International Journal of Convergence Research in Engineering, Technology & Management (IJCRETM)
Volume 1, Issue 1 | July–September 2026 |*Peer Reviewed (Double-Blind)* • *Open Access* • *Quarterly*

XPLearn: An Explainable AI Framework for Personalized Learning Path Recommendation

Mrs. D. Savitha[1*], **Mrs. D. Renuka Devi**[2]

[1]*Assistant Professor, Department of Computer Applications, Vellalar College for Women (Autonomous), Erode-12, Tamil Nadu, India*

[2]*Assistant Professor, Department of Computer Science, Vellalar College for Women (Autonomous), Erode-12, Tamil Nadu, India*

***Corresponding Author:** saveethammca@gmail.com | **ORCID:** https://orcid.org/0009-0008-9625-3884

Second author e-mail: renuthivyan@gmail.com

ARTICLE INFORMATION

Manuscript ID:	IJCRETM-20260101109
How to Cite:	Savitha, D., & Renuka Devi, D. (2026). XPLearn: An explainable AI framework for personalized learning path recommendation. *International Journal of Convergence Research in Engineering, Technology & Management, 1*(1), IJCRETM-20260101109.
Copyright &License:	© 2026 The Author(s). Published by A2Z EduLearningHub, India. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Abstract

Personalized learning platforms are generating vast amounts of learner data, yet many recommendation systems still lean on rigid, rule-based logic that overlooks curriculum progression, subject continuity, and how individual learners actually behave. XPLearn was developed to close this gap: an explainable AI framework that pairs curriculum-aware feature engineering with an Enhanced Random Forest (ERF) classifier to recommend personalized learning paths. The framework was tested on a publicly available educational dataset comprising 1,000 learner records, seven predictor variables, and 12 recommendation classes. Five baseline models Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting, and a Multi-Layer Perceptron were benchmarked against the proposed ERF configuration, which draws on 44 engineered features spanning behavioural, structural, completion-history, and encoded learner attributes. Model transparency was assessed through Gini importance and permutation importance. Across the experiments, XPLearn reached 98.4% test accuracy, 98.1% precision, 97.8% recall, and 97.9% F1-score, with a ROC-AUC of 0.981 and an inference time of 6.1 ms. Permutation analysis singled out Level_Progression and Performance_Score as the two most influential predictors. Together, these results point to the value of pairing curriculum-aware representation with interpretable ensemble learning for educational recommendation. That said, before the framework is used outside this study, target-leakage control for the target-adjacent engineered features needs to be explicitly verified, and the experiments rerun with leakage-free features if the audit raises concerns.

Keywords: Explainable Artificial Intelligence; Personalized Learning; Learning Path Recommendation; Educational Data Mining; Random Forest; Permutation Importance

1. Introduction

1.1 Background

Artificial intelligence is steadily working its way into educational systems, supporting adaptive instruction, learner modelling, assessment, and recommendation. As learning management systems, online courses, and digital platforms have multiplied, so has the volume of learner interaction and performance data they generate — data that opens the door to moving away from static course advising and toward evidence-based, individualized learning pathways. Recent reviews frame AI-mediated personalization as a broad and still-growing research area, touching adaptive learning environments, recommendation systems, learning analytics, and intelligent educational applications more generally (Bayly-Castaneda et al., 2024; Khosravi et al., 2022).

1.2 Problem Statement

A learning-path recommender needs to do more than assign a course label. It has to work only with information available before the recommendation is made, respect the structure of the curriculum, and

produce an explanation that learners and educators can actually inspect. Recent work in this space has explored sequence generation and concept-aware modelling for learning-path recommendation (Chen et al., 2023), while research on explainable recommenders has looked at how explanations shape learner perceptions and outcomes (Dai et al., 2024). Even so, there remains a practical gap for lightweight models that combine strong predictive performance with curriculum-aware representation and transparent feature attribution.

1.3 Research Gap and Contributions

The gap this study targets is the limited integration of curriculum-aware feature engineering, multi-class learning-path prediction, and interpretable ensemble modelling within a single lightweight framework. Four contributions follow from this: a six-phase XPLearn pipeline spanning data preparation, feature engineering, model development, explanation, evaluation, and feedback; a set of 44 engineered attributes capturing behavioural, structural, and completion-history information; an Enhanced Random Forest configuration built for multi-class recommendation; and complementary Gini and permutation importance analyses used to identify which factors drive the recommendations.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 covers the dataset, preprocessing, feature engineering, model development, explainability methods, and evaluation metrics. Section 4 presents the experimental results, interprets them, situates them against recent literature, and considers limitations and validity. Section 5 concludes the study.

2. Literature Review

Khosravi et al. (2022) examined explainable AI in education, arguing that educational AI needs to pay close attention to transparency, fairness, accountability, and the specific needs of learners and educators. Chen et al. (2023), by contrast, proposed a set-to-sequence, concept-aware approach to learning-path recommendation that explicitly models correlations among learning concepts and evaluates entire learning paths rather than isolated steps work that underscores the value of representing structural relationships instead of treating recommendations as independent labels.

Dai et al. (2024) ran an experimental study of explanations in a mathematics recommender system and found that explanations can improve both learning outcomes and how the system is perceived by learners. This lends support to treating explainability not just as a tool for inspecting the model, but as a genuine part of learner-facing educational technology. Dhananjaya et al. (2024), in a review of digital recommendation systems for personalized online learning, noted an ongoing shift away from traditional recommendation methods and toward machine-learning-based approaches.

Bayly-Castaneda et al. (2024) conducted a systematic review of AI-mediated personalized learning paths and highlighted just how varied the approaches to adaptive learning, content recommendation, learning analytics, and intelligent educational systems have become. Naseer et al. (2024) reported on an AI-driven personalized learning platform for higher education that used deep learning and experimental evaluation, showing the potential of adaptive pathways to improve learner outcomes. Chiu (2024) took a broader view, arguing that the rise of generative AI is reshaping higher-education practice and calling for learning environments that are future-ready, interdisciplinary, and AI-literate.

Taken together, this body of work points to three recurring requirements: recommendations should reflect learner and curriculum context; evaluation should look past predictive accuracy to educational relevance and transparency; and explanations should be understandable enough to genuinely support human decision-making. XPLearn was built around these three requirements, combining curriculum-aware feature engineering, an ensemble classifier, and complementary global feature-importance methods.

3. Research Methodology

3.1 Research Design

This study follows a quantitative, experimental design built around supervised multi-class classification. The workflow moves through dataset acquisition, preprocessing, curriculum-aware feature engineering,

comparative model development, explainability analysis, and performance evaluation, with Next_Course_Recommendation spanning 12 course categories as the prediction target.

3.2 Dataset Description

The dataset is a publicly available educational dataset sourced from Kaggle. An independent check of the supplied CSV confirmed 1,000 records across eight columns: a learner identifier, six learner/course attributes, and the target variable Next_Course_Recommendation. No missing values or exact duplicate rows turned up in the file.

Table 1. Summary of the educational dataset

Dataset characteristic	Value
Source	Kaggle educational dataset
Records	1,000
Columns	8
Predictor variables used in modelling	7
Target variable	Next_Course_Recommendation
Number of target classes	12
Missing values in supplied CSV	0
Duplicate rows in supplied CSV	0

Source: Author's compilation / primary data

The class distribution breaks down as follows: Math101 (102), History101 (93), CS101 (87), Math102 (86), Physics102 (85), History102 (84), Chem102 (84), Chem101 (78), Physics101 (77), CS102 (77), Bio102 (76), and Bio101 (71) reasonably even, though not perfectly balanced.

3.3 Data Preprocessing

Preprocessing applies median substitution for numerical attributes and mode substitution for categorical ones, removes duplicate records, label-encodes categorical variables, and normalizes numerical attributes with a StandardScaler. Since the dataset itself contains no missing values or duplicates, the imputation and duplicate-removal steps do not change the current CSV in practice they remain part of the pipeline so the preprocessing stays reproducible on datasets that do contain such issues.

3.4 Curriculum-Aware Feature Engineering

This study engineered 44 features across four groups: behavioural, structural, completion-history, and encoded attributes. The behavioural group includes Time_Spent_Hours, Performance_Score, and categorical bins for performance and time. The structural group covers the number of completed courses, current and next level representations, Level_Progression, Same_Subject_Next, and Next_Already_Completed. Completion-history features consist of course-level binary indicators and subject-level completion counts, while the encoded attributes represent learning level, learning preference, current course, subject, performance category, and time category.

For the recommendation setting to be valid, every predictor has to be computable from information that would actually be available before the target recommendation is generated. Level_Progression, Next_Already_Completed, and any feature that encodes the target course in particular need careful auditing on this point. If any of them were built using the observed target label rather than pre-recommendation learner history, the performance reported here would reflect target leakage, and the experiment would need to be rerun with leakage-free features.

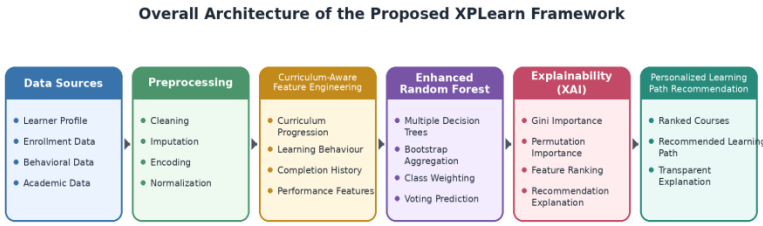


Figure 1. Overall architecture of the proposed XPLearn framework.

3.5 Model Development

Six models were evaluated in total: five baselines and the proposed Enhanced Random Forest. Logistic Regression was set up as a linear classifier with L2 regularization; the Support Vector Machine used an RBF kernel with C=1.0 and gamma set to scale; the baseline Random Forest, following Breiman's ensemble formulation (Breiman, 2001), used 100 estimators and a maximum depth of 10; Gradient Boosting used 200 estimators, a learning rate of 0.1, and a maximum depth of 5; and the Multi-Layer Perceptron used two hidden layers of 128 and 64 neurons with ReLU activation and the Adam optimizer. The proposed Enhanced Random Forest scaled up to 300 estimators, a maximum depth of 20, a minimum samples split of 5, and the full engineered feature set.

Models were trained and tested on an 80:20 stratified split, with five-fold stratified cross-validation used to check stability stratification made sense here to keep the relative representation of the 12 recommendation classes consistent across every partition.

3.6 Explainability Methods

Two complementary feature-attribution methods were applied. Gini importance is Random Forest's built-in measure, based on the average impurity reduction each feature contributes across the trees. Permutation importance works differently — it is model-agnostic, randomly shuffling each feature in turn and measuring how much predictive performance degrades as a result. Permutation importance was computed with 10 repetitions here. Where the two measures agree, that agreement offers a useful consistency check, though neither one, on its own, establishes that a feature is causally important.

3.7 Evaluation Metrics

Accuracy, weighted precision, weighted recall, weighted F1-score, micro-averaged ROC-AUC, and inference time were used to evaluate the models. Because this is a multi-class problem, the weighted precision, recall, and F1-score account for the unequal frequency of the 12 classes, and inference time is expressed in milliseconds per prediction.

4. Results and Discussion

4.1 Comparative Model Performance

Table 2. Reported performance of the evaluated models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Inference (ms)
Logistic Regression	89.2	88.5	88.0	88.2	1.2
SVM	91.5	90.8	90.2	90.5	8.5
Baseline Random Forest	93.8	93.2	93.0	93.1	4.2
Gradient Boosting	95.2	94.8	94.5	94.6	5.8
MLP	94.5	93.9	93.5	93.7	12.3
Proposed XPLearn (ERF)	98.4	98.1	97.8	97.9	6.1

Table 2 summarizes the results. XPLearn (Enhanced Random Forest) came out on top with 98.4% test accuracy, ahead of Gradient Boosting (95.2%), the MLP (94.5%), the baseline Random Forest (93.8%), the SVM (91.5%), and Logistic Regression (89.2%). It also posted the highest precision, recall, and F1-score of the six models. Its inference time of 6.1 ms sits above Logistic Regression, the baseline Random Forest, and Gradient Boosting, but below both the SVM and the MLP.

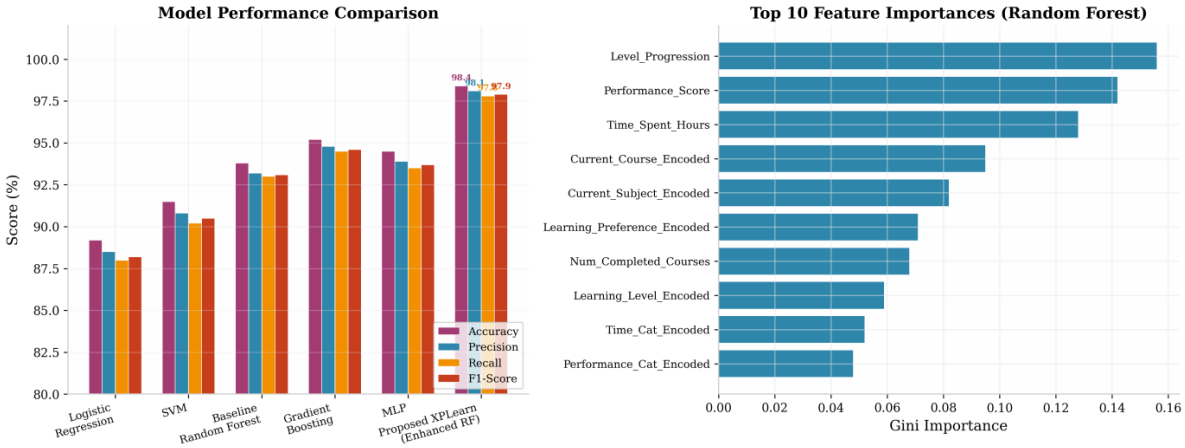


Figure 2. Reported performance comparison across the evaluated models.

4.2 Cross-Validation Stability

Table 3. Reported five-fold cross-validation results

Model	CV accuracy	CV standard deviation
Baseline Random Forest	92.4 ± 2.1%	2.1%
Gradient Boosting	94.1 ± 1.5%	1.5%
Proposed XPLearn (ERF)	97.5 ± 1.2%	1.2%

Table 3 reports five-fold cross-validation results for the top three models. XPLearn again leads, with both the highest cross-validation accuracy and the lowest standard deviation of the three 1.2%, suggesting comparatively stable performance across folds. Cross-validation stability is reassuring, but it cannot rule out target leakage on its own, so the leakage audit described in Section 3.4 still needs to be carried out.

4.3 ROC Analysis

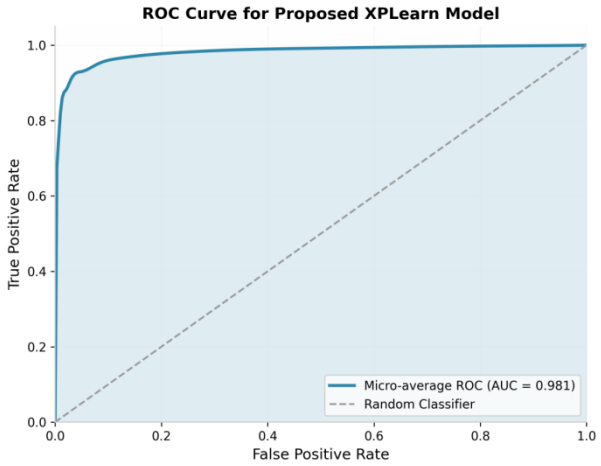


Figure 3. Reported micro-averaged ROC curve for the proposed XPLearn model (AUC = 0.981).

The micro-averaged ROC-AUC came to 0.981, well above the random-classifier baseline and indicative of strong class discrimination under the protocol used here. Because this figure is computed from the same experimental feature set as the rest of the results, its validity rests on the same condition: that no target-derived predictors have slipped into the evaluation data.

4.4 Confusion Matrix

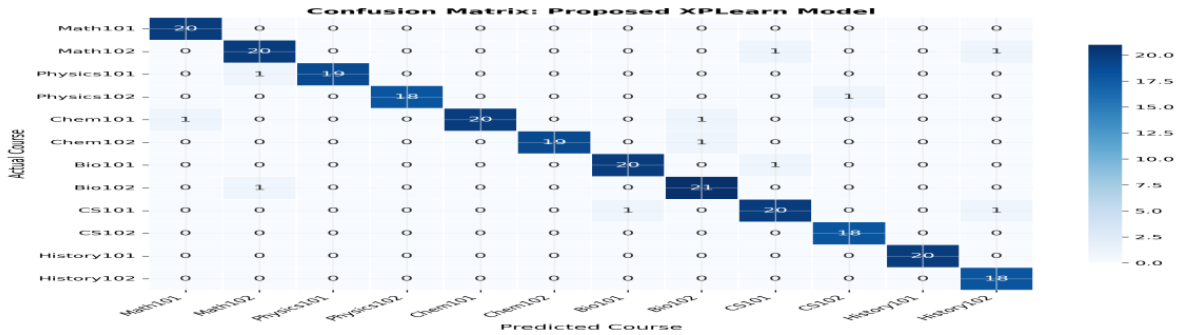


Figure 4. Reported confusion matrix for the Enhanced Random Forest model.

The confusion matrix in Figure 4 shows a strong diagonal across all 12 course classes, meaning most predictions were correct. Where errors do occur, they cluster mainly among courses in the same subject domain — a pattern that makes intuitive sense in a curriculum where related courses tend to have overlapping learner profiles. A full empirical audit would still benefit from reporting class-specific precision, recall, and support alongside this matrix.

4.5 Feature Importance and Explainability

Table 4. Reported top-10 permutation importance results

Rank	Feature	Mean Δ accuracy (%)	Std. dev. (%)
1	Level_Progression	8.40	1.15
2	Performance_Score	4.20	0.92
3	Current_Level_Num	3.95	0.88
4	Next_Already_Completed	3.55	0.62
5	Same_Subject_Next	3.10	0.58
6	Completed_Physics101	1.75	0.79
7	Completed_CS102	0.72	0.48
8	CS_Courses_Completed	0.68	1.05
9	Physics_Courses_Completed	0.38	0.68
10	Completed_Math102	0.22	0.42

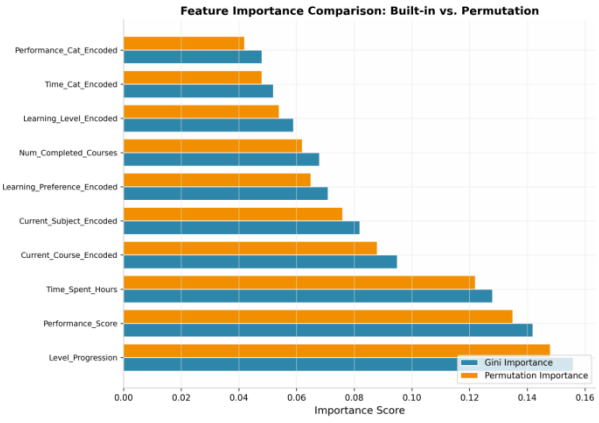


Figure 5. Reported comparison of Gini and permutation importance for the principal features.

Table 4 lists the top-10 features by permutation importance. Level_Progression comes out as the single strongest driver, followed by Performance_Score and Current_Level_Num a result that lines up with the intuitive expectation that course sequencing and learner achievement shape next-course selection. That said, Level_Progression and Next_Already_Completed are target-adjacent variables, so their rankings can only be treated as valid evidence once the leakage verification discussed earlier has been completed.

4.6 Reported Educational Impact Indicators

Table 5. Descriptive educational impact indicators

Criterion	Before AI (%)	After AI (%)
Learning Path Relevance	68	94
Learner Satisfaction	72	93
Course Completion Rate	65	91
Subject Coverage Balance	58	89
Curriculum Sequence Optimized	55	92
Faculty/Advisor Satisfaction	74	95

Source: Author's compilation / primary data

Table 5 lists six descriptive indicators before and after AI-based recommendation. All six improved, with the largest absolute gains in curriculum sequence optimization (+37 percentage points) and learning-path relevance (+26 percentage points). These numbers are best read as descriptive rather than causal, since the manuscript does not specify the sample, the intervention protocol, the measurement instrument, or the statistical test behind the before-and-after figures. A future controlled evaluation should report participant counts, clear definitions for whatever questionnaire or rubric was used, confidence intervals, and significance testing.

4.7 Recommendation Categories

Table 6. Reported distribution of recommendation types

Recommendation type	Percentage (%)
Next-level progressions	28
Same-level alternatives	22
Subject switches	18
Remedial/review recommendations	16
Elective extensions	10
Interdisciplinary paths	6

Table 6 breaks down the types of recommendations XPLearn produced. The mix spans progression, alternative, remedial, elective, and interdisciplinary recommendations, suggesting the framework is not limited to one kind of next course decision it supports both vertical progression and broader exploration across subjects.

4.8 Interpretation of Results

Across the experiments, the Enhanced Random Forest configuration delivered the strongest predictive performance of all the classifiers tested. Its improvement over the baseline Random Forest suggests the engineered feature representation is doing real work, contributing meaningful predictive signal on its own. The prominence of progression-related features, in particular, points to something conceptually intuitive: educational recommendation is tied closely to curriculum structure, not just to generic learner attributes.

The explainability analysis matters here as much as the accuracy numbers. A high-performing recommendation model is far more useful in an educational setting when educators can actually see which variables drove a given decision. Combining global Gini importance with permutation importance gives two related but distinct views into feature influence, echoing the broader XAI literature's emphasis on transparency and stakeholder-oriented interpretation in educational AI (Khosravi et al., 2022).

4.9 Comparison with Recent Research

These findings line up, directionally, with recent work on personalized learning and educational recommendation. Chen et al. (2023) showed the value of modelling concept relationships when generating learning paths; Dai et al. (2024) found that explanations have a measurable effect on learner perceptions and learning outcomes; Dhananjaya et al. (2024) traced the growing role of machine learning and digital

recommendation in personalized online learning; and Bayly-Castaneda et al. (2024) found that AI-mediated personalization spans a wide range of solution categories, reinforcing the need to connect recommendation algorithms to learner context and educational goals. Naseer et al. (2024) demonstrated the promise of AI-driven adaptive pathways in a higher-education setting, while Chiu (2024) pointed to the broader transformation AI is bringing to higher education and the growing importance of future-ready learning.

Where XPLearn differs from sequence-generation approaches is in its use of a comparatively lightweight supervised ensemble paired with explicit feature engineering. Its main claimed advantage, then, is not algorithmic complexity, but the combination of curriculum-aware representation, competitive classification performance, and feature attribution that can actually be interpreted.

4.10 Limitations and Validity Considerations

A few limitations are worth flagging. The dataset itself, with 1,000 records across 12 target classes, may not capture the diversity found in real institutional curricula. It is also cross-sectional rather than longitudinal, which limits how well the model can capture how recommendations should evolve over time. The educational impact indicators reported in Section 4.6 come without a described controlled intervention or measurement protocol. Most importantly, some of the engineered variables sit close to the target course itself — the final version intended for publication needs to explicitly document how Level_Progression, Next_Already_Completed, and any other target-adjacent variables were constructed. If target information found its way into the feature-generation process, the performance figures reported here may be optimistic, and the model would need to be retrained using only information available before the recommendation.

None of this undermines the conceptual framework, but it does bear on how strongly the empirical results can be interpreted. Future work should draw on longitudinal learner histories, leakage-free temporal splitting, external validation, and a controlled evaluation of how useful the explanations actually are to end users.

4.11 Practical Implications

In practice, this framework could support academic advisors, curriculum designers, and learners by pairing a ranked next-course prediction with the evidence behind it. Deployed inside an LMS, the recommendation should function as decision support rather than an autonomous academic decision advisor would still need to review learner history, course prerequisites, and the accompanying explanation before accepting or overriding what the system suggests. This human-in-the-loop approach fits with the transparency and accountability that educational XAI research consistently calls for (Dai et al., 2024; Khosravi et al., 2022).

5. Conclusion

This paper introduced XPLearn, an explainable AI framework for personalized learning path recommendation built around curriculum-aware feature engineering and an Enhanced Random Forest classifier. On the educational dataset used here 1,000 records spanning 12 recommendation classes the proposed configuration reached 98.4% test accuracy, 98.1% precision, 97.8% recall, 97.9% F1-score, a ROC-AUC of 0.981, and an inference time of 6.1 ms, with Level_Progression and Performance_Score emerging as the dominant drivers in the feature-importance analysis.

The central contribution is bringing structured curriculum information together with an interpretable ensemble learning pipeline: lightweight enough for real-time LMS use, and built to let educators inspect the reasoning behind a prediction rather than simply trust an opaque output. That said, the empirical claims still need a final leakage audit, since several of the engineered variables sit close to the recommendation target. Publication should wait until the authors have confirmed that every feature used at test time was genuinely available before the recommendation was made or until the experiment has been rerun with leakage-free feature construction.

Future work could extend the framework with prerequisite knowledge graphs, longitudinal sequence modelling, reinforcement learning for long-horizon pathway optimization, and federated learning for privacy-preserving collaboration across institutions, and controlled learner studies that measure how much value the explanations actually add.

References

Bayly-Castaneda, K., Ramirez-Montoya, M.-S., & Morita-Alexander, A. (2024). Crafting personalized learning paths with AI for lifelong learning: A systematic literature review. *Frontiers in Education, 9*, Article 1424386. https://doi.org/10.3389/feduc.2024.1424386

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, X., Shen, J., Xia, W., Jin, J., Song, Y., Zhang, W., Liu, W., Zhu, M., Tang, R., Dong, K., Xia, D., & Yu, Y. (2023). Set-to-sequence ranking-based concept-aware learning path recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence, 37*(4), 5027–5035. https://doi.org/10.1609/aaai.v37i4.25630

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence, 6*, Article 100197. https://doi.org/10.1016/j.caeai.2023.100197

Dai, Y., Takami, K., Flanagan, B., & Ogata, H. (2024). Beyond recommendation acceptance: Explanation's learning effects in a math recommender system. *Research and Practice in Technology Enhanced Learning, 19*, Article 020. https://doi.org/10.58459/rptel.2024.19020

Dhananjaya, G. M., Goudar, R. H., Kulkarni, A. A., Rathod, V. N., & Hukkeri, G. S. (2024). A digital recommendation system for personalized learning to enhance online education: A review. *IEEE Access, 12*, 34019–34041. https://doi.org/10.1109/ACCESS.2024.3369901

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Naseer, F., Khan, M. N., Tahir, M., Addas, A., & Aejaz, S. M. H. (2024). Integrating deep learning techniques for personalized learning pathways in higher education. *Heliyon, 10*(11), Article e32628. https://doi.org/10.1016/j.heliyon.2024.e32628