# Deep Learning-Based Student Grade Prediction and Interpretability Analysis

Youhui Yu*

Fujian Chuanzheng Communications College, Fuzhou, China

*67654723@qq.com

**Abstract.** Addressing the limited accuracy of traditional machine learning methods in handling complex nonlinear features, as well as the challenges of applying deep learning models to pedagogical practice due to their inherent lack of interpretability, this study proposes a student grade prediction framework that integrates Gated Recurrent Units (GRUs) with SHAP interpretability analysis. Using an educational dataset from Kaggle, an end-to-end prediction model was constructed, selecting study duration, sleep duration, attendance rate, and previous academic performance as input features. Experimental results demonstrate that the GRU model achieves a Root Mean Square Error of 3.0884 in grade prediction, significantly outperforming baseline models such as Support Vector Machine and Random Forest. Furthermore, SHAP analysis elucidates the coupling effects of these four features on academic performance, providing an empirical basis for data-driven precision instruction.

**Keywords:** Deep learning; Student grade prediction; Interpretability analysis; Pedagogical suggestions.

## 1 Introduction

The proliferation of smart campuses and online learning platforms has generated massive volumes of multidimensional student data, ranging from classroom attendance and online interactions to sleep habits. This influx of data has catalysed the emergence of Educational Data Mining (EDM) and Learning Analytics. However, a critical challenge for stakeholders remains: how to extract practical features from these complex datasets to accurately assess learning states and predict academic performance [1–3].

Student grades serve not only as a critical metric for teaching quality but also as a comprehensive reflection of knowledge mastery, learning attitude, and psychological state. However, traditional pedagogical models often suffer from significant latency; instructors typically only identify academic difficulties in students after mid-term or final examinations have concluded. This retrospective evaluation mechanism usually causes educational interventions to miss the optimal time window, leading to academic disengagement among students who fail to receive timely assistance, and potentially increasing the risk of grade retention or dropout. Therefore, constructing an efficient and precise student-grade prediction model has profound practical significance [4].


I. A. Khan et al. (eds.), *Proceedings of the 2026 5th International Conference on Educational Innovation and Multimedia Technology (EIMT 2026)*, Atlantis Highlights in Social Sciences, Education and Humanities 51,
https://doi.org/10.2991/978-94-6239-691-3_20

Early research on grade prediction primarily relied on traditional statistical methods, such as linear or logistic regression. While these methods offer high interpretability, they often struggle to capture non-linear relationships and complex interactions in the data, resulting in limited predictive accuracy [5]. With the rise of artificial intelligence, machine learning algorithms such as Support Vector Machines (SVM) and Random Forests have been widely applied in this domain, significantly enhancing predictive performance [6, 7]. Nevertheless, when applied to increasingly complex, high-dimensional student behavioral data, the feature-extraction capabilities of traditional machine learning models remain insufficient. In recent years, deep learning has achieved breakthroughs in fields like image recognition and natural language processing, owing to its powerful end-to-end feature-learning capabilities. The Gated Recurrent Unit (GRU), a pivotal branch of deep learning, possesses inherent advantages in processing sequential data and mining complex dependency relationships. By leveraging GRU's structural characteristics, it is possible to map non-linear combinatorial relationships within the feature space at a deeper level, thereby uncovering latent patterns that influence academic performance.

While deep learning models excel in predictive accuracy, their inherent "black box" nature hinders practical application in the education sector. Lacking interpretability, it is challenging to translate predictions into concrete pedagogical recommendations [8]. To address this, we propose a deep learning-based framework for student grade prediction and interpretability analysis. We first construct a GRU-based regression model to achieve high-precision prediction using multi-dimensional features, including study duration, sleep time, attendance rate, and prior academic performance. Subsequently, we employ the SHAP method to provide post-hoc explanations, quantifying each feature's marginal contribution to the final grade. By integrating deep learning with interpretability, this approach bridges the gap between algorithms and educational practice, supporting the development of intelligent, personalized, and trustworthy educational systems.

## 2 Dataset Analysis

### 2.1 Overview and Structure

The dataset employed in this study, titled "Student Academic Performance Trends," was sourced from the Kaggle data science competition platform. Designed to capture critical behavioral traits and lifestyle factors influencing student learning outcomes, this dataset comprises sample records from 200 students. As shown in Table 1, behavioral features include study duration, sleep duration, attendance rate, and previous academic performance, with the current exam score serving as the target variable.

### 2.2 Overview and Structure

Figure 1 illustrates the influence of input features on Exam Score. Hours_studied exhibits a significant positive correlation, with data following a clear upward diagonal, reaffirming the "input-output" pedagogical principle. Attendance_percent similarly

displays a steep positive linear trend, underscoring the critical role of classroom engagement. Conversely, sleep_hours shows a dispersed distribution with only a marginal positive trend, suggesting it serves as an auxiliary factor with high individual variance rather than a dominant determinant. Finally, previous_scores demonstrate the strongest linear dependency with tight clustering; this feature reflects baseline capabilities and serves as the core foundation for high-precision prediction.

**Table 1.** Dataset Field Definitions.

| Field Name | Variable Type | Font size and style |
|---|---|---|
| student_id | Identifier | Unique student identity identifier |
| hours_studied | Input Feature | Average duration of study invested in preparation for the exam (hours) |
| sleep_hours | Input Feature | Average daily sleep duration (hours) |
| attendance_percent | Input Feature | Classroom attendance rate (%) |
| previous_scores | Input Feature | Historical scores achieved in past examinations |
| exam_score | Target Variable | Final score achieved in the current examination |

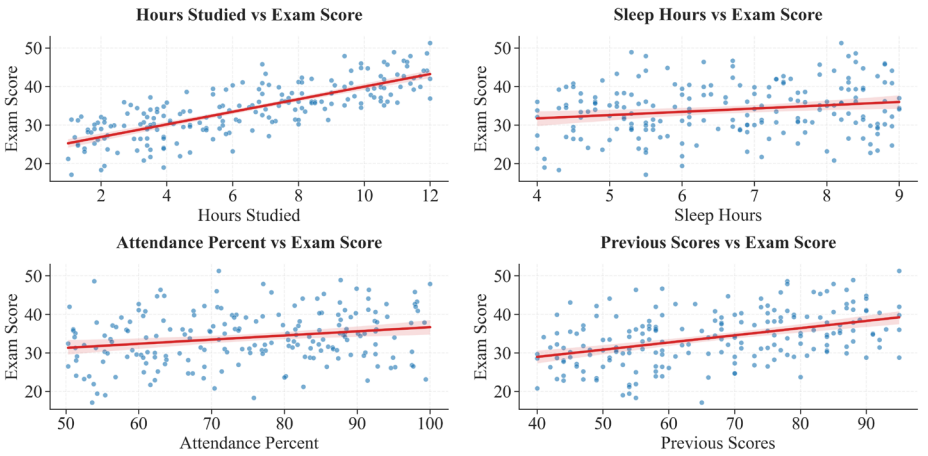


**Fig. 1.** A figure caption is always placed below the illustration.

Correlation analysis reveals complex interdependencies between the aforementioned features and the target variable, confirming their suitability for neural network training. Prior to the initiation of network training, both the four input features and the target labels undergo Min-Max normalization, as formulated in Equation (1).

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{1}$$

# 3 Deep Learning-Based Grade Prediction Model

## 3.1 Gated Recurrent Unit

This study employs the Gated Recurrent Unit (GRU) as the core prediction model [9]. The core of the GRU lies in controlling the update of the hidden state $h_t$ via two gating units: the reset gate and the update gate, as illustrated in Figure 2. For a given time step $t$, with input vector $x_t$ and the previous hidden state $h_{t-1}$, the calculation process of the GRU is as follows:

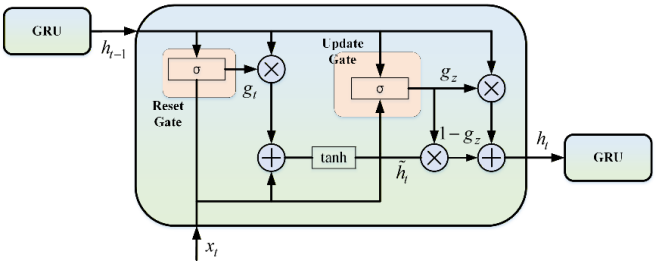


**Fig. 2.** Gated Recurrent Unit

Reset Gate $r_t$: Determines how much information from the previous hidden state needs to be forgotten, as shown in Equation (2).

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{2}$$

Update Gate $z_t$: Determines the amount of state information from the previous moment to be retained for the current moment. It simultaneously assumes the functions of both the forget and input gates found in LSTM, as shown in Equation (3).

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{3}$$

Candidate Hidden State $\tilde{h}_t$: Calculates the candidate value for the new state at the current moment after filtering historical information via the reset gate, as shown in Equation (4).

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \tag{4}$$

Final Hidden State $h_t$: Combines with the update gate to perform linear interpolation between the historical state and the candidate state, yielding the current final output, as shown in Equation (5).

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{5}$$

where, $\sigma$ denotes the Sigmoid activation function; $\odot$ represents the Hadamard product; $W$ and $b$ the corresponding weight matrices and bias terms, respectively. In this

study, we treat each student's feature vector as an input with a sequence length of 1, utilizing the GRU to extract deep interaction patterns among features. The total sample set was randomly partitioned at a ratio of 8:2, with 80% of the data serving as the training set and the remaining 20% allocated as the test set.

The GRU neural network established in this study consists of a three-layer architecture, with neuron counts sequentially set at 128, 32, and 1. Although the input dimension is limited, the relationship between behavioral features and academic performance is highly non-linear and coupled. We employ a wide initial layer to project these features into a high-dimensional space to disentangle latent factors, followed by a bottleneck layer to enforce compact representation learning and prevent overfitting on the small dataset. The network is configured with a learning rate of 0.001 and undergoes 300 training iterations.

### 3.2 Principles of SHAP Interpretability Analysis

This study introduces SHAP theory, aiming to quantify the contribution of each feature to the final predicted grade [10].

The SHAP method originates from Shapley values in cooperative game theory. It interprets the prediction value for a specific sample as the sum of attributions for all feature values of that sample. For a model containing $M$ features, the predicted value $f(x)$ for a specific sample $x$ can be expressed as a linear superposition of the average prediction value $\phi_0$ and the contribution values $\phi_i$ of each feature, as shown in Equation (6):

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{6}$$

To calculate the SHAP value $\phi_i$ for feature $i$, it is necessary to consider the marginal contribution of that feature across all possible feature combinations. Its mathematical definition is as shown in Equation (7):

$$\phi_i = \sum_{S \subseteq F\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f_x(S \cup \{i\}) - f_x(S)] \tag{7}$$

where $F$ is the set of all features; $S$ is a subset of features excluding feature $i$ and $f_x(S)$ represents the model's predicted output when retaining only the features in subset $S$.

## 4 Results and Discussion

### 4.1 Student Grade Prediction Results

Figure 3 illustrates the prediction error distribution of the GRU model, showing that the predicted values exhibit high concordance with the actual grades and maintain a

consistent overall trend. Whether in the data-dense intermediate segments or the edge segments characterized by higher volatility, the model effectively captures the non-linear mapping relationships between input features and grades, showing no significant signs of overfitting or underfitting.

To verify the effectiveness of the proposed GRU model, comparative experiments were conducted against Support Vector Machine (SVM) and Random Forest (RF) on the same test set, utilizing Root Mean Square Error (RMSE) and the coefficient of determination ( $R^2$ ) as evaluation metrics. The Random Forest model was configured with 20 decision trees and no maximum depth constraint, utilizing the standard squared error criterion. The SVM model employed a Radial Basis Function kernel, with the regularization parameter set to $C = 1.0$ and epsilon set to 0.1. Experimental data reveal that traditional machine learning models performed relatively modestly; the RMSE for SVM and Random Forest were 4.2884 and 3.9884, respectively, with $R^2$ values hovering between 0.80 and 0.81. In contrast, the GRU model exhibited significant performance advantages, with RMSE reduced to 3.0884 and R2R2 increased to 0.8903, thereby substantially improving the goodness of fit while significantly reducing prediction error. (See Table 2).

**Table 2.** Academic performance prediction error metric

| error metric | GRU | SVM | RF |
|---|---|---|---|
| RMSE | 3.0884 | 4.2884 | 3.9884 |
| R2 | 0.8903 | 0.8003 | 0.8103 |

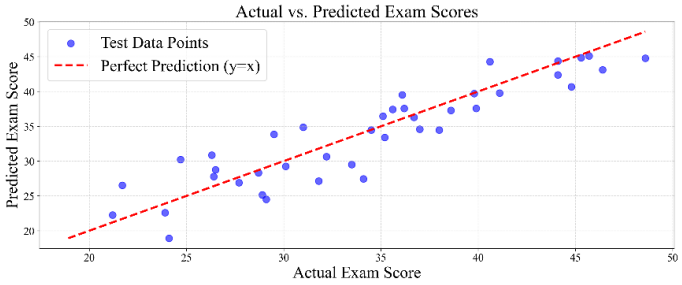


**Fig. 3.** Distribution diagram of GRU prediction error

## 4.2 Interpretability Analysis

To investigate the internal decision-making logic of the GRU model, this study employed the SHAP method to perform quantitative feature attribution. As shown in Figure 4, among the four dimensions influencing final student grades, study duration had the highest marginal contribution, with a mean SHAP value of 0.6967, significantly exceeding those of the other features. This implies that, when capturing temporal dependencies, the model identifies absolute time investment as the primary driver of academic performance prediction. Past historical grades ranked second (SHAP value

0.4321), indicating that the cumulative effect of knowledge and baseline proficiency exerts a substantial influence on the outcome. However, their weight is secondary to current effort levels. In comparison, the contributions of classroom attendance and sleep duration were 0.2627 and 0.1942, respectively.

Based on the comparative analysis of feature importance, we recommend adopting differentiated learning strategies. Since Study Duration has the most significant impact on the model's output, increasing study input is the primary way to improve grades. Meanwhile, considering the observed adverse effects of insufficient Sleep Hours in the SHAP plots, we suggest that study efforts should be balanced with physiological rest. It is worth noting that while current analysis identifies key behavioural levers, determining precise quantitative intervention thresholds (e.g., optimal study duration) remains a direction for our future research, given the complexity of educational dynamics.

While our proposed GRU-SHAP framework demonstrates strong performance in predicting and interpreting, we acknowledge several limitations in the dataset. First, the study utilises a public dataset with a relatively small sample size (N=200). This inevitably introduces sampling bias, as the data may not fully capture the diverse demographic and socio-economic factors that influence student performance globally. The primary value of this study lies in providing a roadmap for intelligent education, utilizing its own small-scale data to identify locally specific determinants of academic success.

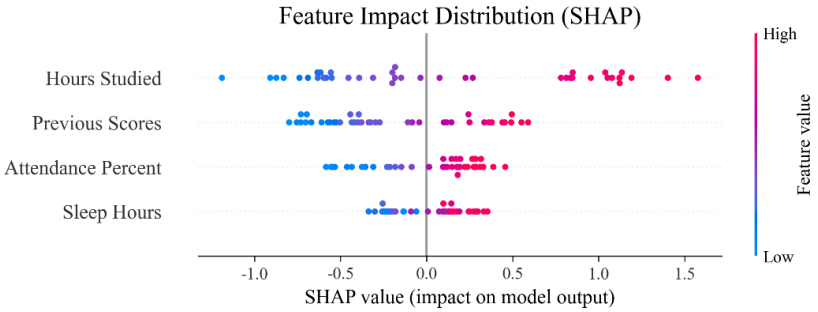


**Fig. 4.** Explainability analysis

## 5 Conclusion

This study constructed a GRU-based grade prediction model and conducted attribution analysis using SHAP theory, leading to the following conclusions:

(1) The GRU model effectively surmounted the limitations of traditional machine learning algorithms, achieving a prediction accuracy (RMSE) of 3.0884, which demonstrates the robustness of deep learning in handling complex educational data;

(2) Interpretability analysis demystified the model's "black box" nature, revealing that study duration and past grades are the core drivers determining academic performance, while attendance and sleep serve as auxiliary regulatory factors;

## Acknowledgments

This work was supported by the 2025 Fujian Provincial Education Science Planning Project (Grant No. FJJKB25167).

## References

1. Trujillo-Torres, J.-M., Hossein-Mohand, H., Gómez-García, M., Hossein-Mohand, H., Hinojo-Lucena, F.-J.: Estimating the Academic Performance of Secondary Education Mathematics Students: A Gain Lift Predictive Model. Mathematics. 8, 2101 (2020). https://doi.org/10.3390/math8122101.
2. Yao, H., Lian, D., Cao, Y., Wu, Y., Zhou, T.: Predicting Academic Performance for College Students: A Campus Behavior Perspective. ACM Trans. Intell. Syst. Technol. 10, 1–21 (2019). https://doi.org/10.1145/3299087.
3. Doz, D., Cotič, M., Felda, D.: Random Forest Regression in Predicting Students' Achievements and Fuzzy Grades. Mathematics. 11, 4129 (2023). https://doi.org/10.3390/math11194129.
4. Yang, F., Li, F.W.B.: Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. Computers & Education. 123, 97–108 (2018). https://doi.org/10.1016/j.compedu.2018.04.006.
5. Brown, T., Robinson, L., Gledhill, K., Yu, M.-L., Isbel, S., Greber, C., Parsons, D., Etherington, J.: Predictors of undergraduate occupational therapy students' academic performance during the Covid-19 pandemic: A hierarchical regression analysis. Scandinavian Journal of Occupational Therapy. 30, 475–487 (2023). https://doi.org/10.1080/11038128.2022.2123854.
6. Sarwat, S., Ullah, N., Sadiq, S., Saleem, R., Umer, M., Eshmawi, A.A., Mohamed, A., Ashraf, I.: Predicting Students' Academic Performance with Conditional Generative Adversarial Network and Deep SVM. Sensors. 22, 4834 (2022). https://doi.org/10.3390/s22134834.
7. Chen, M., Liu, Z.: Predicting performance of students by optimizing tree components of random forest using genetic algorithm. Heliyon. 10, e32570 (2024). https://doi.org/10.1016/j.heliyon.2024.e32570.
8. Liang, M., Zhou, G., He, W., Chen, H., Qian, J.: A Student Performance Prediction Model Based on Hierarchical Belief Rule Base with Interpretability. Mathematics. 12, 2296 (2024). https://doi.org/10.3390/math12142296.
9. Kord, A., Aboelfetouh, A., Shohieb, S.M.: Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques. J Comput High Educ. (2025). https://doi.org/10.1007/s12528-024-09426-0.
10. Eun Choi, J., Won Shin, J., Wan Shin, D.: Vector SHAP Values for Machine Learning Time Series Forecasting. Journal of Forecasting. 44, 635–645 (2025). https://doi.org/10.1002/for.3220.