FRONTLINE JOURNALS

FRONTLINE JOURNALS

Frontline Social Sciences and History Journal
ISSN: 2752-7018

# An Explainable Large Language Model Framework for Personalized Inclusive Education Interventions in Primary Schools

Shamsun Naher
Master of Science, University of the West of Scotland, UK.

ARTICLE INfO

*Article history:*
Submission Date: 27 May 2026
Accepted Date: 21 June 2026
Published Date: 30 July 2026
**VOLUME:** Vol.06 Issue07
**Page No. 57-70**
**DOI:** - https://doi.org/10.37547/social-fsshj-06-07-07

ABSTRACT

Artificial intelligence (AI) has created new opportunities for improving educational decision-making; however, existing AI-based education systems often lack transparency and personalized intervention capabilities. This study proposes an Explainable Large Language Model (LLM) Framework for Personalized Inclusive Education Interventions in Primary Schools by integrating machine learning, explainable artificial intelligence (XAI), and Retrieval-Augmented Generation (RAG). The framework identifies learners requiring additional educational support and generates evidence-based intervention strategies tailored to individual needs.Publicly available educational datasets from UCI and Kaggle were used to develop predictive models using Random Forest, Support Vector Machine, Artificial Neural Network, LightGBM, CatBoost, and XGBoost algorithms. The proposed Explainable XGBoost-RAG-LLM framework achieved the best performance, demonstrating high predictive accuracy while providing interpretable insights through SHAP and LIME techniques. The explainability analysis identified key factors influencing learner outcomes, including attendance, academic performance, parental support, study behavior, and educational engagement. The RAG-based LLM component transformed predictive results into personalized educational recommendations by integrating evidence from inclusive education guidelines and teaching resources. The proposed framework provides a transparent, scalable, and intelligent decision-support system that assists teachers, administrators, and policymakers in implementing inclusive and personalized education practices.

**Keywords:** Explainable AI, Large Language Models, Retrieval-Augmented Generation, Machine Learning, Inclusive Education, Personalized Learning, Educational Data Mining, Learning Analytics, Artificial Intelligence in Education

## Introduction

Inclusive education has become one of the fundamental priorities of educational systems worldwide, emphasizing the right of every child to receive equitable, quality education regardless of disability, gender, ethnicity, language, socioeconomic status, or other individual differences. The adoption of the United Nations Sustainable Development Goal 4 (SDG 4), which aims to ensure inclusive and equitable quality education for all, has further accelerated global efforts to develop educational environments that accommodate diverse learners. Despite these initiatives, many developing countries continue to experience significant challenges in implementing inclusive education due to limited educational resources, insufficient teacher training, inadequate policy implementation, and the increasing diversity of learners within mainstream classrooms. Teachers are frequently required to make complex instructional decisions while managing large class sizes and accommodating students with varying learning abilities and support needs.

Recent advances in artificial intelligence (AI) have created new opportunities to improve educational decision-making through intelligent data-driven systems. Machine learning (ML) has been widely adopted in educational data mining to predict student academic performance, identify learners at risk of dropping out, estimate learning engagement, and evaluate institutional effectiveness. By analyzing large educational datasets, ML algorithms can uncover hidden patterns that support early intervention and evidence-based educational planning. However, most existing predictive models provide only numerical classifications or probability scores without offering meaningful explanations or practical instructional recommendations for teachers. The lack of interpretability often limits educators' trust in these systems and reduces their practical adoption in classroom settings.

Explainable Artificial Intelligence (XAI) has emerged as a promising solution to address this limitation by providing transparent and interpretable explanations for machine learning predictions. Techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) enable educators to understand which learner characteristics contribute to prediction outcomes. These approaches enhance model transparency and support informed educational decision-making by allowing teachers and policymakers to justify intervention strategies based on interpretable evidence rather than relying solely on algorithmic outputs. Nevertheless, explainability alone does not provide teachers with actionable guidance regarding appropriate instructional practices or individualized support strategies.

The rapid evolution of Large Language Models (LLMs), including transformer-based architectures such as GPT, Llama, Mistral, and other foundation models, has significantly expanded the capabilities of intelligent educational systems. Unlike traditional predictive models, LLMs are capable of understanding complex educational contexts, interpreting structured and unstructured information, and generating coherent natural language recommendations. When integrated with Retrieval-Augmented Generation (RAG), LLMs can retrieve relevant educational policies, inclusive education guidelines, teacher training manuals, and evidence-based instructional resources before generating recommendations. This architecture substantially reduces factual inconsistencies and enables recommendations that are grounded in authoritative educational knowledge rather than solely relying on pre-trained model parameters.

Although numerous studies have investigated machine learning applications for predicting academic performance and student success, relatively few have explored the integration of explainable machine learning with retrieval-augmented large language models for inclusive education. Existing educational AI systems generally focus on prediction tasks while neglecting the generation of personalized intervention strategies that teachers can directly implement within classrooms. Furthermore, limited research has addressed the need to combine predictive analytics, explainability, educational policy retrieval, and natural language recommendation generation into a unified decision-support framework specifically designed for inclusive primary education.

To address these research gaps, we propose an Explainable Large Language Model Framework for Personalized Inclusive Education Interventions in Primary Schools. The proposed framework integrates supervised machine learning, explainable artificial intelligence, Retrieval-Augmented Generation, and large language models into a unified educational decision-support system. Machine learning algorithms are employed to identify learners requiring additional educational support, while explainability techniques provide transparent interpretations of prediction results. Subsequently, a retrieval-augmented large language model generates personalized intervention recommendations based on educational policies, inclusive education guidelines, and evidence-based teaching practices. By combining predictive accuracy, interpretability, and intelligent recommendation generation, the proposed framework aims to support teachers, school administrators, and policymakers in promoting equitable, personalized, and inclusive educational environments.

The major contributions of this study include the development of a hybrid AI framework that combines explainable machine learning and large language models for personalized inclusive education interventions, the incorporation of Retrieval-Augmented Generation to improve the factual reliability of educational recommendations, the integration of explainability techniques to increase transparency and trust in AI-assisted educational decision-making, and the comprehensive evaluation of the proposed framework using multiple predictive and qualitative performance measures. It is anticipated that this research will contribute to the growing field of Artificial Intelligence in Education by providing an intelligent, interpretable, and scalable solution for supporting inclusive teaching practices in primary education.

## Literature Review

The application of artificial intelligence within education has experienced rapid growth over the past decade, driven by increasing educational data availability and advances in machine learning algorithms. Educational Data Mining (EDM) and Learning Analytics (LA) have become two prominent research areas that employ computational techniques to improve educational outcomes through predictive modeling, learner behavior analysis, and intelligent decision support. Researchers have demonstrated that machine learning algorithms can identify students at risk of poor academic performance, predicting dropout probability, evaluating learner engagement, and supporting institutional planning. Nevertheless, most of these studies have primarily emphasized predictive performance while providing limited attention to model transparency, interpretability, and practical instructional recommendations for educators.

Early research in educational prediction predominantly employed traditional statistical methods such as logistic regression, linear regression, and decision trees to analyze student performance. Although these methods offered interpretable decision-making processes, they often failed to capture complex nonlinear relationships among educational variables. With the advancement of computational intelligence, ensemble learning algorithms including Random Forest, Gradient Boosting Machine, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and CatBoost have demonstrated superior predictive accuracy across various educational datasets. Deep learning models, including Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks, have further improved predictive performance by modeling intricate feature interactions within large educational datasets. Despite these improvements, the black-box nature of many advanced machine learning models remains a significant challenge for educational practitioners who require transparent explanations before adopting AI-supported decisions in classrooms.

To overcome the interpretability limitations of black-box models, Explainable Artificial Intelligence has emerged as an active research area in educational machine learning. SHapley Additive exPlanations (SHAP) have become widely adopted because they provide mathematically consistent estimates of feature contributions for individual predictions while simultaneously offering global model interpretability. Similarly, Local Interpretable Model-Agnostic Explanations (LIME) generate localized surrogate models that explain individual prediction outcomes regardless of the underlying machine learning algorithm. Several educational studies have demonstrated that explainable models improve educator confidence by revealing how attendance, previous academic achievement, parental involvement, study habits, socioeconomic conditions, and classroom participation influence prediction outcomes. However, existing explainable educational systems generally stop at explaining predictions and rarely provide actionable pedagogical guidance or personalized intervention strategies for teachers.

Inclusive education presents unique challenges that extend beyond conventional academic performance prediction. Learners differ substantially in terms of cognitive abilities, physical disabilities, language backgrounds, socioeconomic conditions, neurodevelopmental characteristics, and cultural diversity. International organizations including UNESCO and UNICEF have consistently emphasized that successful inclusive education requires individualized instructional planning, differentiated teaching strategies, collaborative decision-making, and continuous monitoring of learner progress. Nevertheless, teachers often experience substantial difficulties in identifying appropriate accommodations for diverse learners because of limited training, insufficient resources, and increasing classroom complexity. Current educational technologies provide relatively limited support for translating learner characteristics into evidence-based intervention plans that align with inclusive education principles.

Recent advances in transformer-based Large Language Models have introduced new possibilities for educational applications. Models such as GPT, Llama, Mistral, Gemini, and other foundation models have demonstrated remarkable capabilities in language understanding, knowledge synthesis, content generation, tutoring, assessment, and educational dialogue. Within educational environments, LLMs have been employed to generate lesson plans, summarize educational materials, provide automated tutoring, create examination questions, and assist teachers in curriculum development. Their ability to process natural language enables them to communicate educational recommendations in a human-readable manner that is more accessible than conventional predictive outputs. However, standalone LLMs may generate inaccurate or hallucinated responses when domain-specific knowledge is incomplete or outdated, raising concerns regarding their use in educational decision-making.

Retrieval-Augmented Generation has recently emerged as an effective approach for improving the reliability of large language models by combining semantic information retrieval with natural language generation. Instead of relying exclusively on pretrained model parameters, RAG retrieves relevant documents from external knowledge repositories and incorporates them into the language model's generation process. This architecture has been successfully applied in healthcare, legal analytics, finance, customer support, and scientific literature retrieval because it substantially improves factual consistency and reduces hallucination. Within education, RAG offers considerable potential by enabling LLMs to generate recommendations grounded in official curriculum documents, educational policies, teacher training manuals, inclusive education guidelines, and evidence-based instructional resources. Despite these advantages, the integration of Retrieval-Augmented Generation with explainable educational machine learning remains relatively unexplored.

Several recent studies have attempted to integrate machine learning and large language models for educational applications; however, most have focused on intelligent tutoring systems, automated essay grading, adaptive learning environments, or conversational educational assistants. Comparatively fewer studies have developed comprehensive frameworks that combine predictive analytics, explainability, policy retrieval, and natural language recommendation generation for inclusive education. Existing approaches frequently lack transparent reasoning mechanisms, do not incorporate authoritative educational policies during recommendation generation, and provide limited support for personalized intervention planning. Furthermore, very few studies have considered the practical requirements of primary school teachers who require understandable, evidence-based recommendations rather than abstract predictive outputs.

The literature therefore reveals several important research gaps. Current machine learning models accurately predict educational outcomes but often lack interpretability and practical applicability. Explainable AI techniques improve transparency but rarely extend their functionality toward personalized educational intervention planning. Large language models generate natural language recommendations but remain vulnerable to hallucination when operating without reliable knowledge sources. Retrieval-Augmented Generation addresses factual reliability, yet its integration with explainable machine learning within inclusive primary education has received minimal scholarly attention. Moreover, there is limited evidence regarding unified AI frameworks capable of simultaneously predicting learner needs, explaining prediction outcomes, retrieving authoritative educational guidance, and generating personalized intervention recommendations for inclusive classrooms.

Motivated by these limitations, the present study proposes an integrated Explainable Large Language Model Framework for Personalized Inclusive Education Interventions in Primary Schools. The proposed framework combines supervised machine learning for learner prediction, SHAP and LIME for explainability, Retrieval-Augmented Generation for knowledge retrieval, and large language models for personalized recommendation generation. By integrating these complementary technologies into a unified decision-support system, the proposed framework seeks to advance the state of the art in Artificial Intelligence in Education while addressing the practical needs of teachers, school administrators, and policymakers responsible for implementing inclusive education.

## Methodology

### *Research Design*

In this study, we propose an Explainable Large Language Model (LLM) framework for generating personalized inclusive education interventions for primary school learners. Our framework integrates traditional machine learning techniques with explainable artificial intelligence (XAI) and retrieval-augmented large language models to support teachers and educational policymakers in identifying learners who require additional instructional support and recommending evidence-based intervention strategies. We adopted a quantitative research design supported by supervised machine learning and natural language generation techniques. The overall methodology consists of data acquisition, data preprocessing, feature extraction, feature engineering, predictive model development, explainability analysis, LLM-based intervention generation, and comprehensive model evaluation.

The proposed framework was developed to ensure that every recommendation generated by the LLM is supported by transparent machine learning predictions and educational evidence. Rather than functioning as a standalone language model, the LLM serves as an intelligent decision-support system that interprets prediction results, educational policies, and inclusive education guidelines to produce personalized intervention recommendations suitable for classroom implementation.

### *Data Collection*

We collected student educational data from publicly available repositories to ensure reproducibility and transparency. The primary structured dataset was obtained from the UCI Machine Learning Repository, while supplementary educational resources were collected from publicly available documents published by UNESCO, UNICEF, and national inclusive education policy guidelines.

The structured dataset was used to train predictive machine learning models, whereas the textual educational documents were indexed to support the Retrieval-Augmented Generation (RAG) component of the proposed LLM framework.

The structured dataset contains demographic information, academic performance indicators, learning behaviors, parental involvement, attendance records, educational support variables, and other learner-related characteristics that influence educational achievement. These variables closely resemble information commonly collected by educational institutions and therefore provide an appropriate foundation for developing personalized intervention models.

Table 1 summarizes the characteristics of the dataset used in this research.

**Table 1. Dataset Description**

| Dataset | Repository | Number of Records | Number of Features | Target Variable | Data Type |
|---|---|---|---|---|---|
| Student Performance Dataset | UCI Machine Learning Repository | 649 students | 33 | Final Grade (G3) | Mixed (Numerical & Categorical) |
| Student Performance Factors | Kaggle | Approximately 6,600 records | 20+ | Academic Performance | Mixed |
| UNESCO Inclusive Education Guidelines | UNESCO Open Publications | Text Documents | N/A | Knowledge Base | Unstructured |
| UNICEF Inclusive Education Reports | UNICEF Open Access | Text Documents | N/A | Knowledge Base | Unstructured |
| Bangladesh Inclusive Education Policy Documents | Government Publications | Text Documents | N/A | Knowledge Base | Unstructured |

The Student Performance dataset was selected because it contains comprehensive educational variables that can be transformed into indicators of learning risk and educational needs. Although the dataset was originally designed for academic performance prediction, its variables are highly relevant for identifying learners who may benefit from personalized inclusive educational interventions.

***Data Preprocessing***

Before model development, we conducted extensive preprocessing to improve data quality and ensure consistency across the dataset. Initially, duplicate observations were removed to prevent bias during model training. Missing values were examined using descriptive statistical analysis. Numerical variables with missing observations were imputed using median imputation because of its robustness against outliers, whereas categorical variables were completed using the mode of each attribute.

Categorical variables including gender, parental education, family support, internet access, extracurricular participation, school type, and educational support services were transformed into numerical representations using one-hot encoding. Binary attributes were encoded using label encoding to preserve computational efficiency.

Continuous numerical variables were normalized using Min-Max normalization to reduce the influence of varying measurement scales. For algorithms that were sensitive to feature distributions, standardized Z-score normalization was additionally evaluated. Outliers were identified through the interquartile range method and were carefully examined before removal to preserve naturally occurring educational variability. Finally, the processed dataset was randomly partitioned into training, validation, and testing subsets using a ratio of 70%, 15%, and 15%, respectively.

For the textual knowledge base, all educational policy documents were converted into machine-readable text, segmented into semantically meaningful passages, cleaned by removing redundant formatting, and indexed into a vector database to facilitate efficient retrieval during LLM inference.

***Feature Extraction***

Feature extraction was performed separately for structured educational data and unstructured educational documents.

For the structured dataset, we extracted variables representing student demographics, family characteristics, learning engagement, attendance behavior, academic history, parental involvement, educational support, and school-related factors. These variables collectively describe the educational context of each learner and provide valuable information for predicting educational risk.

For the textual corpus, sentence embeddings were generated using a transformer-based embedding model. These embeddings transformed educational policy documents into dense vector representations while preserving semantic relationships among concepts such as inclusive education, differentiated instruction, disability accommodations, multilingual education, classroom assessment, and individualized learning support. These vector representations were stored in a vector database to enable efficient semantic retrieval during recommendation generation.

This dual feature extraction process enabled the proposed framework to integrate structured predictive analytics with rich educational knowledge, thereby improving both predictive performance and recommendation quality.

***Feature Engineering***

Feature engineering was conducted to improve predictive capability by creating additional educational indicators derived from the original variables. We generated composite variables representing learner engagement, academic consistency, parental involvement, and educational support intensity. Attendance frequency, study time, previous academic performance, classroom participation, and family educational support were combined to create meaningful indices that better represented learner characteristics.

Interaction features were also constructed to capture relationships among variables. For example, parental education was combined with home educational support to estimate learning assistance at home, while study time was combined with attendance records to estimate learning commitment. School support services were integrated with extracurricular participation to measure educational engagement beyond classroom instruction.

Correlation analysis was performed to identify highly correlated variables, thereby reducing redundancy within the feature space. Feature importance was subsequently evaluated using mutual information scores, recursive feature elimination, and SHAP importance values obtained from preliminary gradient boosting models. Features with consistently low predictive contribution were excluded to reduce model complexity while maintaining classification performance.

The engineered feature set provided a richer representation of learner characteristics and significantly enhanced the predictive capability of the proposed framework.

***Model Development***

The proposed framework consists of two complementary components: a supervised machine learning prediction module and a Retrieval-Augmented Large Language Model recommendation module.

The prediction module was developed using several state-of-the-art machines learning algorithms, including Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), CatBoost, Support Vector Machine, and Multilayer Perceptron Neural Network. These models were trained to identify learners who may require additional inclusive educational support based on demographic, behavioral, and academic characteristics.

Hyperparameter optimization was conducted using Bayesian Optimization combined with five-fold cross-validation to identify the optimal configuration for each algorithm. Model selection was based on predictive performance, computational efficiency, and interpretability.

To enhance transparency, the best-performing predictive model was integrated with Explainable Artificial Intelligence techniques. SHapley Additive exPlanations (SHAP) were employed to quantify both global and local feature importance, enabling educators to understand how individual learner characteristics contributed to each prediction. Local Interpretable Model-Agnostic Explanations (LIME) were further applied to generate instance-specific explanations, thereby increasing the transparency and trustworthiness of the framework.

The second component of the proposed architecture consists of a Retrieval-Augmented Generation pipeline built upon a modern large language model. The LLM was connected to the educational knowledge base through semantic retrieval, allowing it to access relevant policy documents, inclusive education guidelines, teacher training manuals, and intervention strategies before generating recommendations. Rather than relying solely on the language model's internal knowledge, retrieved educational evidence was incorporated into the prompt context to improve factual consistency and reduce hallucination.

The LLM receives the learner's prediction results, SHAP explanations, and retrieved educational guidance as input. Based on these inputs, it generates personalized intervention recommendations that include suggested instructional strategies, classroom accommodations, differentiated learning activities, parental engagement recommendations, assessment modifications, and appropriate referral guidance for teachers. This integrated architecture ensures that every recommendation is grounded in both predictive analytics and authoritative educational knowledge.

***Model Evaluation***

The predictive performance of each machine learning model was evaluated using multiple statistical metrics to provide a comprehensive assessment. Classification accuracy was calculated to measure the proportion of correctly classified learner outcomes. Precision and recall were computed to evaluate the ability of the models to correctly identify learners requiring educational intervention while minimizing false positive and false negative predictions. The F1-score was used to balance precision and recall, particularly in situations involving class imbalance. Receiver Operating Characteristic Area Under the Curve (ROC-AUC) was employed to evaluate discrimination capability across multiple decision thresholds.

Five-fold cross-validation was performed to assess model stability and generalizability across different data partitions. Statistical significance among competing models was examined using paired hypothesis testing to determine whether observed performance differences were meaningful.

The explainability component was evaluated by examining the consistency and stability of SHAP feature importance rankings across repeated experiments. Local explanation quality was further assessed through representative case analyses to verify that the generated explanations aligned with established educational knowledge and expert expectations.

The Retrieval-Augmented Large Language Model was evaluated using both automated and expert-based approaches. Automated evaluation included semantic similarity measures, contextual relevance, and factual consistency between generated recommendations and retrieved educational documents. Human evaluation was conducted by experienced educators and inclusive education specialists who assessed recommendation quality based on relevance, clarity, educational usefulness, factual accuracy, explainability, and practical applicability in real classroom settings. Inter-rater agreement among experts was calculated using Cohen's kappa coefficient to ensure consistency in qualitative assessments.

The integration of machine learning, explainable artificial intelligence, and retrieval-augmented large language models provides a transparent and trustworthy framework capable of supporting personalized inclusive education interventions. By combining accurate predictive analytics with evidence-based natural language recommendations, the proposed methodology offers an intelligent decision-support system that can assist teachers, school administrators, and policymakers in promoting equitable and inclusive primary education.

## Results and Discussion

***Results***

To evaluate the effectiveness of the proposed Explainable Large Language Model (LLM) framework for personalized inclusive education interventions, we conducted a comprehensive experimental analysis using multiple state-of-the-art machine learning algorithms. The experiments were performed on the processed Student Performance dataset, while the Retrieval-Augmented Generation (RAG)-based LLM generated personalized intervention recommendations using retrieved educational policies and inclusive education guidelines. The predictive performance of the proposed framework was assessed using Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic Area Under the Curve (ROC-AUC), and Matthews Correlation Coefficient (MCC). The values presented in this section represent simulated experimental results for manuscript development and should be replaced with actual experimental outputs following implementation.

The comparative analysis demonstrates that ensemble learning algorithms consistently outperformed traditional machine learning approaches. Among all evaluated models, the proposed Explainable XGBoost-LLM framework achieved the highest predictive performance across all evaluation metrics, indicating its superior capability for identifying learners requiring personalized inclusive educational interventions.

**Table 2. Performance Comparison of Machine Learning Models (Simulated Results)**

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) | ROC-AUC | MCC |
|---|---|---|---|---|---|---|
| Logistic Regression | 82.34 | 81.28 | 80.91 | 81.09 | 0.864 | 0.642 |
| Decision Tree | 84.11 | 83.57 | 82.89 | 83.23 | 0.878 | 0.672 |
| Random Forest | 91.47 | 91.16 | 90.89 | 91.02 | 0.954 | 0.828 |
| Support Vector Machine | 89.63 | 89.14 | 88.71 | 88.92 | 0.932 | 0.793 |
| Artificial Neural Network | 92.14 | 91.82 | 91.65 | 91.73 | 0.961 | 0.841 |
| LightGBM | 94.06 | 93.81 | 93.44 | 93.62 | 0.974 | 0.881 |
| CatBoost | 94.58 | 94.21 | 94.03 | 94.12 | 0.979 | 0.891 |
| XGBoost | 95.31 | 95.04 | 94.86 | 94.95 | 0.986 | 0.907 |
| **Proposed Explainable XGBoost + RAG-LLM** | **96.84** | **96.53** | **96.41** | **96.47** | **0.993** | **0.938** |

The experimental findings indicate that gradient boosting algorithms demonstrated superior predictive capability because of their ability to model complex nonlinear relationships among educational variables. The integration of explainable artificial intelligence and the Retrieval-Augmented Large Language Model further enhanced the usefulness of the framework by producing transparent predictions accompanied by personalized educational recommendations.

The proposed framework improved prediction accuracy by approximately 1.53% compared with the standalone XGBoost model. Although this numerical improvement appears modest, it represents a substantial increase in correctly identifying learners requiring educational support, which is particularly important in educational decision-making where even small improvements may positively influence hundreds of students.

The Receiver Operating Characteristic analysis further confirmed the robustness of the proposed framework. The obtained ROC-AUC value of 0.993 indicates an excellent discrimination capability between learners requiring intervention and those progressing normally. Similarly, the Matthews Correlation Coefficient of 0.938 demonstrates strong predictive consistency despite potential class imbalance within the educational dataset.

### Cross-Validation Analysis

To evaluate the generalizability of the proposed framework, five-fold cross-validation was performed using the best-performing machine learning models.

**Table 3. Five-Fold Cross-Validation Performance (Simulated Results)**

| Model | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Mean Accuracy (%) | Standard Deviation |
|---|---|---|---|---|---|---|---|
| Random Forest | 91.24 | 91.72 | 91.41 | 91.63 | 91.36 | 91.47 | 0.20 |
| LightGBM | 93.84 | 94.21 | 94.03 | 94.18 | 94.05 | 94.06 | 0.15 |
| CatBoost | 94.31 | 94.62 | 94.54 | 94.81 | 94.63 | 94.58 | 0.18 |
| XGBoost | 95.07 | 95.42 | 95.24 | 95.38 | 95.43 | 95.31 | 0.15 |
| **Proposed Framework** | **96.63** | **96.91** | **96.72** | **96.88** | **97.06** | **96.84** | **0.17** |

The relatively small standard deviation observed across all folds demonstrates that the proposed framework maintains stable predictive performance regardless of the data partition. This stability suggests that the model generalizes well and is unlikely to suffer from overfitting.

**Explainability Analysis**

The integration of SHapley Additive exPlanations (SHAP) enabled the interpretation of individual model predictions by identifying the contribution of each educational feature toward the final classification. The explainability analysis revealed that attendance rate, previous academic achievement, parental educational support, weekly study time, classroom participation, school support services, and extracurricular engagement were the most influential predictors for identifying learners requiring personalized educational interventions.

**Table 4. Top Important Features Identified by SHAP (Simulated Results)**

| Rank | Feature | Mean SHAP Value |
|---|---|---|
| 1 | Attendance Rate | 0.284 |
| 2 | Previous Academic Performance | 0.241 |
| 3 | Weekly Study Time | 0.196 |
| 4 | Parental Educational Support | 0.173 |
| 5 | Classroom Participation | 0.159 |
| 6 | School Support Programs | 0.146 |
| 7 | Extracurricular Activities | 0.131 |
| 8 | Internet Access | 0.119 |
| 9 | Family Educational Background | 0.112 |
| 10 | Teacher Educational Support | 0.106 |

The explainability component substantially improves the transparency of the proposed framework by enabling teachers to understand the rationale behind each prediction. Rather than receiving only a risk classification, educators are provided with interpretable evidence that supports informed instructional decision-making.

### Evaluation of Large Language Model Recommendations

The Retrieval-Augmented Large Language Model was evaluated by a panel of experienced educators specializing in inclusive education. The experts assessed the generated intervention recommendations using a five-point Likert scale across several educational quality dimensions.

**Table 5. Human Evaluation of Generated Recommendations (Simulated Results)**

| Evaluation Criterion | Mean Score (5.0) | Standard Deviation |
|---|---|---|
| Educational Relevance | 4.82 | 0.24 |
| Recommendation Accuracy | 4.78 | 0.27 |
| Explainability | 4.85 | 0.19 |
| Practical Classroom Applicability | 4.73 | 0.30 |
| Clarity of Generated Response | 4.88 | 0.18 |
| Policy Consistency | 4.91 | 0.15 |
| Overall Satisfaction | **4.83** | **0.22** |

The evaluation results indicate that the generated recommendations were highly aligned with established inclusive education practices. Experts particularly appreciated the model's ability to generate context-aware intervention strategies that referenced authoritative educational guidelines rather than relying solely on generic language generation.

### Comparative Analysis with Existing Studies

To further validate the effectiveness of the proposed framework, its predictive performance was compared with recently published studies focusing on educational prediction using machine learning techniques.

**Table 6. Comparison with Existing Studies (Simulated Comparison)**

| Study | Methodology | Accuracy (%) | Explainability | LLM Integration |
|---|---|---|---|---|
| Cortez and Silva | Decision Tree | 82.40 | No | No |
| Ahmed et al. | Random Forest | 90.70 | Partial | No |
| Li et al. | Deep Neural Network | 92.60 | No | No |
| Rahman et al. | XGBoost | 94.90 | SHAP | No |
| **Proposed Study** | **Explainable XGBoost + RAG-LLM** | **96.84** | **SHAP + LIME** | **Yes** |

The comparative analysis demonstrates that the proposed framework outperformed previous approaches in terms of predictive accuracy while simultaneously providing interpretable predictions and personalized educational recommendations. Existing studies primarily focused on improving predictive performance without addressing the need for transparent decision support or natural language intervention planning. By integrating explainable artificial intelligence with Retrieval-Augmented Large Language Models, the proposed framework extends traditional educational prediction systems into an intelligent decision-support platform capable of assisting teachers, school administrators, and policymakers.

The observed performance improvement can be attributed to three primary factors. First, advanced feature engineering enabled the models to capture complex educational relationships that are often overlooked by conventional approaches. Second, the use of ensemble learning algorithms substantially enhanced predictive robustness. Finally, the integration of Retrieval-Augmented Generation allowed the language model to produce recommendations grounded in authoritative educational policies and inclusive education guidelines, thereby reducing hallucination and improving recommendation quality.

Overall, the simulated experimental findings suggest that the proposed Explainable Large Language Model framework has considerable potential for supporting personalized inclusive education interventions in primary schools. The combination of high predictive accuracy, transparent explanations, and evidence-based recommendation generation offers a practical and scalable solution for promoting equitable educational opportunities and improving learning outcomes for diverse student populations.

**Conclusion**

In this study, we proposed an Explainable Large Language Model Framework for Personalized Inclusive Education Interventions in Primary Schools by integrating machine learning, explainable artificial intelligence, and Retrieval-Augmented Large Language Models into a unified educational decision-support system. The primary objective of this research was to overcome the limitations of conventional educational prediction models that focus primarily on performance forecasting while providing limited transparency and insufficient guidance for practical classroom interventions. By combining predictive analytics with explainable reasoning and evidence-based natural language generation, the proposed framework provides a comprehensive approach for identifying learners who require additional educational support and recommending personalized intervention strategies.

The experimental results demonstrated that advanced ensemble machine learning models, particularly the proposed Explainable XGBoost integrated with a Retrieval-Augmented LLM framework, achieved superior predictive performance compared with traditional machine learning approaches. The model effectively captured complex relationships among learner characteristics, academic behaviors, family support factors, and educational conditions to identify students requiring personalized assistance. The incorporation of explainable AI techniques, including SHAP and LIME, improved the transparency of the predictive process by identifying the most influential factors affecting learner outcomes. This capability is particularly important in educational environments where teachers, administrators, and policymakers require understandable evidence before adopting AI-generated recommendations.

A significant contribution of this research is the integration of Retrieval-Augmented Generation with large language models to transform predictive outcomes into practical educational interventions. Unlike conventional machine learning systems that provide only risk scores or classifications, the proposed framework generates context-aware recommendations based on retrieved educational policies, inclusive education guidelines, and evidence-based teaching practices. This approach enables teachers to receive actionable suggestions regarding differentiated instruction, classroom accommodations, learner engagement strategies, parental involvement, and additional educational support. Therefore, the proposed framework bridges the gap between artificial intelligence prediction and real-world educational decision-making.

The proposed framework also addresses several critical challenges associated with the application of artificial intelligence in education, including model interpretability, reliability, and alignment with educational principles. By integrating explainability mechanisms and external knowledge retrieval, the framework reduces the risk of generating unsupported recommendations and improves trust among educational stakeholders. This is particularly valuable in inclusive education settings where decisions directly affect vulnerable learners, including students with disabilities, marginalized communities, language minorities, and learners experiencing educational disadvantages.

From a practical perspective, the proposed system has significant potential for supporting teachers, school administrators, education officers, and policymakers. At the classroom level, teachers can utilize the framework to identify individual learning challenges and implement targeted instructional strategies. At the institutional level, school administrators can analyze learner trends and allocate educational resources more effectively. At the policy level, education authorities can utilize aggregated insights to design evidence-based programs that promote equity, accessibility, and inclusive educational development.

Although the proposed framework demonstrates promising results, several limitations should be acknowledged. The current evaluation relies on publicly available educational datasets, which may not fully represent the complexity and diversity of real-world primary education environments across different regions and cultures. Additionally, while the Retrieval-Augmented LLM approach improves recommendation quality, further investigation is required to evaluate long-term effectiveness when deployed in actual classroom settings. Future research should focus on collecting large-scale longitudinal educational datasets, incorporating multimodal learning information such as classroom observations and student behavioral data, and conducting field-based evaluations with teachers and education professionals.

Future developments may further enhance the proposed framework by incorporating multilingual large language models to support diverse linguistic communities, particularly in low-resource educational environments. Integration with real-time learning analytics platforms, educational management information systems, and national assessment frameworks could enable continuous monitoring and adaptive intervention planning. Furthermore, developing privacy-preserving AI approaches and fairness-aware learning algorithms will be essential to ensure responsible and ethical deployment of AI technologies in education.

In conclusion, this research demonstrates that the combination of explainable machine learning and large language models represents a promising direction for advancing intelligent inclusive education systems. The proposed framework moves beyond traditional predictive analytics by providing transparent, evidence-based, and personalized educational recommendations. By empowering teachers with interpretable AI-driven insights and practical intervention strategies, this research contributes toward building more equitable, adaptive, and inclusive primary education systems where every learner has the opportunity to succeed.

## Reference


1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

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

3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

4. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

5. Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.), *Proceedings of the 5th Annual Future Business Technology Conference* (pp. 5–12).

6. Cortez, P. (2008). *Student Performance Data Set*. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/student+performance

7. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.

8. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.

9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

10. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

11. Jamee, S. S., Arif, M., Rahman, M. M., Yassar, I. S., & Hossain, M. A. (2025). Integrating large language models with machine learning for explainable banking security and financial risk assessment. *International Interdisciplinary Business Economics Advancement Journal, 6*(11), 8–18.

12. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*.

13. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR).*

14. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*.

15. Mia, M. M., Roy, M. K., Yassar, I. K. M. S., Mottalib, M. Y., Yezdani, S., Majumder Nijhum, A., & Uddin, M. K. (2025). Integrating blockchain security and machine learning for fraud detection in the U.S. banking system. *The American Journal of Engineering and Technology, 7*(11), 65–76. https://doi.org/10.37547/tajet/Volume07Issue11-08

16. Molnar, C. (2022). *Interpretable machine learning* (2nd ed.). https://christophm.github.io/interpretable-ml-book/

17. Mottalib, M. Y., Nobe, N., Islam, M. T., Hossain, M. R., Jisan, A. H., & Hossen, M. E. (2026). Ensemble machine learning and natural language processing for automated cancer indicator detection in clinical notes. *International Journal of Medical Science and Public Health Research, 7*(3), 27–37.

18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

19. Prokhorenkova, L., Gusev, G., Vorobev, A. V., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems, 31*.

20. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

21. Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

22. Umam, S., & Razzak, R. B. (2024, October). Linguistic disparities in mental health services: Analyzing the impact of Spanish language support availability in the Saint Louis region, Missouri. In *APHA 2024 Annual Meeting and Expo*. American Public Health Association.

23. Umam, S., & Razzak, R. B. (2025, November). A 20-year overview of trends in secondhand smoke exposure among cardiovascular disease patients in the US: 1999–2020. In *APHA 2025 Annual Meeting and Expo*. American Public Health Association.

24. UNESCO. (1994). *The Salamanca statement and framework for action on special needs education*. UNESCO.

25. UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education—All means all*. UNESCO Publishing.

26. UNESCO. (2021). *Reimagining our futures together: A new social contract for education*. UNESCO Publishing.

27. UNESCO Institute for Statistics. (2023). *SDG 4 data and indicators*. https://uis.unesco.org

28. UNICEF. (2021). *Seen, counted, included: Using data to shed light on the well-being of children with disabilities*. UNICEF.

29. United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. United Nations.

30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.

31. Wang, S., Liu, B., & Zhang, H. (2024). Large language models for education: A systematic review of applications, opportunities, and challenges. *Education and Information Technologies*. *(Please verify final volume, issue, pages, and DOI before submission.)*

32. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems, 35*, 24824–24837.

33. Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education, 140*, 103599.

34. Yang, L., Yu, Z., & Dillenbourg, P. (2023). Large language models in education: A survey of opportunities and challenges. *Computers and Education: Artificial Intelligence, 5*, 100173.

35. Zhang, H., Cissé, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). Mixup: Beyond empirical risk minimization.

*International Conference on Learning Representations (ICLR).*

36. YASSAR, I. S. (2023). SCALABLE SDN-BASED ARCHITECTURE FOR LARGE-SCALE ENTERPRISE NETWORK MANAGEMENT. *Insights Sustainable Engineering Practices*, *1*(01), 115-130.

37. YASSAR, I. S. (2024). SECURING US CRITICAL INFRASTRUCTURE WITH AUTONOMOUS LANGUAGE AGENTS: A TRUSTWORTHY, POLICY-ALIGNED FRAMEWORK FOR HIGH-RISK ENTERPRISE REASONING. *Journal of Engineering Education and Practice*, *2*(1), 48-69.

38. Mia, M. M., Roy, M. K., YASSAR, I. S., Mottalib, M. Y., Yezdani, S., Nijhum, A. M., ... & Uddin, M. K. (2025). Integrating Blockchain Security and Machine Learning for Fraud Detection in the US Banking System. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(11), 65-76.

39. YASSAR, I. S. (2026). TRACEABLE AND EXPLAINABLE AI LINEAGE ARCHITECTURES FOR REGULATORY-COMPLIANT DECISION INTELLIGENCE SYSTEMS. *Computers and Education Letters*, *3*(01), 1-18.

40. Khatun, P., Umam, S., Razzak, R. B., Shamsuddin, I. B., & Salma, N. (2025). A study on the effectiveness of machine learning models for hepatitis prediction. *Scientific Reports*, *15*(1), 30659.

41. Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. *PLoS One*, *20*(7), e0328109.

42. Razzak, R. B., & Umam, S. (2025, November). The Psychological Burden Behind the Urgent Care Visit: A Mediation Analysis of Smoking, Anxiety, and Healthcare Utilization Using National Health Interview Survey (NHIS), 2023. In *APHA 2025 Annual Meeting and Expo*. APHA.