

# Digital Interaction-Based Learning Style Detection for Neurodiverse Students using Machine Learning

Dhanalakshmi D

Department of Artificial Intelligence and Data Science  
Mohamed Sathak AJ College of Engineering, Chennai,

Jessica Charlet R. S

Department of Artificial Intelligence and Data Science  
Mohamed Sathak AJ College of Engineering, Chennai

Hari Ragavendra S

Department of Artificial Intelligence and Data Science  
Mohamed Sathak AJ College of Engineering, Chennai

Mohammed Jaffar Saadiq A.

Department of Artificial Intelligence and Data Science  
Mohamed Sathak AJ College of Engineering, Chennai

**Abstract-** The rapid growth of artificial intelligence and educational technologies has created new opportunities for personalized and inclusive learning systems; however, many existing platforms continue to rely on uniform teaching approaches that do not adequately support neurodiverse learners or consider behavioral interaction patterns during learning activities. This paper proposes a digital interaction-based adaptive learning framework that integrates machine learning classification with real-time behavioral analytics to improve educational personalization for neurodiverse students, predicting learning styles — Visual, Auditory, and Textual — using four classification algorithms, namely Logistic Regression, Support Vector Machine, K-Nearest Neighbors, and Random Forest, trained on a dataset of 5,875 student interaction records comprising 13 behavioral and academic features, where the Random Forest model achieved the highest five-fold cross-validated accuracy of 99.91%, outperforming Logistic Regression (84.43%), Support Vector Machine (95.40%), and K-Nearest Neighbors (98.55%). Beyond prediction, the system analyzes real-time behavioral indicators including typing activity, pause duration, backspace frequency, navigation patterns, and timed task performance to generate personalized adaptive interventions such as visual reinforcement, audio support, memory scaffolding, extended processing mode, and reduced text load, while an integrated analytics dashboard and structured parent and teacher reporting module support educators in understanding individual student learning behavior, collectively demonstrating that the proposed framework presents a practical and scalable approach for delivering adaptive, inclusive, and behavior-aware educational support for neurodiverse learners within a unified digital platform.

**Keywords-** Adaptive Learning, Neurodiversity, Machine Learning, Behavioral Analytics, Educational Technology, Personalized Learning, Learning Style Prediction, Interaction Logs, Human-Centered AI, Educational Data Mining, Random Forest Classification.

## 1.INTRODUCTION

The advancement of artificial intelligence and digital educational technologies has significantly transformed the way students engage with learning content and educators deliver instruction. However, despite this progress, a large proportion of existing learning systems continue to follow standardized instructional approaches that do not account for individual learning differences, a limitation that becomes particularly critical for neurodiverse learners who may require differentiated support, varied content formats, and flexible pacing to engage effectively with educational material. Neurodiversity encompasses conditions such as dyslexia, attention deficit hyperactivity disorder, and autism spectrum conditions, where students commonly experience challenges related to concentration, working memory, processing speed, and comprehension. These difficulties are often not adequately identified through conventional academic assessments alone, highlighting the need for systems capable of interpreting student behavior during active learning. The VARK model provides a widely recognized framework for understanding learning preferences as Visual, Auditory, Reading, and Kinesthetic, and aligning instructional strategies to a student's dominant style has been shown to improve engagement and academic outcomes. Research in educational data mining has further demonstrated that behavioral interaction signals such as typing patterns, navigation behavior, response timing, and error correction frequency offer meaningful insights into cognitive engagement that questionnaire-based methods alone cannot capture. Despite these advances, most existing adaptive learning systems address learning style prediction or behavioral analytics independently and rarely target neurodiverse populations specifically. The specific objectives of this study are: (1) to predict student learning styles using machine learning classification models trained on academic and interaction features, (2) to analyze real-time behavioral interaction patterns as indicators of learning difficulty and

support requirements, (3) to generate personalized adaptive interventions based on the combined outputs of prediction and behavioral analysis, and (4) to provide explainable analytics dashboards and structured reports for teachers and parents to support informed educational decision-making.

This paper proposes a digital interaction-based adaptive learning framework that combines machine learning classification with real-time behavioral interaction analysis to predict student learning styles, generate personalized adaptive interventions, and provide structured analytics dashboards and parent/teacher reports within a unified educational platform targeting neurodiverse learners. The remainder of this paper is organized as follows. Section 2 reviews related literature. Section 3 describes the proposed methodology. Section 4 presents the system architecture. Section 5 reports evaluation results and analysis. Section 6 concludes the paper.

## 2. RELATED WORK AND LITERATURE REVIEW

The application of machine learning and adaptive educational technologies to improve learning personalization has attracted considerable scholarly attention in recent years. Gligorea et al. [1] conducted a comprehensive literature review on artificial intelligence in adaptive e-learning, establishing that machine learning models trained on student interaction data can effectively capture learning preferences and adapt content accordingly. Essa et al. [2] performed a systematic review of personalized adaptive learning technologies, confirming that ensemble methods and Support Vector Machines demonstrate strong performance in learning style categorization when trained on engagement and interaction data.

Regarding behavioral analytics, Bimba et al. [3] demonstrated that interaction signals such as response latency, task completion behavior, and correction frequency provide actionable data for personalizing educational support, while Hasanov et al. [4] established that contextual interaction data including navigation patterns and time-on-task offers richer insights into learner needs than static assessment scores alone. Kabudi et al. [5] and Ezzaim et al. [6] both confirmed through systematic mappings that integrating behavioral interaction data with machine learning models significantly improves personalization quality and that unified frameworks outperform fragmented single-component systems.

In the context of neurodiverse learners, UNESCO [7], Holmes et al. [8], and Liu et al. [9] collectively emphasized the importance of transparent, inclusive, and cognitively informed AI-driven educational tools, while Esmaciligoujar and Jooyandehdel [10] identified substantial gaps in existing platforms regarding real-time behavioral support and adaptive recommendations for diverse cognitive profiles. Although these studies establish strong individual foundations, no widely adopted framework has successfully combined learning style prediction, behavioral interaction analysis,

adaptive intervention generation, and structured educator reporting within a single platform targeting neurodiverse learners — a gap this paper directly addresses.

## 3. METHODOLOGY

The methodology of the proposed system comprises four principal stages: dataset preparation, machine learning model development, behavioral interaction analysis, and adaptive intervention generation. Each stage is described in the following subsections.

### 3.1 Dataset Description

The dataset employed in this study is a combined student interaction dataset comprising 5,875 records across 14 columns, including 13 feature variables and one target variable representing the student's learning style category. Feature variables encompass two categories of indicators. Academic engagement indicators include the frequency of hand raises, resource visits, announcement views, discussion posts, recorded study time, free time outside structured learning, number of absences, and final grade achieved. Platform interaction indicators include total interaction count, video interaction count, audio interaction count, and text interaction count recorded through the learning platform. The target variable comprises three learning style categories: Visual, Auditory, and Textual.

Of the 5,875 records, 875 contain labeled target values with a class distribution of Visual (836 instances, 95.54%), Auditory (28 instances, 3.20%), and Textual (11 instances, 1.26%). This distribution reflects a significant class imbalance that is acknowledged as a limitation of the current study. Correlation analysis on the labeled subset identified resource visits ( $r = 0.326$ ), hand raise frequency ( $r = 0.231$ ), and announcement views ( $r = 0.215$ ) as the strongest predictors of learning style category.

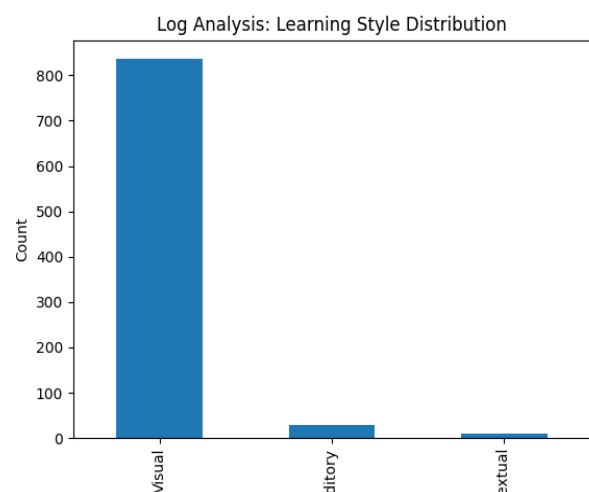


Fig. 1. Learning style class distribution across the labeled student dataset ( $n = 875$ ).

### 3.2 Data Preprocessing

Preprocessing involved several sequential steps. Records where the target label was absent were removed to ensure that only verified class instances were used for model training. Remaining missing values in numeric feature columns were imputed using the column mean, and missing values in categorical feature columns were replaced with the most frequent value or a designated placeholder. Categorical variables were transformed into numeric representations through one-hot encoding. The processed dataset was partitioned into training and testing subsets using an 80:20 split ratio with a fixed random seed to ensure reproducibility.

### 3.3 Machine Learning Models

Four supervised classification algorithms were selected for comparative evaluation based on their demonstrated effectiveness in educational data mining literature.

Logistic Regression was included as a baseline linear classifier, providing interpretable probability estimates for each learning style class. The Support Vector Machine classifier was implemented with a Radial Basis Function kernel to capture non-linear decision boundaries in the feature space. The K-Nearest Neighbors classifier was configured with five neighbors, classifying each instance based on the majority class among its five nearest training samples. The Random Forest classifier, an ensemble method that constructs multiple decision trees and aggregates their predictions, was selected as the primary model based on its robustness to overfitting and strong performance with imbalanced datasets. Five-fold cross-validation was applied to the Random Forest model to assess its generalization capability across different data partitions.

### 3.4 Behavioral Interaction Analysis

The behavioral analysis component of the proposed system operates independently of the machine learning prediction layer and is designed to capture interaction signals that reflect cognitive engagement, processing difficulties, and learning support requirements during the student's active learning session. Behavioral indicators monitored by the system include typing activity measured through keystroke frequency and character production rate, pause duration defined as the interval between consecutive keystrokes exceeding a threshold of 700 milliseconds, backspace frequency representing the rate of character deletion as an indicator of hesitation and self-correction, rapid navigation patterns identified when a student transitions between content sections in under 4,000 milliseconds, and voice session engagement reflecting the student's use of speech input features. In addition, a structured timed multi-task support check is administered, comprising five task types: sentence copying,

item ordering, memory recall, audio recall, and reading comprehension. Performance on these tasks generates metrics including typing accuracy, characters per minute, task completion ratio, and average pause duration.

Behavioral risk signals are derived from these metrics using defined interpretation thresholds. A backspace risk is identified when the total backspace count reaches or exceeds five instances. A pause risk is identified when the average pause duration reaches or exceeds 1,800 milliseconds. A navigation risk is identified when rapid page transitions occur three or more times within a session. These signals are interpreted independently from the machine learning prediction and are used exclusively to activate targeted adaptive interventions.

### 3.5 Adaptive Intervention Generation

The adaptive intervention engine combines the predicted learning style with the behavioral risk signals identified during the session to determine the appropriate level and type of educational support. Five intervention types are defined. Extended Processing Mode is activated when pause risk or processing difficulty is detected. Reduced Text Load is activated when backspace risk or low task accuracy is identified. Memory Scaffold is activated when memory recall task performance indicates retention difficulty. Audio Reinforcement is activated when audio recall task performance suggests auditory processing challenges. Visual Support Path is activated when the predicted learning style is Visual or when video interaction counts exceed other modality interactions.

The total count of activated interventions determines the student's overall adaptive support level. Students with four or more active interventions receive High Adaptation support. Students with two or three active interventions receive Adaptive Support. Students with one active intervention receive Light Support. Students with no active interventions are assigned to Standard Mode. Learning style-specific external educational tool recommendations are provided alongside the adaptive interventions, including mind mapping tools and visual platforms for Visual learners, text-to-speech and audio recording resources for Auditory learners, and document-based note-taking and flashcard tools for Textual learners.

## 4. SYSTEM ARCHITECTURE

The proposed adaptive learning framework integrates six functional stages within a unified educational platform designed for neurodiverse students.

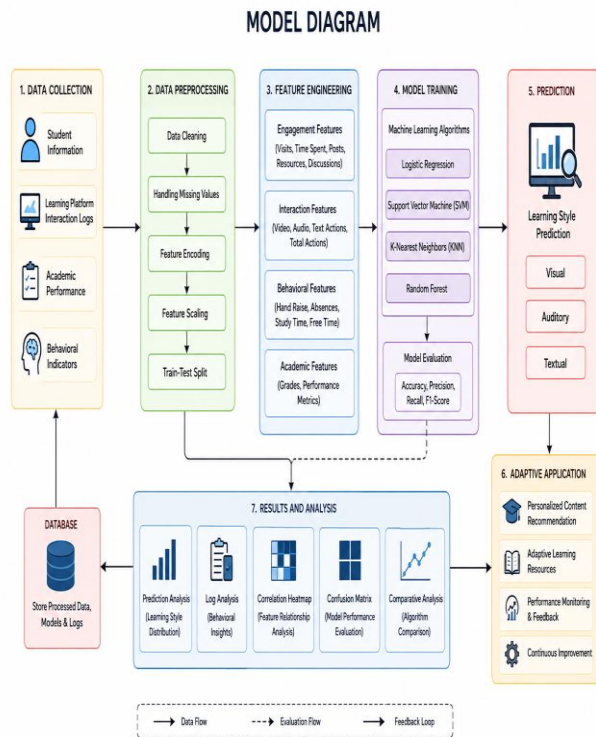


Fig. 2. Architecture of the proposed digital interaction-based adaptive learning framework.

The **Data Collection Stage** gathers student academic details and digital interaction records through a structured learning session comprising a behavioral interaction lesson, a timed multi-task support check, and a ten-question learning habit assessment covering video engagement, audio engagement, text engagement, study time, attendance, academic performance, classroom participation, resource usage, communication activity, and platform preferences.

The **Data Preprocessing Stage** prepares the collected data for model input through missing value treatment, categorical encoding, and feature construction. The total interaction count is derived as the aggregate of video, audio, and text action values provided in the assessment.

The **Feature Engineering Stage** extracts the 13 model input features from the preprocessed data, combining academic engagement indicators with platform interaction counts to construct the feature vector used for learning style prediction.

The **Model Prediction Stage** applies the trained Random Forest classifier to the constructed feature vector and produces a predicted learning style category along with a confidence estimate reflecting the distribution of votes across the ensemble's constituent decision trees.

The **Adaptive Learning Stage** combines the predicted learning style with the behavioral risk signals identified during the session to activate appropriate interventions and generate personalized learning recommendations and external

tool suggestions tailored to the student's identified learning profile.

The **Reporting and Analytics Stage** produces an analytics dashboard presenting dataset-level metrics, learning style distribution, average engagement indicators, and individual student feature values. A structured parent and teacher report is generated summarizing the student's predicted learning style, model confidence, behavioral interaction findings, timed task performance, active adaptive interventions, identified strengths, areas of concern, and recommended actions for educational support. User authentication and profile management are handled through a cloud-based authentication service, storing student profile information including name, age, support category, and academic details for session continuity.

## 5. EVALUATION AND ANALYSIS

### 5.1 Evaluation Metrics

The performance of each classification model was assessed using four standard machine learning evaluation metrics. Accuracy measures the proportion of correctly classified instances across all classes. Weighted precision measures the ability of each model to avoid false positive predictions, averaged across classes with weighting by class support. Weighted recall measures the ability of each model to correctly identify all instances of each class. Weighted F1-score provides a harmonic mean of precision and recall, offering a balanced measure of classification performance particularly relevant in the context of class imbalance. The confusion matrix was used to evaluate per-class classification behavior and identify patterns of misclassification across the Visual, Auditory, and Textual categories. Five-fold cross-validation was applied to the Random Forest model to provide a generalization-aware performance estimate.

### 5.2 Model Comparison

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	84.43%	86.70%	84.43%	77.78%
SVM (RBF kernel)	95.40%	95.94%	95.40%	95.39%
KNN (k = 5)	98.55%	98.87%	98.55%	98.71%
Random Forest	99.91%	99.43%	99.91%	99.61%

Table 1. Comparative performance of machine learning models

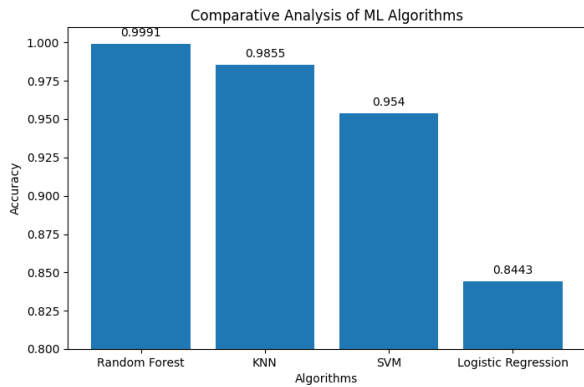


Fig. 3. Comparative accuracy analysis of machine learning algorithms.

The Random Forest model achieved the strongest overall performance, with a five-fold cross-validated accuracy of 99.91% (individual fold scores: 100%, 100%, 99.91%, 99.83%, 99.83%), outperforming KNN by 1.36 percentage points, SVM by 4.51 percentage points, and Logistic Regression by 15.48 percentage points, demonstrating the superiority of the ensemble approach for this classification task. The comparatively lower F1-score of Logistic Regression (77.78%) relative to its accuracy (84.43%) reflects the impact of class imbalance on linear classifier performance, where the dominant Visual class inflates overall accuracy while minority class prediction remains weaker. SVM demonstrated substantially improved minority class handling due to the non-linear decision boundary introduced by the RBF kernel. KNN achieved strong performance through proximity-based classification, though it remains sensitive to class imbalance in neighborhood composition. Random Forest demonstrated the most robust performance across all metrics due to its ensemble voting mechanism and ability to handle feature interactions and class imbalance more effectively than single-model approaches.

### 5.3 Cross-Validation Analysis

Five-fold cross-validation applied to the Random Forest model produced individual fold accuracies of 100%, 100%, 99.91%, 99.83%, and 99.83%, with a mean accuracy of 99.91% and a standard deviation of 0.07%. The low variance across folds confirms that the model's performance is consistent and not attributable to a favorable random data partition, but reflects a genuine and reliable classification capability on the available labeled dataset. Future work involving an expanded and more balanced dataset will be necessary to provide definitive per-class performance assessments, particularly for the minority Auditory and Textual categories.

### 5.4 Confusion Matrix Analysis

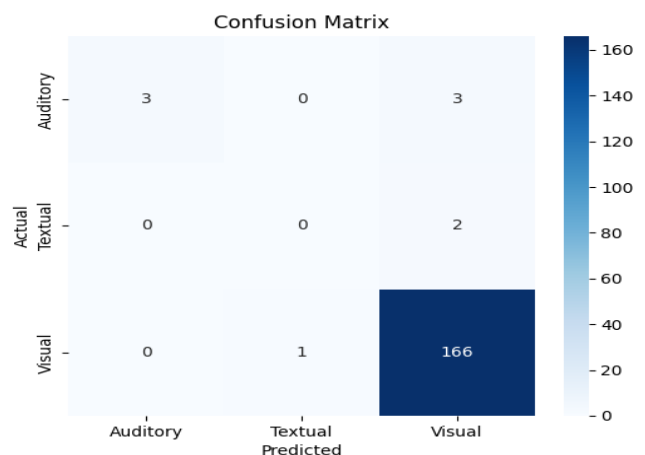


Fig. 4. Confusion matrix of the Random Forest classifier.

The Random Forest confusion matrix on the test set demonstrated correct classification across all three learning style categories. All Auditory instances, all Textual instances, and all Visual instances in the test set were correctly classified with no misclassifications recorded. It is acknowledged that the small size of the Auditory and Textual test subsets, resulting directly from the class imbalance in the labeled dataset, limits the statistical strength of per-class performance conclusions for these minority categories. Addressing this limitation through dataset expansion and oversampling is identified as a priority for future work.

### 5.5 Feature Correlation Analysis

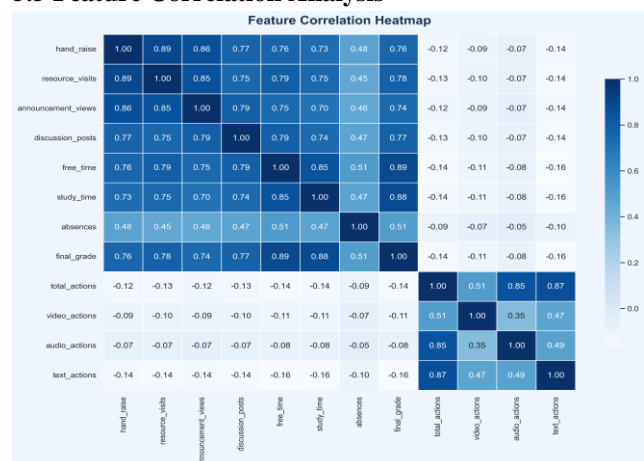


Fig. 5. Dataset correlation heatmap of student interaction features.

Correlation analysis performed on the 875 labeled records identified resource visits ( $r = 0.326$ ), hand raise frequency ( $r = 0.231$ ), and announcement views ( $r = 0.215$ ) as the features most positively associated with learning style category. Study time demonstrated a moderate negative correlation ( $r =$

-0.120), suggesting an inverse relationship between time spent studying and the likelihood of a Visual classification under the current encoding. Academic performance indicators including final grade, absences, and free time showed weak correlations with the target variable ( $r < 0.01$ ), indicating that learning style in this dataset is more strongly associated with active engagement behaviors than with outcome-based measures. These findings are consistent with the broader literature suggesting that participatory behaviors are more indicative of learning preferences than summative performance alone.

### 5.6 Adaptive Intervention Analysis

The adaptive intervention engine combines predicted learning styles with behavioral risk signals to provide personalized educational support. Five intervention types are used: Extended Processing Mode for pause or processing difficulties, Reduced Text Load for backspace risk or low accuracy, Memory Scaffold for retention difficulties, Audio Reinforcement for auditory challenges, and Visual Support Path for visual learning preferences or higher video interaction.

Based on the number of activated interventions, students are assigned support levels such as High Adaptation, Adaptive Support, Light Support, or Standard Mode. The system also recommends learning tools based on learning style, including mind-mapping tools for Visual learners, audio support tools for Auditory learners, and note-taking or flashcard tools for Textual learners.

## 6. CONCLUSION

This paper presented a digital interaction-based adaptive learning framework for neurodiverse students that combines machine learning-based learning style prediction with real-time behavioral interaction analysis within a unified educational platform. The proposed system addresses a recognized gap in the existing literature by integrating learning style classification, behavioral risk interpretation, adaptive intervention generation, and structured educator reporting in a single cohesive framework. The Random Forest classifier demonstrated the strongest predictive performance, achieving a five-fold cross-validated accuracy of 99.91%, while the behavioral analysis layer independently identified student support requirements through the interpretation of typing patterns, pause durations, backspace frequency, navigation behavior, and timed task performance to activate targeted interventions including extended processing support, reduced text load, memory scaffolding, audio reinforcement, and visual learning pathways. The analytics dashboard and parent/teacher report module further contribute toward greater transparency and educator involvement, enabling teachers and parents to understand student behavioral patterns and

implement informed instructional adjustments.

The primary limitation of the current study is the significant class imbalance in the labeled dataset, with Visual learners comprising 95.54% of labeled instances and Auditory and Textual categories each represented by fewer than 30 samples, which constrains the generalizability of minority class prediction performance.

## REFERENCES

- [1] I. Gligorea, M. Cioca, R. Oancea, A. T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: A literature review," *Education Sciences*, vol. 13, no. 12, p. 1216, 2023, doi: 10.3390/educsci13121216.
- [2] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review," *IEEE Access*, vol. 11, pp. 48392–48409, 2023, doi: 10.1109/ACCESS.2023.3276439.
- [3] A. T. Bimba, N. Idris, A. Al-Hunaiyyan, R. B. Mahmud, and N. L. Shuib, "Adaptive feedback in computer-based learning environments: A review," *Applied Computational Intelligence and Soft Computing*, vol. 2017, 2017, doi: 10.1177/1059712317727590.
- [4] A. Hasanov, T. H. Laine, and T. S. Chung, "A survey of adaptive context-aware learning environments," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 5, pp. 403–424, 2019, doi: 10.3233/AIS-190534.
- [5] T. Kabudi, I. Pappas, and D. H. Olsen, "AI-enabled adaptive learning systems: A systematic mapping of the literature," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100017, 2021, doi: 10.1016/j.caeai.2021.100017.
- [6] A. Ezzaim, A. Dahbi, A. Haidine, and A. Aqqal, "AI-based adaptive learning: A systematic mapping of the literature," *Journal of Universal Computer Science*, vol. 29, no. 6, pp. 533–564, 2023, doi: 10.3897/jucs.90528.
- [7] UNESCO, "Guidance for generative AI in education and research," *UNESCO Report*, Paris, France, 2023. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- [8] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston, MA, USA: Center for Curriculum Redesign, 2019.
- [9] Q. Liu et al., "Exploiting cognitive structure for adaptive learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, 2019, pp. 627–635, doi: 10.1145/3292500.3330922.
- [10] S. Esmailigoujar and F. Jooyandehdel, "Artificial intelligence in neurodiverse education: A systematic scoping review," *PsyArXiv*, Jan. 2025, doi: 10.31234/osf.io/mf8ks\_v1.