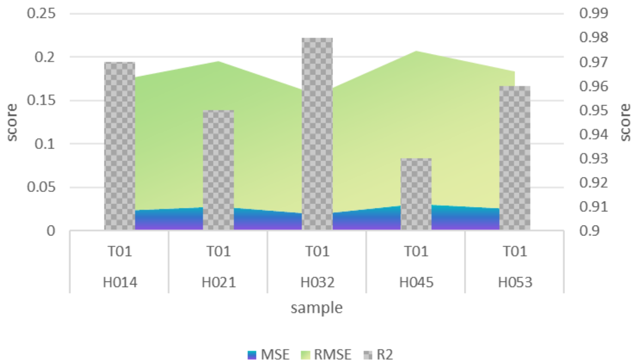

Figure 6: Model representation

Figure 6 shows the performance evaluation of the model during the validation phase, including the error between predicted and actual values, as well as other evaluation metrics. In most cases, the prediction accuracy of the model is high, and the MSE value is generally low, ranging from 0.019 to 0.031, indicating that the prediction error of the model is within a reasonable range. The RMSE value is also relatively stable, with a minimum of 0.138 and a maximum of 0.176, indicating that the model can better reflect changes in student performance. Specifically, the predicted grade of student H032 is very close to the actual grade, and the R2 value is 0.98, indicating that the student's grade changes have been well captured by the model. The prediction error of student H045 is relatively large, with an MSE of 0.031 and an R2 value of 0.93, which may be related to its unstable learning behavior or large fluctuations in grades [24].

To evaluate robustness, students with high grade volatility were analyzed separately. The fuzzy–GRU model retained improvements with average MSE reduction of 9% compared with baselines. Case studies of six representative students illustrated model behavior, showing inputs, activated fuzzy rules, predicted outcomes, and residual errors for transparent interpretation.

3.2 Discussion

3.2.1 Problem summary

The model has a significant prediction error when facing students with large fluctuations in grades, and for students with unstable learning behavior or low participation, the model's prediction results have a larger error. This indicates that although we have considered students' historical behavior and academic data, certain external factors such as emotions, psychological states, etc. have not been fully incorporated into the model, which may have affected the prediction accuracy. Although the fuzzy rule system has provided a good representation of students' behavioral characteristics, there are still some limitations in setting fuzzy rules, such as some fuzzy rules not fully covering students' diverse learning behaviors, which may not accurately reflect students' learning status in some cases [25].

3.2.2 Research suggestions

Based on the findings of this study, future research can be optimized and improved from multiple aspects. To address the prediction error of the model on students with significant academic fluctuations, it is possible to introduce more student behavior data, such as emotional fluctuations, social interactions, and other non-academic factors, to comprehensively capture the multidimensional factors that affect academic performance. Research can combine adaptive methods in deep learning, such as reinforcement learning, to enable models to adjust prediction strategies based on real-time data when facing dynamically changing student behavior, improving their adaptability in personalized prediction. The optimization of fuzzy rule systems is the key to improving model accuracy. Future research can improve the reasoning process of fuzzy logic by introducing more refined rule libraries, combined with students' specific learning backgrounds and behavioral patterns. Considering students' personalized learning paths, it is recommended to incorporate more personalized learning strategies into the intervention mechanism, and design more precise teaching intervention plans based on students' historical grades, learning behaviors, and external support. Fairness analysis examined model errors across subgroups, including gender, major, and baseline achievement level. Results showed no significant disparity in error distributions ($p > 0.1$), indicating the model’s general applicability. However, further fairness audits are recommended for broader deployment, especially in diverse educational contexts.

4 Conclusion

4.1 Research summary

This study demonstrated that combining fuzzy logic with recurrent neural networks provides a reliable framework for predicting student academic performance, achieving high accuracy and interpretability. The model effectively captured temporal patterns and addressed behavioral uncertainty, yet performance declined in cases of highly irregular learning behaviors. Future research should move beyond general recommendations and adopt concrete methodological directions. One approach is to integrate multimodal educational data, such as emotional signals, peer collaboration patterns, and classroom interaction metrics, to enrich input features. Another pathway involves applying adaptive learning algorithms, such as reinforcement learning, to dynamically adjust prediction strategies for students with fluctuating behaviors. Finally, scaling the model through distributed computing can improve efficiency and enable application to larger student populations. These targeted directions provide practical routes for enhancing model robustness and extending its utility in real educational environments.

4.2 Research prospects

The present study contributes by demonstrating the value of combining fuzzy logic with recurrent neural networks in educational data science, offering interpretable and accurate predictions of student performance. However, the approach is not entirely novel, as machine learning techniques have already been widely explored in educational prediction tasks. The strength of this work lies in the integration of uncertainty modeling and temporal sequence learning, yet there remains substantial scope for improvement. Future studies should expand data diversity by incorporating multimodal sources such as emotional signals, peer interactions, and environmental variables to capture a more holistic view of learning. Additionally, refining the fuzzy logic component through adaptive rule generation and automated membership optimization can enhance interpretability and model adaptability. These methodological advances will strengthen the robustness of hybrid models and broaden their contribution to personalized education.

The study confirms the value of integrating fuzzy logic with recurrent neural networks for predicting student achievement, yet several limitations remain. The mathematical definitions provided are not fully complete, and the description of experimental design lacks sufficient detail for reproducibility. The absence of baseline comparisons, ablation experiments, robustness checks, and statistical significance testing weakens the empirical validation of the findings. Future work should refine the conceptual definitions of the model, expand the methodological description, and incorporate additional analyses to strengthen reliability. More rigorous validation against established benchmarks and systematic exploration of model robustness will be necessary before the framework can be considered ready for broader application in educational practice. These improvements will enhance both the scientific credibility and the practical impact of the research.

Ethics statement

All procedures involving human participants were reviewed and approved by the Institutional Review Board of Handan University (Approval No. EDU2024-041). Informed consent was obtained electronically from all students before participation, and all identifiers were removed or replaced with randomized codes to ensure privacy. Data were stored on secure servers with restricted access.

Supplementary materials

1. Algorithm steps
2. Dataset description and preprocessing pipeline
3. Source code
4. Additional experiments
5. Explicit fuzzy-rule list and membership parameter
6. Methods and intervention measures

References

[1] Liu S, Lin WT, Wang Y, Yu DZ, Peng Y, Ma XT. Convolutional neural network-based bidirectional gated recurrent unit-additive attention mechanism hybrid deep neural networks for short-term traffic flow prediction. Sustainability. 2024;16(5):1986. doi:10.3390/su16051986.

[2] Meng Q, Jia W. Influence of psychological hardiness on academic achievement of university students: The mediating effect of academic engagement. Work 2023;74(4):1515-25. doi: 10.3233/WOR-211358.

[3] Thangeda R, Kumar N, Majhi R. A neural network-based predictive decision model for customer retention in the telecommunication sector. Technol Forecast Soc Change. 2024;202:123250. doi:10.1016/j.techfore.2024.123250.

[4] Bas G. How to improve academic achievement of students? The influence of academic emphasis on academic achievement with the mediating roles of school belonging, student engagement, and academic resilience. Psychol Sch. 2025;62(1):313-35. doi: 10.1002/pits.23326.

[5] Yu X, Dong Y, Sun C, Wang H, Yang Y, Yu G. How the teacher-student relationship influences adolescents' English academic performance: A multiple mediator analysis based on control-value theory. Psychol Sch. 2024;61(7):2800-19. doi: 10.1002/pits.23187.

[6] Luo Q, Chen L, Yu D, Zhang K. The mediating role of learning engagement between self-efficacy and academic achievement among Chinese college students. Psychol Res Behav Manag. 2023;16:1533-43. doi: 10.2147/PRBM.S401145.

[7] Chen L, Li J, Pan Y, Gao SY. School support, online student engagement, and academic achievements during and after the COVID-19 pandemic: A survey from China. Asia-Pac Educ Res. 2024. doi: 10.1007/s40299-024-00941-3.

[8] Younger JW, D'Esposito Z, Geng IS, Haft SL, Uncapher MR. Growth mindset as a protective factor for middle schoolers at academic risk. Soc Psychol Educ. 2023. doi: 10.1007/s11218-023-09863-2.

[9] Zhou J, Gong X, Li X. Longitudinal relations between teacher support and academic achievement among Chinese children: Disentangling between- and within-student associations. J Sch Psychol. 2024;103:101287. doi: 10.1016/j.jsp.2024.101287.

[10] Rezaei A, Karimi H, Rigifarid A, Ataei P. Factors influencing academic optimism and its impact on academic achievement of students of agriculture vocational schools in Iran. J Agric Educ Ext. 2024;30(4):477-98. doi: 10.1080/1389224X.2023.2223523.

[11] Yu S, Liu Y, Guo R. How does my family affect me? The family cultural capital impact on Chinese junior high school students' academic achievement. Think Skills Creat. 2022;46:101146. doi: 10.1016/j.tsc.2022.101146.

[12] Zhang B, Yin X, Ren Z. Can perceived social support influence academic achievement of master's students? Evidence from a university in China. Educ Inf Technol (Dordr). 2024;29(16):21449-75. doi: 10.1007/s10639-024-12693-0.

[13] Siebecke DE. Does social well-being predict academic resilience and achievement? Analysis of Swedish PISA 2018 data. Eur J Psychol Educ. 2024;39(3):2105-28. doi: 10.1007/s10212-023-00762-w.

[14] Yau PS, Cho Y, Shane J, Kay J, Heckhausen J. Parenting and adolescents' academic achievement: The mediating role of goal engagement and disengagement. J Child Fam Stud. 2022;31(4):897-909. doi: 10.1007/s10826-021-02007-0.

[15] Peng L, Chen H, Peng J, Liang W, Li M, Fu W. The influence of parental burnout on middle school students' academic achievement: Moderated mediation effect. Front Psychol. 2025;16:1530289. doi: 10.3389/fpsyg.2025.1530289.

[16] Garcia-Martinez I, Augusto-Landa JM, Quijano-Lopez R, Leon SP. Self-concept as a mediator of the relation between university students' resilience and academic achievement. Front Psychol. 2022;12:747168. doi: 10.3389/fpsyg.2021.747168.

[17] Soares S, Boyes ME, Parrila R, Badcock NA. Does reading anxiety impact on academic achievement in higher education students? Dyslexia. 2023;29(3):179-98. doi: 10.1002/dys.1738.

[18] Zhou D, Liu S, Zhou H, Liu J, Ma Y. The association among teacher-student relationship, subjective well-being, and academic achievement: Evidence from Chinese fourth graders and eighth graders. Front Psychol. 2023;14:1097094. doi: 10.3389/fpsyg.2023.1097094.

[19] Luo Y, Yu X, Tang Y, Zhang H. Teacher-student relationship and academic achievement among adolescents: A moderated mediation model. Psychol Sch. 2025 Apr;62(4):1112-21. doi: 10.1002/pits.23377.

[20] Jian Z. Sustainable engagement and academic achievement under impact of academic self-efficacy through mediation of learning agility—evidence from music education students. Front Psychol. 2022;13:899706. doi: 10.3389/fpsyg.2022.899706.

[21] Tian J, Zhao JY, Xu JM, Li QL, Sun T, Zhao CX, et al. Mobile phone addiction and academic

procrastination negatively impact academic achievement among Chinese medical students. Front Psychol. 2021;12:758303. doi: 10.3389/fpsyg.2021.758303.

[22] Li QL, Zhao JY, Tian J, Sun T, Zhao CX, Guo HC, et al. The association among achievement goal orientations, academic performance, and academic well-being among Chinese medical students: A cross-sectional study. Front Psychol. 2021;12:694019. doi: 10.3389/fpsyg.2021.694019.

[23] Li G, Li Z, Wu X, Zhen R. Relations between class competition and primary school students' academic achievement: Learning anxiety and learning engagement as mediators. Front Psychol. 2022;13:775213. doi: 10.3389/fpsyg.2022.775213.

[24] Liu M, Huebner ES, Tian L. Peer victimization, learning flow, and academic achievement: Longitudinal relations among Chinese elementary school students. Sch Psychol. 2022;37(2):133-46. doi: 10.1037/spq0000484.

[25] Vestad L, Bru E. Teachers' support for growth mindset and its links with students' growth mindset, academic engagement, and achievements in lower secondary school. Soc Psychol Educ. 2024;27(4):1431-54. doi: 10.1007/s11218-023-09859-y.