

# Educational Data Mining of Student Learning Gains: Feature Relevance and Limits of Prediction

Nur Azlina Mohamed Mokmin<sup>1</sup>, Muhammad Aminuddin Akmal Mohd Hamizi<sup>1\*</sup>

<sup>1</sup>Centre for Instructional Technology and Multimedia, Universiti Sains Malaysia, 11800 Penang, Malaysia

\*Corresponding author: Muhammad Aminuddin Akmal Mohd Hamizi, aminmizi@student.usm.my

## HOW TO CITE:

Nur Azlina Mohamed Mokmin,  
Muhammad Aminuddin Akmal Mohd  
Hamizi (2026).  
Educational Data Mining of Student  
Learning Gains: Feature Relevance  
and Limits of Prediction.  
International Journal of Special  
Education, 41(24s), 754-766

## COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.  
Open access publication under the  
terms and conditions  
of the Creative Commons Attribution  
(CC BY)  
license (<http://creativecommons.org/licenses/by/4.0/>).

## ABSTRACT:

Educational Data Mining (EDM) is widely used to predict achievement, yet less is known about its ability to predict learning gain. This study examined whether previous academic achievement, prior knowledge, study time, and pedagogical intervention were associated with and predictive of learning gain among 120 secondary school students participating in Guided Inquiry-Based Learning (GIBL) or Problem-Based Learning (PBL). Gain Score was calculated as post-test minus pre-test. A two-way ANOVA tested intervention and prior-knowledge effects, while Linear Regression, Support Vector Machine, k-Nearest Neighbors, Random Forest, and Decision Tree models were evaluated with 10-fold cross-validation. RReliefF assessed predictor relevance and RMSE, MAE, and  $R^2$  assessed model performance. The ANOVA showed no significant effects of intervention, prior knowledge, or their interaction. Previous Grade was the most relevant predictor, followed by Study Hour, Pedagogical Intervention, and Prior Knowledge. Linear Regression yielded the lowest errors (RMSE = 2.666; MAE = 2.130), but all models had negative  $R^2$  values. The findings show that relative predictor relevance and model superiority do not necessarily indicate adequate individual-level prediction; richer process-level learning data are needed.

**Keywords:** Educational Data Mining, learning gain, academic achievement, predictive modeling, regression.

## 1. Introduction

Educational Data Mining (EDM) applies statistical and machine-learning techniques to educational data to identify patterns and predict student outcomes (Romero & Ventura, 2012; Sghir et al., 2022; Kalita et al., 2025). Its educational value, however, depends not only on identifying the best-performing algorithm but also on determining whether the available

variables contain enough information for reliable prediction (Ersozlu et al., 2024; Hidayat et al., 2026). This distinction is especially important when models rely on routinely available learner and instructional information that teachers could use without sophisticated analytics infrastructure (Yağcı, 2022; Reddy, 2025).

Much EDM research predicts final grades, success, failure, or dropout, whereas learning gain focuses on improvement during instruction. In this study, Gain Score is defined as post-test minus pre-test performance, so students with similar final achievement may still show different amounts of learning (Westphale et al., 2021; Aliyev et al., 2026). Predicting gain may therefore require information about students' academic history, prior knowledge, study behaviour, and instructional conditions rather than a single background characteristic (Hidayat et al., 2026; Erdem et al., 2025).

The present study evaluates Previous Grade, Prior Knowledge, Study Hour, and Pedagogical Intervention within one explanatory-predictive framework. A 2 x 2 ANOVA examines group-level differences in Gain Score, while Linear Regression, Support Vector Machine (SVM), k-Nearest Neighbours (kNN), Random Forest, and Decision Tree models are compared using 10-fold cross-validation. RReliefF is used to rank predictor relevance, and RMSE, MAE, and  $R^2$  are used to evaluate predictive performance. The main contribution is to distinguish relative feature relevance and model ranking from predictive adequacy at the individual level.

Accordingly, the study addresses the following research questions:

RQ1. How do students' learning gains vary across pedagogical interventions and levels of prior knowledge?

RQ2. How accurately can students' learning gains be predicted using Previous Grade, Prior Knowledge, Study Hour, and Pedagogical Intervention?

RQ3. How do different regression-based machine-learning models compare in predicting students' learning gains?

RQ4. Which learner and instructional attributes show the greatest relevance to students' learning gains?

### 1.1 Educational Data Mining in Educational Research

EDM includes classification, regression, clustering, pattern mining, and feature-selection methods (Romero & Ventura, 2020; Batool et al., 2022). Because Gain Score is continuous, regression is appropriate in the present study (Chicco et al., 2021; Plevris et al., 2022). Model evaluation should use several indicators: RMSE and MAE describe prediction error, while  $R^2$  indicates whether a model improves on a mean-based reference (Chicco et al., 2021; Plevris et al., 2022). A model may rank first among alternatives yet still have weak practical predictive value if its overall performance remains close to or below that baseline (Shaikhanova et al., 2025; Alkan et al., 2025).

Feature-relevance analysis provides a complementary perspective by identifying which available attributes contain comparatively more information about an outcome without implying causation (Batool et al., 2022; Ersozlu et al., 2024). Accordingly, the present study considers predictor ranking, comparative model performance, and predictive adequacy together rather than treating the lowest error as evidence of successful educational prediction.

### 1.2 Learning Gain as a Distinct Predictive Outcome

Learning gain represents change between two measurement points and is operationalised here as  $\text{Gain Score} = \text{Post-test Score} - \text{Pre-test Score}$ . Positive values indicate improvement, zero indicates no change, and negative values indicate lower post-test performance (Westphale et al., 2021; Aliyev et al., 2026). Gain scores shift attention from achievement status to improvement, but they can be influenced by baseline performance and measurement characteristics, making individual-level prediction difficult (Navarrete, 2018; Simonsmeier et al., 2021).

### 1.3 Previous Academic Achievement and Learning Gain

Previous academic achievement is a plausible predictor because it reflects accumulated

educational performance and experience (Sghir et al., 2022; Erdriani et al., 2026). Yet high previous achievement does not necessarily imply larger subsequent gains; students with similar academic histories may respond differently to the same learning experience, and initially higher-performing students may have less room to improve (Westphale et al., 2021; Bittermann et al., 2023). In this study, Previous Grade comes from an examination separate from the pre-test and therefore represents prior academic history rather than part of the Gain Score itself.

#### 1.4 Prior Knowledge and Learning Gain

Prior knowledge influences how learners interpret and integrate new information (Brod, 2021; Fromm et al., 2025). Its effect may depend on the instructional environment, and students within the same broad prior-knowledge category can still differ substantially (Schneider & Simonsmeier, 2025). The present study represents Prior Knowledge as Low or High and examines it both as a factor in the 2 x 2 design and as a predictor of individual Gain Score.

#### 1.5 Study Time as a Learning Behavior Predictor

Study time represents behavioural engagement during learning, but duration alone does not indicate the quality of learning. Students who

report similar study hours may differ in concentration, strategy use, self-regulation, and interaction with instructional resources (Manganelli et al., 2019; Wang & Luo, 2024). Study Hour is therefore treated as one behavioural predictor alongside academic-history, learner, and instructional variables.

#### 1.6 Pedagogical Intervention as an Instructional Predictor

Learning occurs within an instructional environment, so Pedagogical Intervention may add information beyond learner characteristics (Westphale et al., 2021; Wang & Luo, 2024). The study compares Guided Inquiry-Based Learning (GIBL) and Problem-Based Learning (PBL), while avoiding the assumption that intervention labels alone are sufficient to explain or predict individual improvement.

#### 1.7 Research Gaps

The literature suggests that learning improvement reflects a combination of academic history, prior knowledge, learning behaviour, and instructional conditions, yet these variables are often considered separately. The present study integrates all four within one EDM framework, compares five regression algorithms, and ranks predictor relevance to examine both the value and the limits of readily interpretable educational information for predicting Gain Score.

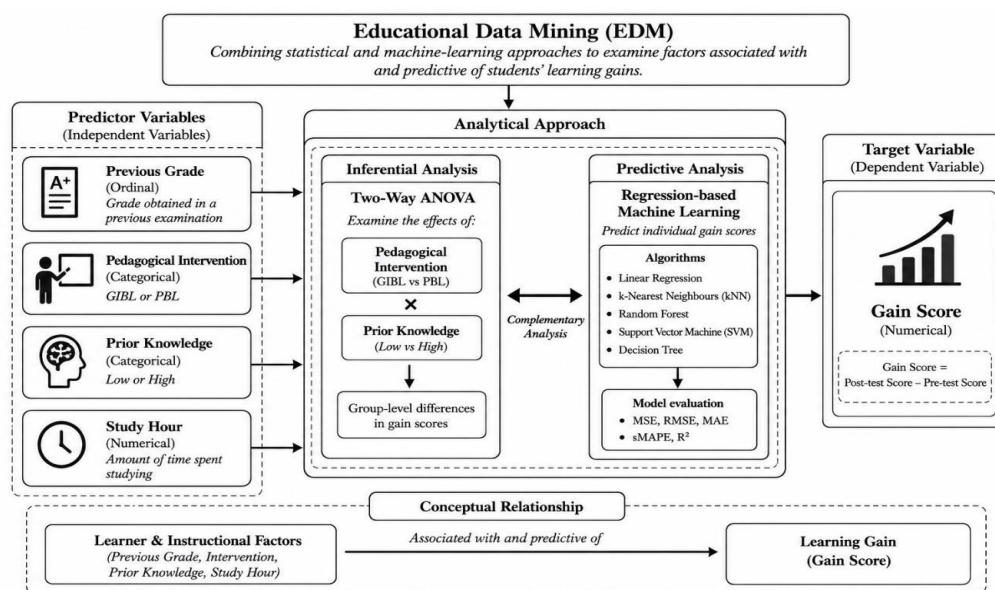


Figure 1: Conceptual Framework

The framework links learner and instructional variables to complementary inferential and predictive analyses of learning gain.

## 2. Methodology

### 2.1 Research Design

This study used a quantitative explanatory-predictive EDM design. The explanatory component examined whether Gain Score differed by Pedagogical Intervention (GIBL vs. PBL), Prior Knowledge (Low vs. High), or their interaction. The predictive component examined whether Previous Grade, Prior Knowledge, Study Hour, and Pedagogical Intervention could estimate individual Gain Scores. Inferential and predictive evidence were treated as complementary rather than interchangeable.

The analytical process followed the logic of Knowledge Discovery in Databases: data preparation, descriptive and inferential analysis, predictive modelling, cross-validation, feature-

relevance analysis, and integrated interpretation. Predictive performance was explicitly compared with a simple mean-based reference so that the best-performing candidate model would not automatically be interpreted as adequate.

### 2.2 Participants and Educational Context

The dataset comprised 120 Form 4 students who participated in a biology learning intervention. Thirty students were represented in each condition: GIBL-Low, GIBL-High, PBL-Low, and PBL-High. Low and High refer specifically to prior-knowledge levels rather than general academic achievement.

Previous Grade was obtained from an examination conducted before the intervention and was distinct from the pre-test. Prior Knowledge represented knowledge before instruction, Study Hour represented time devoted to study, and Pedagogical Intervention identified GIBL or PBL. Student identification numbers were excluded from all analyses.

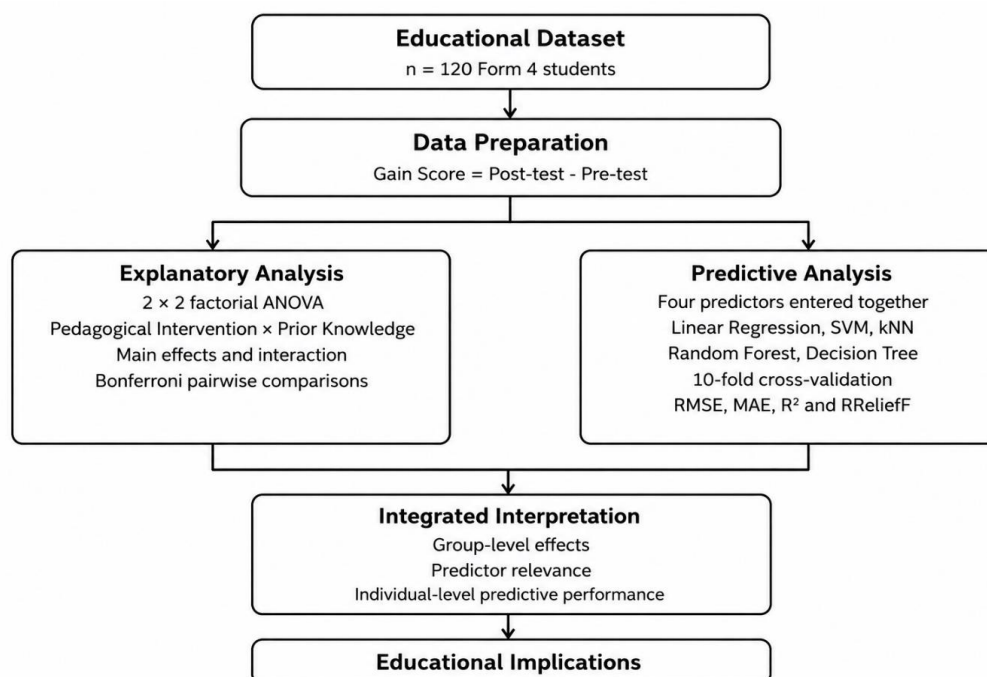


**Figure 2. Explanatory and Predictive Research Design for Learning-Gain Analysis**

### 2.3 Variables and Operationalization of Learning Gain

The outcome variable was Gain Score, calculated as Post-test Score - Pre-test Score and retained as a continuous measure. Four

predictors were used: Previous Grade (ordinal academic history), Prior Knowledge (Low/High), Study Hour (numerical), and Pedagogical Intervention (GIBL/PBL).

***Gain Score = Post-test Score – Pre-test Score***

Pre-test and post-test scores were used only to construct Gain Score and were not entered as

predictors; including post-test performance would create direct target leakage. Previous Grade remained eligible because it was obtained from a separate earlier examination.

**Table 1 Description and Analytical Role of Study Variables**

| Variable                 | Analytical Role      | Information Domain            | Measurement Type     | Operational Description                                                                       |
|--------------------------|----------------------|-------------------------------|----------------------|-----------------------------------------------------------------------------------------------|
| Student ID               | Identifier           | —                             | Nominal              | Unique identifier assigned to each student; excluded from statistical and predictive analyses |
| Previous Grade           | Predictor            | Academic history              | Ordinal categorical  | Grade obtained in a previous examination conducted independently of the intervention pre-test |
| Prior Knowledge          | Predictor            | Learner characteristic        | Categorical          | Students' prior-knowledge level, categorized as Low or High                                   |
| Study Hour               | Predictor            | Learning behaviour            | Numerical            | Amount of time spent studying during the learning period                                      |
| Pedagogical Intervention | Predictor            | Instructional condition       | Categorical          | Instructional approach experienced by the student: GIBL or PBL                                |
| Pre-test Score           | Outcome construction | Baseline performance          | Numerical            | Assessment score obtained before the pedagogical intervention; used to calculate Gain Score   |
| Post-test Score          | Outcome construction | Post-intervention performance | Numerical            | Assessment score obtained after the pedagogical intervention; used to calculate Gain Score    |
| Gain Score               | Target variable      | Learning improvement          | Continuous numerical | Difference between post-test and pre-test scores                                              |

Table 1 summarises the analytical role and operationalisation of the study variables.

## 2.4 Data Preparation and Preliminary Examination

The 120 records were screened for missing values, duplicates, invalid values, inconsistent categories, and potentially influential observations. Categorical variables were represented according to their educational meaning, and any model-specific scaling was

estimated within training folds before being applied to validation data to avoid information leakage.

Descriptive statistics were calculated for the four intervention-prior-knowledge conditions. A 2 x 2 factorial ANOVA then tested the main effects of Pedagogical Intervention and Prior Knowledge and their interaction on Gain Score. Residual diagnostics and homogeneity of variance were examined, partial eta squared was

reported for effect size, and Bonferroni-adjusted pairwise comparisons were used where appropriate.

For prediction, Linear Regression, kNN, SVM, Random Forest, and Decision Tree models used the same four predictors and were evaluated with 10-fold cross-validation. Performance was assessed using RMSE, MAE, and  $R^2$ ; MAPE was not used because Gain Score could equal zero. RReliefF ranked

predictor relevance for the continuous outcome. Relevance scores were interpreted as associations, not causal effects, and were considered alongside overall model performance.

The final interpretation integrated the ANOVA, feature-relevance ranking, and cross-validated regression results to distinguish group-level effects, relative predictor information, and individual-level predictive adequacy.

### 3. Results

#### 3.1 Descriptive Statistics

The sample contained 120 students, with 30 in each of the GIBL-Low, GIBL-High, PBL-Low, and PBL-High conditions (Table 2).

**Table 2 Distribution of Participants by Pedagogical Intervention and Prior Knowledge**

| Pedagogical Intervention | Prior Knowledge | n   |
|--------------------------|-----------------|-----|
| GIBL                     | Low             | 30  |
| GIBL                     | High            | 30  |
| PBL                      | Low             | 30  |
| PBL                      | High            | 30  |
| Total                    |                 | 120 |

Mean Gain Score was highest in GIBL-Low ( $M = 4.70$ ,  $SD = 2.73$ ), followed by GIBL-High ( $M = 4.13$ ,  $SD = 2.47$ ), PBL-Low ( $M = 3.83$ ,  $SD = 2.78$ ), and PBL-High ( $M = 3.13$ ,  $SD = 2.42$ )

(Table 3). These descriptive differences were evaluated inferentially rather than treated as intervention effects.

**Table 3 Descriptive Statistics of Gain Score by Pedagogical Intervention and Prior Knowledge**

| Pedagogical Intervention | Prior Knowledge | n  | Mean Gain Score | SD   |
|--------------------------|-----------------|----|-----------------|------|
| GIBL                     | Low             | 30 | 4.70            | 2.73 |
| GIBL                     | High            | 30 | 4.13            | 2.47 |
| PBL                      | Low             | 30 | 3.83            | 2.78 |
| PBL                      | High            | 30 | 3.13            | 2.42 |

### 3.2 Effects of Pedagogical Intervention and Prior Knowledge

The Pedagogical Intervention x Prior Knowledge interaction was not significant,  $F(1, 116) = 0.02$ ,  $p = .89$ , partial  $\eta^2 = .000$ . Neither main effect reached significance: Pedagogical Intervention,  $F(1, 116) = 3.85$ ,  $p = .052$ , partial  $\eta^2 = .032$ , and Prior Knowledge,  $F(1, 116) = 1.77$ ,  $p = .19$ , partial  $\eta^2 = .015$ . Estimated means were higher for GIBL ( $M = 4.42$ ,  $SE = 0.34$ ) than PBL ( $M = 3.48$ ,  $SE = 0.34$ ) and for Low ( $M = 4.27$ ,  $SE = 0.34$ ) than High Prior

Knowledge ( $M = 3.63$ ,  $SE = 0.34$ ), but these differences were not statistically significant.

Bonferroni-adjusted comparisons were likewise nonsignificant: GIBL vs. PBL differed by 0.87 points for Low Prior Knowledge ( $p = .20$ ) and 1.00 point for High Prior Knowledge ( $p = .14$ ); Low vs. High Prior Knowledge differed by 0.57 points within GIBL ( $p = .40$ ) and 0.70 within PBL ( $p = .30$ ).

### 3.3 Predictor Relevance

RReliefF ranked the four predictors by their relative relevance to Gain Score (Table 4).

**Table 4 RReliefF Ranking of Predictors for Gain Score**

| Rank | Predictor                | RReliefF |
|------|--------------------------|----------|
| 1    | Previous Grade           | 0.521    |
| 2    | Study Hour               | 0.219    |
| 3    | Pedagogical Intervention | 0.217    |
| 4    | Prior Knowledge          | 0.131    |

Previous Grade had the highest relevance value (0.521), followed by Study Hour (0.219), Pedagogical Intervention (0.217), and Prior

Knowledge (0.131). The values indicate relative association within the available predictor set and do not imply causality.

### 3.4 Predictive Model Performance

Five regression models were evaluated using the same four predictors (Table 5).

**Table 5 Predictive Performance of Regression Models for Gain Score**

| Model             | MSE   | RMSE  | MAE   | sMAPE  | R <sup>2</sup> |
|-------------------|-------|-------|-------|--------|----------------|
| Linear Regression | 7.105 | 2.666 | 2.130 | 62.001 | -0.033         |
| SVM               | 7.133 | 2.671 | 2.164 | 62.242 | -0.037         |
| kNN               | 7.581 | 2.753 | 2.212 | 64.435 | -0.102         |
| Random Forest     | 7.626 | 2.762 | 2.292 | 67.472 | -0.108         |
| Decision Tree     | 8.871 | 2.978 | 2.394 | 69.481 | -0.289         |

Linear Regression produced the lowest errors ( $MSE = 7.105$ ,  $RMSE = 2.666$ ,  $MAE = 2.130$ ) and the highest  $R^2$  (-0.033), followed closely by SVM ( $RMSE = 2.671$ ,  $MAE = 2.164$ ,  $R^2 = -$

0.037). kNN, Random Forest, and Decision Tree produced  $RMSE$  values of 2.753, 2.762, and 2.978 and  $R^2$  values of -0.102, -0.108, and -0.289, respectively. Thus, more complex models

did not improve prediction over the simpler linear model.

All  $R^2$  values were negative, indicating performance below the mean-based reference. Consequently, although Linear Regression ranked first among the candidate algorithms, none of the models provided adequate individual-level prediction from the four available predictors. MAPE was excluded because zero Gain Scores make percentage error undefined.

#### 4. Discussion

The study produced a consistent distinction between relative information and predictive adequacy. Descriptive Gain Scores were higher for GIBL and for Low Prior Knowledge, but the factorial effects were not significant. Previous Grade was the most relevant predictor, yet all five cross-validated regression models had negative  $R^2$  values. The findings therefore show that an attribute can be comparatively informative without making the available predictor set sufficient for accurate student-level prediction.

##### 4.1 Pedagogical Intervention, Prior Knowledge, and Learning Gain

GIBL-Low showed the highest mean Gain Score (4.70) and PBL-High the lowest (3.13), but neither Pedagogical Intervention nor Prior Knowledge nor their interaction was statistically significant. The pairwise comparisons were also nonsignificant. This pattern is consistent with evidence that learning improvement reflects multiple learner and instructional influences rather than instructional condition alone (Simonsmeier et al., 2021; Brod, 2021; Bittermann et al., 2023). The broad Low/High prior-knowledge classification may also have obscured variation among students within each category (Schneider & Simonsmeier, 2025).

##### 4.2 Previous Achievement and the Prediction of Learning Gain

Previous Grade had the highest RReliefF value (0.521), consistent with research showing that prior academic performance often contains useful information about later educational

outcomes (Batool et al., 2022; Yağcı, 2022; Molnár & Kocsis, 2023). However, it was not sufficient for predicting learning gain: models that included Previous Grade still failed to achieve a positive  $R^2$ . This demonstrates why feature relevance and predictive performance should be interpreted separately. Previous grades may be useful background information, but they do not necessarily capture how much an individual student will improve during a particular learning experience (Brod, 2021; Bittermann et al., 2023).

##### 4.3 Study Time and Instructional Information

Study Hour ranked second (RReliefF = 0.219), but time spent studying provides only a coarse indication of learning behaviour. Students with similar study duration may differ in concentration, strategy use, persistence, self-regulation, and engagement with learning resources (Neroni et al., 2019; Fong et al., 2021). Pedagogical Intervention showed a similar relevance value (0.217), but this does not establish that GIBL was superior to PBL because the corresponding ANOVA effect was not significant.

Recent studies of technology-enhanced instruction further suggest that a broad intervention label may conceal design features that influence learning processes. In design education, AI-generated virtual instructors enhanced affective engagement, while a conversational teaching style improved academic performance, reduced cognitive load, and supported engagement (Chen, Mokmin, Shen, et al., 2025). In middle school history, a generative AI-powered arts-based learning system similarly improved historical knowledge and motivation while reducing cognitive load (Chen, Mokmin, & Qi, 2025). Although these contexts differ from the present biology intervention, they illustrate why future EDM models may benefit from representing specific instructional features and learner responses rather than using Pedagogical Intervention only as a binary category.

#### 4.4 Predictive Performance and Its Limitations

Linear Regression had the lowest RMSE (2.666) and MAE (2.130), with SVM only marginally behind, but every model produced a negative  $R^2$ . The predictor set therefore contained insufficient information for accurate individual prediction. Likely contributors include the small number of predictors, the difficulty of predicting change scores, and the modest sample size for flexible nonlinear algorithms (López-Zambrano et al., 2021; Batool et al., 2022; Lou & Colvin, 2025). The near-identical performance of Linear Regression and SVM also provides little evidence that added model complexity improved prediction.

#### 4.5 Explanatory and Predictive Evidence

The inferential and predictive analyses addressed different questions. ANOVA evaluated whether group means differed relative to within-group variation, whereas predictive modelling evaluated performance on observations not used to fit a model (López-Zambrano et al., 2021; Lou & Colvin, 2025). Pedagogical Intervention illustrates the distinction: it had a non-zero relevance score but no significant factorial effect. Similarly, Previous Grade ranked highest in RReliefF without producing adequate overall prediction. Reporting both forms of evidence helps avoid interpreting statistical association or relative feature ranking as proof of practical predictive accuracy (Batool et al., 2022; Yağcı, 2022).

#### 4.6 Implications for Educational Research and Data Mining

For EDM research, the central question is not simply which algorithm produces the smallest error, but whether available information supports prediction that is meaningfully better than a simple reference. The present findings suggest that conventional variables such as previous grades, broad prior-knowledge categories, study hours, and intervention labels may need to be supplemented with process-level measures such as engagement, attendance, strategy use, time on task, and interaction patterns.

The importance of contextual and learner-level information is also evident in augmented-reality research with students with dyslexia, where teachers and students identified both opportunities and implementation challenges in technology-supported physical education (Mokmin et al., 2025). For predictive modelling, such evidence supports collecting variables that describe accessibility, interaction quality, learner engagement, and instructional conditions when the goal is to explain differences in learning improvement. Models should also demonstrate generalisation and outperform an appropriate baseline before being used for student-level identification or intervention.

#### 4.7 Directions for Future Research

Future studies should use larger and more diverse samples and collect richer process data, including engagement, motivation, self-regulation, learning strategies, attendance, task completion, and interaction with digital resources. Prior Knowledge could be retained as a continuous measure where appropriate, and Study Hour could be supplemented with indicators of learning quality. Researchers should also compare gain-score prediction with approaches that model post-test performance while accounting for baseline performance.

#### 4.8 Limitations

Several limitations constrain interpretation. The sample of 120 students restricts the information available for training and evaluating machine-learning models, particularly flexible nonlinear approaches. Only four predictors were available, and Prior Knowledge was reduced to Low/High categories, which may have discarded meaningful within-group variation.

Study Hour measured quantity rather than quality of study, and Gain Score may be affected by baseline performance and measurement variability. These limitations do not show that learning gain is inherently unpredictable; rather, they indicate that the variables available in this dataset were insufficient for accurate individual prediction. Future studies should preserve richer

behavioural and continuous measures and evaluate generalisation on independent data.

## 5. Conclusion

This study examined learning gain using complementary inferential and predictive analyses. The ANOVA found no significant effects of Pedagogical Intervention, Prior Knowledge, or their interaction, while RReliefF ranked Previous Grade as the most relevant predictor. Linear Regression produced the lowest error among five models, but all models had negative  $R^2$  values and failed to improve on the mean-based reference. The findings show that feature relevance and relative model ranking do not necessarily translate into adequate individual-level prediction. For EDM applications, richer process-level data and explicit baseline comparison may be more important than increasing algorithmic complexity.

## 6. Acknowledgements

This work was supported by the Ministry of Higher Education Malaysia under the Prototype Research Grant Scheme (Project Code: R503-KR-PRG754-766002412-K134).

## 7. Author Declaration

Originality and conflicts of interest. The authors confirm that this manuscript is original, has not

been published previously, and is not under consideration for publication elsewhere. All authors have read and approved the manuscript. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data availability.** The data supporting the findings of this study are available from the corresponding author upon reasonable request.

**Ethical approval.** This study was conducted in accordance with institutional ethical guidelines. Informed consent was obtained from all participants prior to participation in the study.

**Use of generative AI and AI-assisted technologies.** During the preparation of this manuscript, the authors used generative artificial intelligence tools to refine language and improve grammar. The authors reviewed, edited, and verified all generated content and take full responsibility for the accuracy, integrity, and originality of the manuscript.

This study is funded by Prototype Research Grant Scheme provided by the Ministry of Higher Education Malaysia [R503-KR-PRG754-766002412-K134] and DANA INOVASI DAN KAJIAN SOSIAL [R503-KR-DNS754-766002224-K134].

## References

- Ahmed, E. (2024). Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, 2024, Article 4067721. <https://doi.org/10.1155/2024/4067721>
- Ahmed, W., Wani, M. A., Pławiak, P., Meshoul, S., Mahmoud, A., & Hammad, M. (2025). Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Scientific Reports*, 15. <https://doi.org/10.1038/s4754-766-1754-766>
- Alam, T. (2019). Comparative analysis between pre-test/post-test model and post-test-only model in achieving the learning outcomes. *Pakistan Journal of Ophthalmology*. <https://doi.org/10.36351/pjo.v35i1.855>
- Alhazmi, E., & Sheneamer, A. M. (2023). Early predicting of students performance in higher education. *IEEE Access*, 11, 27579–27589. <https://doi.org/10.1109/ACCESS.2023.3250702>
- Aliyev, R., Backhaus, J., Hammer, S., & König, S. (2026). Comparing the weighted gain score and a Rasch-based approach for estimating learning outcomes in medical education: Quantitative study. *JMIR Medical Education*, 12. <https://doi.org/10.2196/75516>
- Alkan, B. B., Kuzucuk, S., Alkan, N., & Sinan, A. (2025). Using machine learning to predict student outcomes for early intervention and formative assessment. *Scientific Reports*, 15. <https://doi.org/10.1038/s4754-766-23409-w>

- Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H.-Y., & Hussain, A. (2022). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28, 905–971. <https://doi.org/10.1007/s1754-766-11152-y>
- Bittermann, A., McNamara, D., Simonsmeier, B. A., & Schneider, M. (2023). The landscape of research on prior knowledge and learning: A bibliometric analysis. *Educational Psychology Review*, 35, 1–39. <https://doi.org/10.1007/s1754-766-0754-766>
- Brod, G. (2021). Toward an understanding of when prior knowledge helps or hinders learning. *npj Science of Learning*, 6. <https://doi.org/10.1038/s4754-766-00103-w>
- Cents-Boonstra, M., Lichtwarck-Aschoff, A., Denessen, E., Aelterman, N., & Haerens, L. (2020). Fostering student engagement with motivating teaching: An observation study of teacher and student behaviours. *Research Papers in Education*, 36, 754–779. <https://doi.org/10.1080/02671522.2020.1767184>
- Chen, J., Mokmin, N. A. M., & Qi, S. (2025). Generative AI-powered arts-based learning in middle school history: Impact on achievement, motivation, and cognitive load. *The Journal of Educational Research*, 118(6), 688–700. <https://doi.org/10.1080/00220671.2025.2510395>
- Chen, J., Mokmin, N. A. M., Shen, Q., & Su, H. (2025). Leveraging AI in design education: Exploring virtual instructors and conversational techniques in flipped classroom models. *Education and Information Technologies*, 30, 16441–16461. <https://doi.org/10.1007/s1754-766-13458-z>
- Chicco, D., Warrens, M., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7. <https://doi.org/10.7717/peerj-cs.623>
- Christman, E., Miller, P., & Stewart, J. (2024). Beyond normalized gain: Improved comparison of physics educational outcomes. *Physical Review Physics Education Research*. <https://doi.org/10.1103/PhysRevPhysEducRes.20.010123>
- Direnga, J., Timmermann, D., Kieckhäfer, F., Brose, A., & Kautz, C. (2018). The discriminative learning gain: A two-parameter quantification of the difference in learning success between courses. *Australasian Journal of Engineering Education*, 23, 71–82. <https://doi.org/10.1080/22054952.2018.1553357>
- Diseth, Å., Pallesen, S., Brunborg, G., & Larsen, S. (2010). Academic achievement among first semester undergraduate psychology students: The role of course experience, effort, motives and learning strategies. *Higher Education*, 59, 335–352. <https://doi.org/10.1007/s1754-766-754-766>
- Dong, A., Jong, M., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.591203>
- Erdem, C., Kaya, M., Toptaş, H. T., & Altunbaşak, İ. (2025). Problem-based learning and student outcomes in higher education: A second-order meta-analysis. *Studies in Higher Education*, 51, 950–971. <https://doi.org/10.1080/03075079.2025.2498084>
- Erdriani, D., D., & Hidayanti, A. (2026). The overriding influence of prior academic performance: Why GPA matters more than psychological factors in predicting student achievement. *Journal of Applied Research in Higher Education*. <https://doi.org/10.1108/JARHE-754-766-0780>
- Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, 29, 22125–22145. <https://doi.org/10.1007/s1754-766-1754-766>
- Fong, C. J., Krou, M., Johnston-Ashton, K., Hoff, M. A., Lin, S., & Gonzales, C. (2021). LASSI's great adventure: A meta-analysis of the Learning and Study Strategies Inventory and academic outcomes. *Educational Research Review*. <https://doi.org/10.1016/j.edurev.2021.100407>
- Fromm, Y. M., Rasch, J., & Ifenthaler, D. (2025). Technology-enhanced personalized learning environments based on prior knowledge in continuing education: An experimental study. *Smart Learning Environments*, 12. <https://doi.org/10.1186/s4754-766-00407-z>
- Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research*, 102, 101586. <https://doi.org/10.1016/j.ijer.2020.101586>

- Hidayat, A., Maharani, S., Pratama, D., Ramadiani, R., Sujito, S., Septyawan, A., & Sjuchro, D. (2026). Predicting student academic success using machine learning models: A learning analytics approach in higher education. *Daengku: Journal of Humanities and Social Sciences Innovation*. <https://doi.org/10.35877/454RI.daengku4881>
- Kalita, E., Oyelere, S., Gaftandzhieva, S. N., Rajesh, K., Jagatheesaperumal, S., Mohamed, A., Elbarawy, Y. M., Desuky, A. S., Hussain, S., Çifçi, M., Theodorou, P., Hilčenko, S., Hazarika, J., & Ali, N. T. (2025). Educational data mining: A 10-year review. *Discover Computing*, 28. <https://doi.org/10.1007/s1754-766-09589-z>
- Kim, J., & Sakir, N. A. I. (2020). Enhancing students' learning activity and outcomes via implementation of problem-based learning. *Eurasia Journal of Mathematics, Science and Technology Education*. <https://doi.org/10.29333/ejmste/9344>
- Lin, Y., Chen, H., Xia, W., Lin, F., Wang, Z., & Liu, Y. (2023). A comprehensive survey on deep learning techniques in educational data mining. *Data Science and Engineering*, 10, 564–590. <https://doi.org/10.1007/s4754-766-00303-z>
- López-Zambrano, J., Torralbo, J. L. A., & Romero, C. (2021). Early prediction of student learning performance through data mining: A systematic review. *Psicothema*, 33(3), 456–465. <https://doi.org/10.7334/psicothema2021.62>
- Lou, Y., & Colvin, K. F. (2025). Performance prediction using educational data mining techniques: A comparative study. *Discover Education*, 4. <https://doi.org/10.1007/s4754-766-00502-w>
- Mamun, M. A. A., & Lawrie, G. (2023). Student-content interactions: Exploring behavioural engagement with self-regulated inquiry-based online learning modules. *Smart Learning Environments*, 10, 1–31. <https://doi.org/10.1186/s4754-766-00221-x>
- Manganelli, S., Cavicchiolo, E., Mallia, L., Biasi, V., Lucidi, F., & Alivernini, F. (2019). The interplay between self-determined motivation, self-regulated cognitive strategies, and prior achievement in predicting academic performance. *Educational Psychology*, 39, 470–488. <https://doi.org/10.1080/01443410.2019.1572104>
- Mokmin, N. A. M., Rassy, R. P., & Yie, D. L. (2025). Evaluating augmented reality in physical education for dyslexic students from the perspectives of teachers and students. *Scientific Reports*, 15, 7682. <https://doi.org/10.1038/s4754-766-9754-766>
- Molnár, G., & Kocsis, Á. (2023). Cognitive and non-cognitive predictors of academic success in higher education: A large-scale longitudinal study. *Studies in Higher Education*, 49, 1610–1624. <https://doi.org/10.1080/03075079.2023.2271513>
- Möser, M., Hermkes, R., Filmann, N., Harsch, S.-Y., Rüttermann, S., & Gerhard-Szép, S. (2023). Does prior knowledge affect interaction dynamics and learning achievement in digital problem-based learning? A pilot study. *GMS Journal for Medical Education*, 40. <https://doi.org/10.3205/zma001651>
- Nabizadeh, S., Hajian, S., Sheikhan, Z., & Rafiei, F. (2019). Prediction of academic achievement based on learning strategies and outcome expectations among medical students. *BMC Medical Education*, 19. <https://doi.org/10.1186/s1754-766-754-766>
- Namoun, A., & Alshangiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*. <https://doi.org/10.3390/app11010237>
- Navarrete, J. A. (2018). Addressing old mysteries of gain scores in a pretest-posttest educational setting. *Cognitive Science*.
- Navarrete-Ulloa, J. A. (2024). Learning rates: A correction of gain scores to assess mat learning interventions. *The Journal of Experimental Education*, 93, 649–665. <https://doi.org/10.1080/00220973.2024.2352768>
- Neroni, J., Meijs, C., Gijssels, H., Kirschner, P., & De Groot, R. H. M. (2019). Learning strategies and academic performance in distance education. *Learning and Individual Differences*. <https://doi.org/10.1016/j.lindif.2019.04.007>
- Plevris, V., Solorzano, G., Bakas, N., El, M., & Seghier, A. B. (2022). Investigation of performance metrics in regression analysis and machine learning-based prediction models. 8th European Congress on Computational Methods in Applied Sciences and Engineering. <https://doi.org/10.23967/eccomas.2022.155>

- Reddy, S. M. (2025). Predictive modeling of student academic outcomes using machine learning. *International Journal of Scientific Research in Engineering and Management*. <https://doi.org/10.55041/IJSREM49913>
- Romero, C., & Ventura, S. (2012). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3. <https://doi.org/10.1002/widm.1075>
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10. <https://doi.org/10.1002/widm.1355>
- Schneider, M., & Simonsmeier, B. A. (2025). How does prior knowledge affect learning? A review of 16 mechanisms and a framework for future research. *Learning and Individual Differences*. <https://doi.org/10.1016/j.lindif.2025.102744>
- Sghir, N., Adadi, A., & Lahmer, M. (2022). Recent advances in predictive learning analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, 1–35. <https://doi.org/10.1007/s1754-766-1754-766>
- Shaikhanova, A., Kuznetsov, O., Iklassova, K., Tokkuliyeva, A., & Sugurova, L. (2025). Interpretable predictive modeling for educational equity: A workload-aware decision support system for early identification of at-risk students. *Big Data and Cognitive Computing*, 9, 297. <https://doi.org/10.3390/bdcc9110297>
- Shapiro, A. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, 41, 159–189. <https://doi.org/10.3102/00028312041001159>
- Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2021). Domain specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, 57, 31–54. <https://doi.org/10.1080/00461520.2021.1939700>
- Sirianansopa, K. (2024). Evaluating students' learning achievements using the formative assessment technique: A retrospective study. *BMC Medical Education*, 24. <https://doi.org/10.1186/s1754-766-0754-766>
- Ssemugenyi, F. (2023). Teaching and learning methods compared: A pedagogical evaluation of problem-based learning (PBL) and lecture methods in developing learners' cognitive abilities. *Cogent Education*, 10. <https://doi.org/10.1080/2331186X.2023.2187943>
- Wang, S., & Luo, B. (2024). Academic achievement prediction in higher education through interpretable modeling. *PLOS ONE*, 19. <https://doi.org/10.1371/journal.pone.0309838>
- Westphale, S., Backhaus, J., & König, S. (2021). Quantifying teaching quality in medical education: The impact of learning gain calculation. *Medical Education*, 56, 312–320. <https://doi.org/10.1111/medu.14694>
- Wijnia, L., Noordzij, G., Arends, L. R., Rikers, R. M. J. P., & Loyens, S. M. M. (2024). The effects of problem-based, project-based, and case-based learning on students' motivation: A meta-analysis. *Educational Psychology Review*, 36. <https://doi.org/10.1007/s1754-766-0754-766>
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9. <https://doi.org/10.1186/s4754-766-00192-z>