International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

IDEAL: A Machine Learning Framework for the Automated Detection of Early Academic Learning Disabilities

Alfiya N*, Arunima Sony#, Devi A&,
Nowrine Fathima@, Abin Roy+, Mary Priyanka$
Department of CSE,College of Engineering Kidangoor, Kottayam, Kerala, India
*nalfiya1203@gmail.com,#arunimatessa@gmail.com, &devitheckathu@gmail.com,
@nowrinfathima8@gmail.com,+abinroy21@gmail.com,$marypriyanka@ce-kgr.org

***Abstract*— With the increasing need for early identification of learning disorders among children, there is a growing demand for intelligent, scalable, and child-friendly diagnostic systems. Learning disabilities such as dyslexia, ADHD, and related cognitive challenges often go undetected due to reliance on time-consuming manual assessments and subjective evaluations. The proposed system engages children through adaptive tasks to assess cognitive abilities such as memory, attention, logical reasoning, and language comprehension, reducing anxiety while ensuring accurate data collection. The system incorporates an intelligent framework that combines interactive assessment techniques with data-driven analysis to support early identification of learning disorders. A machine learning–based backend utilizes classifiers like Random Forest, Extra Trees, and Logistic Regression to analyze performance data and detect potential learning difficulties, with comparative evaluation to determine the most effective model. Additionally, a dedicated dashboard enables educators and specialists to visualize results and track progress over time, providing a scalable and efficient solution for continuous monitoring and support.**

***Index Terms*—learning disabilities, early detection, machine learning, classification models.**

I. INTRODUCTION

Learning disabilities and neurodevelopmental disorders such as Dyslexia and Attention Deficit Hyperactivity Disorder (ADHD) and other related learning disorders significantly affect a child's ability to read, write, communicate, and solve mathematical problems, even when they possess normal intelligence. Early detection of these conditions is essential to ensure timely intervention and effective educational support. However, traditional assessment methods are often manual, time-consuming, and dependent on expert evaluation, which can delay accurate diagnosis.

With advancements in machine learning and intelligent systems, automated approaches have been developed to improve the efficiency of early screening. These systems collect data from children through cognitive tests, handwriting samples, speech inputs, and interactive activities designed to capture behavioral and psychomotor skills. The collected data is then preprocessed and analyzed to extract significant features such as response accuracy, handwriting patterns, and behavioral traits. Machine learning algorithms are applied to classify and detect potential learning disabilities, providing reliable and data-driven insights.

In this context, the proposed Intelligent System for Early Detection of Learning Disabilities introduces a gamified platform for cognitive assessment. Each learning disorder is evaluated through a dedicated game-based module that collects user performance data, which is analyzed to predict disorder-specific risk levels. The results are made available to child support facilitators, including educators and specialists, through an interactive dashboard to support early intervention.

II. LITERATURE SURVEY

Recent studies have explored the application of machine learning for early detection of learning and neurodevelopmental disorders. Various models, including Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN),

Na¨ıve Bayes (NB), and Stochastic Gradient Descent (SGD), have been widely evaluated. Among these, Logistic Regression often provides a strong balance between interpretability and performance, while ensemble methods such as Random Forest and Gradient Boosting offer higher adaptability at the cost of increased computational complexity.

Paduthala *et al.* [4] compared multiple models for speech disorder detection and highlighted the effectiveness of deep learning approaches such as CNN, BiLSTM, and Wave2Vec2 in handling noisy and unstructured speech data. Similarly, Sindhu *et al.* [1] showed that transformer-based models outperform traditional methods in speech and voice disorder detection. In addition, Karunasekara *et al.* [8] proposed a web-based system for speech disorder identification and therapy support, demonstrating the potential of accessible and user-friendly platforms for early intervention.

Giri *et al.* [5] applied Random Forest for dyscalculia detection using cognitive assessment data, achieving high accuracy and reduced diagnostic effort. Lalithadevi *et al.* [3] utilized handwriting-based features for Autism Spectrum Disorder detection, where Random Forest achieved superior performance due to its robustness in handling complex feature sets.

Malathi *et al.* [2] employed NLP techniques with SVM for detecting language disorders, achieving high accuracy through effective feature extraction. Raj *et al.* [9] also identified SVM as a strong performer for voice disorder classification due to its ability to handle high-dimensional data.

Santhiya *et al.* [6] and Leeon *et al.* [7] demonstrated the effectiveness of machine learning in early detection of learning disorders using cognitive and behavioral data, with Decision Tree and Random Forest achieving high accuracy and interpretability. Toki *et al.* [10] further showed that Logistic Regression provides reliable and interpretable predictions when applied to game-based cognitive data.

Overall, existing works indicate that while deep learning models achieve high accuracy in domain-specific tasks such as speech and handwriting analysis, traditional machine learning models remain effective for structured cognitive data due to their interpretability and lower computational requirements. However, most approaches focus on single modalities or specific disorders, highlighting the need for a unified, scalable system that integrates multiple assessments for comprehensive early detection.

III. METHODOLOGY

The methodology of the IDEAL system defines a structured approach for designing and implementing an ML-based platform for early identification of learning disorders in children. The system integrates gamified test modules, machine learning-based analysis, and a monitoring dashboard to provide accurate and timely insights for educators and child support facilitators. The methodology outlines the design, implementation, and testing phases involved in building an efficient and user-friendly detection system.

A. SYSTEM ARCHITECTURE

The project has been mainly divided into four modules:

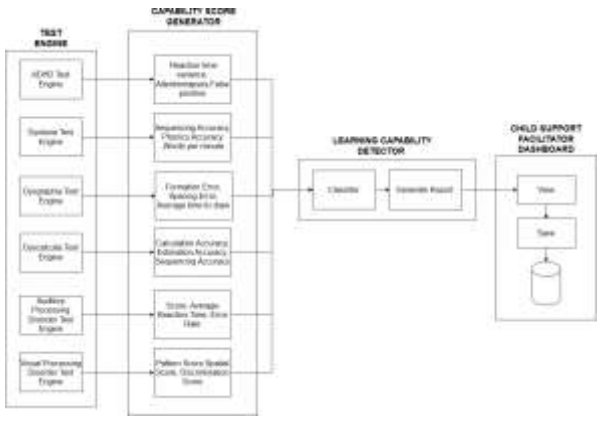

Fig. 1. SYSTEM ARCHITECTURE

1) ***Test Engine Module****:*

- Provides a user-friendly platform for children.
- Child interacts with different test engines.
- Captures response data.
- ADHD Test Engine: Measures attention and reaction behavior.
  TEST: Continuous Performance Task
- Dyslexia Test Engine: Evaluates reading and phonics ability.
  TESTS:

Part 1: Letter Sequencing – Arrange scrambled letters to form words.
Part 2: Phonics Matching – Listen to a word and choose the correct spelling.
Part 3: Typing Speed – Type a short passage to measure WPM.

- Dysgraphia Test Engine: Analyzes writing formation and spacing.
  TESTS:
  Part 1: Shape Copying – Draw shapes following dotted outlines.
  Part 2: Letter Tracing – Trace letters following guides.
  Part 3: Timed Typing – Type sentences accurately.

- Dyscalculia Test Engine: Checks numerical and sequencing skills.
  TESTS:
  Part 1: Number Sequencing – Find missing numbers.
  Part 2: Simple Calculations – Solve basic operations.
  Part 3: Estimation – Estimate number of objects.

- Auditory Processing Disorder Test Engine: Evaluates listening and response accuracy.
  TEST: Sound discrimination

- Visual Processing Disorder Test Engine: Assesses pattern and spatial recognition.
  TESTS:
  Part 1: Pattern Recognition – Complete visual sequences.
  Part 2: Spatial Memory – Recall grid positions.
  Part 3: Visual Discrimination – Identify matching symbols.

- Output: Raw performance data for further analysis.

*2) **Capability Score Generator**:*

- Extracts performance parameters from each test.
- Calculates accuracy, error rates, and response metrics.
- Standardizes scores for model input.
  GENERATED PARAMETERS:
- ADHD Test Engine:Reaction time variance, attention lapses, false positives.
- Dyslexia Test Engine:Sequencing accuracy, phonics accuracy.
- Dysgraphia Test Engine:Formation error, spacing error, average writing time.
- Dyscalculia Test Engine:Calculation, sequencing accuracy and estimation accuracy.
- Auditory Processing Disorder Test Engine:Score, average reaction time, error rate.
- Visual Processing Disorder Test Engine:Pattern score, spatial score, discrimination score.
  Output: Structured capability score dataset.

*3) **Learning Capability Detector**:*

- Receives the capability scores generated from different tests.
- Classifiers such as Extra Trees ,Logistic Regression, RandomForest used as the main model to classify learning disorders.
- Compares multiple classifiers and selects the best-performing model to accurately detect learning disabilities.
- Flags early-stage learning disability indicators such as dyslexia, dyscalculia, ADHD, dysgraphia, auditory processing disorder, and visual processing disorder.
- Generates individualized Results.

*4) **Child Support Facilitator Dashboard**:*

- Allows facilitators to review results.
- Saves student progress to database.
- Enables long time performance tracking.

B. SYNTHETIC DATA

*1) **Synthetic Data Generation**:*

Due to the high sensitivity and privacy requirements associated with children's health and educational data, synthetic data was generated to train and validate the classification models. Data was modeled using Gaussian distributions to represent two distinct groups:

- **Low-Risk Group:** Modeled with high mean accuracy scores and low variance (e.g., $\mu = 90$, $\sigma = 5$).
- **High-Risk Group:** Modeled with lower mean scores and higher variance to simulate the

struggle associated with specific disorders (e.g., $\mu = 50, \sigma = 15$).

Gaussian noise was injected into the dataset to test model robustness and simulate real-world environmental factors or ”bad-day” testing scenarios.

Synthetic Dataset Generation Algorithm with Noise Modeling

1) Start
2) Create Model object
3) Initialize trainer using ModelTrainer
4) Set model = None
5) Set model_name = None
6) Set accuracy = 0.0
7) Set is_trained = False
8) Set random seed for reproducibility
9) Define total number of samples N
10) Compute number of low-risk samples
11) Generate low-risk data
12) Generate high-risk data
13) Limit all generated feature values within the range 0 to 100
14) Combine low-risk and high-risk feature data into a single dataset
15) Combine corresponding labels to form the target vector
16) Create dataset containing features
17) For each feature column in the dataset:
    - Insert noise
    - Add generated noise to original feature values
18) Clip updated feature values to maintain valid data limits
19) Obtain final synthetic dataset enhanced with Gaussian noise
20) Stop

IV. IMPLEMENTATION AND RESULTS

The implementation was carried out in Python using the `scikit-learn` library. The core of the system is a dynamic `ModelTrainer` class that automates the selection of the most accurate classifier for each specific disorder.

A. MACHINE LEARNING MODELS

Three distinct classifiers were implemented and compared:

1) **Random Forest (RF):** An ensemble method using bagging to reduce variance and provide high accuracy.
2) **Extra Trees (ET):** Similar to RF but with increased randomness in split points, often leading to better performance in noisy synthetic datasets.
3) **Logistic Regression (LR):** A baseline linear model used to determine the complexity of the decision boundary.

B. MODEL SELECTION LOGIC

For each disorder (Dyslexia, ADHD, etc.), the system executes a 5-fold cross-validation across all three models. The classifier achieving the highest mean cross-validation accuracy is automatically selected as the primary engine for generating real-time diagnostic predictions.

TABLE I
MODEL ACCURACY COMPARISON TABLE

Disorders	Random Forest	Extra Trees	Log. Reg.
Dyslexia	95.7	96.1	**96.8**
Dysgraphia	96.7	96.7	**97.9**
Dyscalculia	95.9	95.8	**96.8**
ADHD	**99.8**	**99.8**	99.6
Auditory Processing Disorder	95.0	95.1	**95.4**
Visual Processing Disorder	97.1	96.6	**97.4**

Fig. 2. System generated model comparison interface

The accuracy variations of the different models across various disorder modules are represented in Table I. This is followed by a visual comparison of the performance metrics as captured from the system's analytical interface in Fig. 2. The reported model accuracies are evaluated on synthetically generated data with added Gaussian noise, and therefore reflect performance under controlled variability rather than real-world data conditions.

V. CONCLUSIONS

This paper presents IDEAL, an intelligent and scalable system for the early detection of learning disabilities in children. The proposed framework integrates multiple game-based assessment modules to evaluate cognitive, linguistic, and behavioral abilities, enabling automated and data-driven risk prediction for various learning disorders. Machine learning models, including Logistic Regression, Random Forest, and Extra Trees, are employed to analyze extracted performance features and identify potential learning difficulties with high accuracy. The system also provides a facilitator dashboard for monitoring progress and supporting personalized intervention strategies. A representative implementation, such as the dyslexia assessment module, demonstrates how user interactions are captured through structured tasks and transformed into meaningful performance indicators for risk evaluation. Similar approaches are applied across all modules to ensure consistency and scalability. Although the system shows promising results on synthetically generated data, future work will focus on validation using real-world datasets, improving model generalization, and incorporating multimodal inputs such as speech and handwriting analysis. The proposed system is designed as an early screening tool that predicts the risk of learning disabilities, assisting educators and specialists in identifying children who may require further clinical evaluation.

Fig. 3. Input Page

Fig. 4. Letter Sequencing test

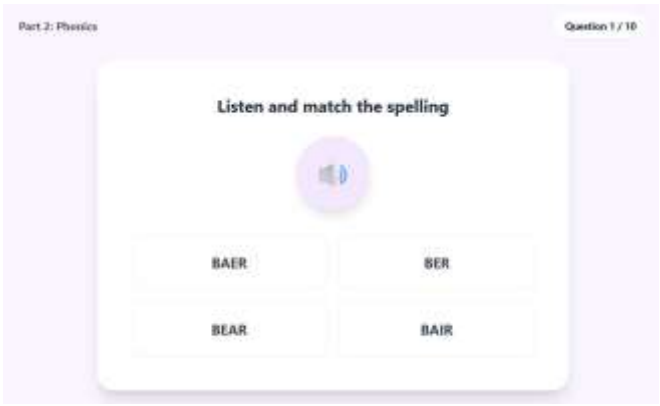

Fig. 5. Phonics matching test

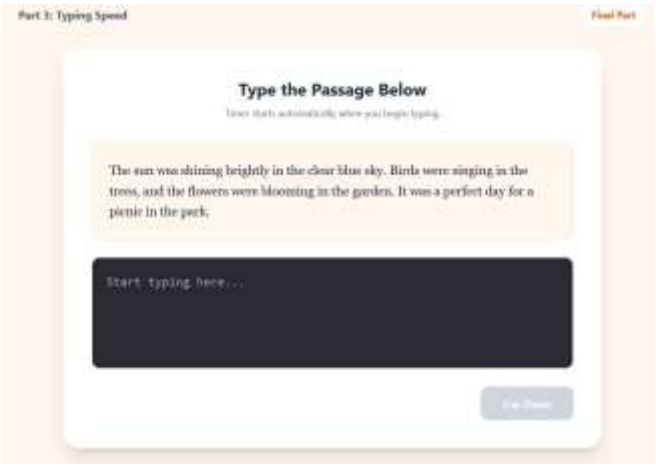

Fig. 6. Typing speed test

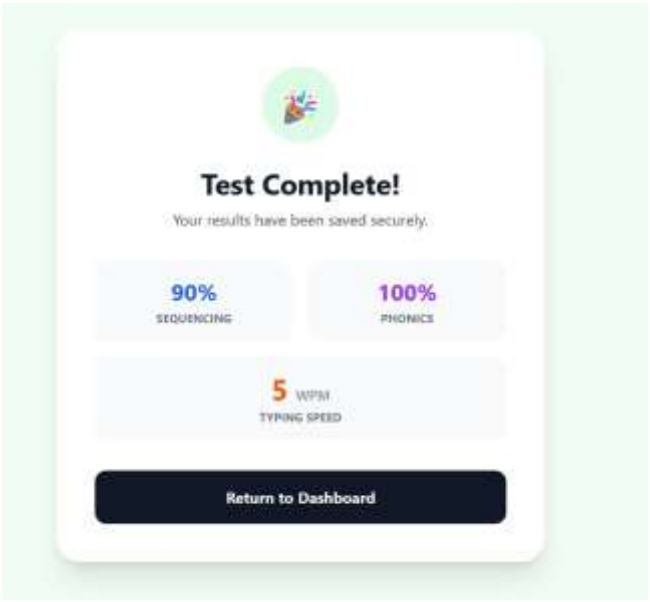

Fig. 7. Test completed

Fig. 8. Output screen.

REFERENCES

[1] I. Sindhu and M. S. Sainin, "Automatic Speech and Voice Disorder Detection Using Deep Learning—A Systematic Literature Review," in IEEE Access, vol. 12, pp. 49667-49681, 2024.

[2] P. Malathi, N. Legapriyadharshini, S. S. Nair, M. P. Sujatha, S. R and T.Thirumalaikumari, "Automated Detection of Language Disorders in Children Using NLP and Machine Learning," 2024 International Conference on Recent Innovation in Smart and Sustainable Technology (ICRISST), Bengaluru, India, 2024.

[3] B. Lalithadevi, N. S. S. Mahalakshmi and N. Varun, "Handwriting-Based Autism Spectrum Disorder Detection in Children using Machine Learning Techniques," 2025 7th International Conference on Intelligent Sustainable Systems (ICISS), India, 2025.

[4] J. M. Paduthala, N. A. Varghese, N. Joseph, S. Mahesh, and S. Thomas, "Comparative Analysis of Machine Learning Models for Speech Disorder Detection," 2025 2nd International Conference on Trends in Engineering Systems and Technologies (ICTEST), Kottayam, India, 2025.

[5] N. Giri et al., "Detection of Dyscalculia Using Machine Learning," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020.

[6] S. Santhiya, S. Priyanka, S. Keerthika, M. K, M. R. M and D. K. B, "Early Detection and Support for Learning Disabilities: A Machine Learning Approach Empowering Educators," 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), Chennai, India, 2023.

[7] Shakirul Islam Leeon, S. A. M. Zahin Abdal, Fahrin Hossain Sunaira, Shanila Nehlin, and Sifat Momen, "A Machine Learning Approach for Early Detection of Learning Disorders in Pediatrics," Proceedings of the 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), Dhaka, Bangladesh, Mar. 2024.

[8] P. Karunasekara, D. Karunarathna, S. Deshitha, S. Lokuliyana, D. De Alwis, and N. Gamage, "Phrasefluent – An Automated Solution for Children's Speech Disorders Identification and Therapy Treatment," Proceedings of the 2023 5th International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 2023..

[9] A. Raj and P. M. O, "Voice Disorder Detection and Classification Using Machine Learning Techniques and Feature Selection Methods," Proceedings of the 2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Bangalore, India, 2024.

[10] E. I. Toki, I. G. Tsoulos, V. Santamato, and J. Pange, "Machine Learning for Predicting Neurodevelopmental Disorders in Children," Applied Sciences, vol. 14, no. 2, p. 837, Jan. 2024.