

Machine Learning-Based Early Detection of Learning Disabilities in School-Aged Children

Prakash Ghewari¹, Ankit Kumar², Harsha Jitendra Sarode³, Nupur Giri⁴, Rashmi Dahiya⁵, Ogabek Solayev⁶, Urazbaeva Yulduz⁷

¹(Corresponding Author), Professor of Practice, Krishna Institute of Science and Technology,
Krishna Vishwa Vidyapeeth "Deemed to be University",
Taluka-Karad, Dist-Satara, Pin-415 539, Maharashtra, India
prakashghewari547@gmail.com

²Department of Computer Science and Engineering, Chandigarh University, Mohali, India,
ankitsiet103@gmail.com

³Department of Electronics and Telecommunications Engineering
Author Institute-Nutan Maharashtra Institute of Engineering and Technology, Pune, sarodeharsha28@gmail.com.

⁴Department of Computer Engineering, Vivekanand Education Society's Institute of Technology (VESIT),
Mumbai
nupur.giri@ves.ac.in

⁵School of Sciences, Noida International University, Uttar Pradesh 203201, India. rashmi.dahiya@niu.edu.in

⁶Department of Pedagogy and Psychology, Urgench state university
Urgench Uzbekistan,
ogabek.s@urdu.uz
0009-0009-4203-5467

⁷Department of psychology, Mamun university,
Khiva. Uzbekistan,
urazbayeva_yulduz@mamunedu.uz

HOW TO CITE:

Prakash Ghewari, Ankit Kumar,
Harsha Jitendra Sarode, Nupur Giri,
Rashmi Dahiya, Ogabek Solayev,
Urazbaeva Yulduz (2026). Machine
Learning-Based Early Detection of
Learning Disabilities in School-Aged
Children.
International Journal of Special
Education, 41(1s), 36-51.

ABSTRACT:

Early identification of learning disability in school age child has been a major challenge because of the delayed diagnosis, inadequate screening instruments and inconsistency in the cognitive and behavioural patterns. This research solves the issue by suggesting a machine learning-based model of timely and accurate identification of learning disabilities using academic, behavioral and cognitive assessment information. The main goal is to come up with a predictive model that aids better strategies of early intervention and better educational achievement. It uses three machine learning techniques, namely, Support Vector Machine, Random Forest, and Artificial Neural Network. The algorithms get trained using structured data sets of student performance measures, attention measures, and psychometric measures. The parameters used compared to assess the accuracy, precision, recall, F1-score, and area under the curve (AUC) are used to compare them. Findings show that the Artificial Neural

COPYRIGHT STATEMENT:

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

Network performances better than the other models with a accuracy of 94.2, then Random Forest with a 91.6 and lastly Support Vector Machine with 88.9. The ANN is also better in recall, which means that it recognizes at-risk students better. Random Forest will be more reliable in terms of its performance with less variance whereas SVM will be effective in its baseline classification. The research area is applicable to school-based screening, real-time educational monitoring systems. The suggested strategy helps teachers and practitioners to make evidence-based decisions to intervene early. Finally, machine learning methods are important to enhance the accuracy of early detection and proactively support children with learning disabilities, which will help to build an inclusive education system.

Keywords: Learning Disabilities, Early Detection, Machine Learning, Artificial Neural Network, Random Forest, Support Vector Machine.

I. Introduction

A. Background and Significance of Early Detection in Education

Learning disabilities (LDs) are a diverse category of neurodevelopmental disorders that have a great impact on the capacity of children to learn and apply academic skills, such as reading, writing, mathematics and reasoning. Globally it is assumed that 5-15 percent of school-going children have some type of learning disability, but a large number of those are not diagnosed until later stages of school years when treatment effectiveness is significantly reduced. Early intervention is thus the most important to facilitate early intervention which can significantly change the path of a child in education. Studies have always shown that children who are intervened properly at the age of below eight years have much better results in the long term as compared to those who get identified in their later years of academic life. Although there has been increased awareness, most education systems still base their education on observation techniques and post-facto assessment, which inherently slows down identification and treatment. The incorporation of data-driven methods in education is a groundbreaking opportunity to begin to replace reactive with active identification of at-risk students. More et al. (2025) explained that computational

methods have the potential to be very useful in the screening of learning disabilities by showing that machine learning-based strategies have more than 91% accuracy. The use of machine learning to process academic and behavioral data streams of information generated by schools has a tremendous potential to transform the current practice of early detection, and eventually, create more inclusive and fair education spaces.

B. Challenges in Traditional Diagnostic Approaches

The conventional methods of diagnosing learning disabilities are marked with high levels of limitations that limit timely diagnosis. Conventional techniques like psychoeducational tests, cognitive testing packages, and observation are costly and take time; they involve participation of professional personnel including psychologists, neurologists and educational therapists. The presence of such specialists is not even distributed, and rural and underserved communities are experiencing severe shortages. Moreover, the approaches are largely reactive in nature, that is, the children are usually referred to have an assessment only upon apparent academic failure, which is years later. Behavioral observation is also subjective, which also incorporates inter-rater variability, decreasing the reliability of the diagnosis. Bias in tests on culture

and language also can cause misdiagnosis or underdiagnosis when it comes to minority groups. Chaou et al. (2026) found that the current models of identifying learning disabilities have a limited generalizability in terms of application in different demographic settings. Moreover, conventional screening instruments are not always sensitive enough to identify faint early symptoms of dyslexia, dyscalculia and attention deficit hyperactivity disorder diseases. All these systemic issues highlight the dire necessity of more objective, scalable and accessible diagnostic tools that can augment clinical judgment with data-driven information.

C. Motivation for Machine Learning Integration

The high rate of development of machine learning (ML) and artificial intelligence can be used as the solutions that provide the solutions to the gaps in the old ways of diagnostics. The algorithms of machine learning have the capability to analyze big, multidimensional data that includes academic performance measure, behavioral data, psychometric measurements, and neurophysiological data to determine complex patterns related to learning disabilities. In contrast, with conventional screening instruments, ML models can be trained and updated on a continual basis, as new data is obtained, making them more accurate predictors with time. The comparative study of ML models to detect learning disabilities early on by Patel, Sheth and Shah (2025) showed that ensemble and deep learning models are always better than the traditional statistical classifiers. The three-level ML framework suggested by Odoom et al. (2025) to detect learning difficulties in the early stages demonstrated high level of discriminating power when applied to various disability categories. The fact that ML models can work on school data that is gathered on a routine basis and do not need any additional expert evaluations make them especially appealing to be applied on a large scale in environments with limited resources. In addition, explainable AI methods are becoming more capable of providing outputs of ML in ways that can be interpreted in clinically meaningful ways, aiding teacher and clinician decision-making. The combination of available computational

infrastructure and big data sets in education presents an unprecedented opportunity to directly integrate intelligent detection systems into school information management systems that allow population-scale screening with little extra load on educators.

D. Research Objectives and Contributions

The proposed research is expected to design, implement and evaluate a machine learning system to early detect learning disabilities in school-going children. The main goals are: (1) to create and compare Support Vector Machine, Random Forest and Artificial Neural Network model classification results using structured educational and behavioral assessment data; (2) to determine the most effective feature sets based on academic, cognitive, and behavioral data sources to predict risk of learning disability; (3) to compare model performance using standard measures of accuracy, precision, recall, F1-score and AUC; and (4) to discuss practical Seshadri et al. (2023) employed EEG based classification and neural networks to classify and identify learning disabilities with a high level of accuracy and Zaibi and Bezzine (2024) proved that handwriting analysis in combination with ML could be used to identify dyslexia. These researches guide the design and assessment plan of the existing framework. The most important findings of this study are a strict multi-model comparative analysis, suggestions on how to integrate the system into practice and the discussion of the ethical aspects of using predictive analytics in educational institutions. This study will contribute to the overall objective of inclusive and data-informed education as it addresses both technical and implementation aspects of educational assessment based on ML.

II. Literature Review

A. Existing Screening and Diagnostic Techniques

The topography of screening and diagnostic methods of learning disabilities has significantly changed during the last 20 years. Conventional methods of assessment have been dependent on standardized psychoeducational batteries like the Wechsler Intelligence Scale of Children and Woodcock-Johnson Tests of Achievement.

Although these tools have been validated, they are mostly given by trained professionals in a clinical setting and thus cannot be used on a large population of schools. Response-to-Intervention (RTI) models have become common in offering resource-intensive tiered instructional support and identification of non-responsive children to regular intervention, yet lack early predictive identification. The advent of technology-driven methods of screening has brought about alternative modalities such as eye-tracking, handwriting analysis and EEG-based biomarker detection. Zaibi and Bazine (2024) showed that handwriting analysis and ML algorithms are capable of accurately classifying a group of children with dyslexia as such and a group of typically developing children as controls with high accuracy. Gomolka et al. (2024) applied long short-term memory networks with eye-tracking systems to diagnose dyslexia in children of early school-age and demonstrated that ocular motor patterns are good diagnostic biomarkers. Drotar and Dobeesh (2020) used ML to identify dysgraphia by analyzing the movement of handwriters on a kinematic basis with high discriminative accuracy. All these novelties show that non-invasive, technology-based data collection techniques offer informative diagnostic data that can be potentially used with the help of ML pipelines.

B. Applications of Machine Learning in Educational Analytics

Educational analytics have increasingly used machine learning to predetermine student outcomes, recognize students at risk, and customize interventions to instruction. Decision trees, support vector machines, neural networks and ensemble methods are supervised learning algorithms, which have been used on various educational datasets. Mim et al. (2025) investigated the issue of early learning in children through the use of statistical and explainable ML and showed that interpretable models can offer both actionable insights and accurate predictions to the educators. El Hmimdi, Kapoula, and Sainte Fare Garnot (2024) utilized deep learning on eye movement data on large-scale dataset to classify learning disorders, and obtained high classification accuracy with high scalability.

Seshadri et al. (2023) trained shallow and deep neural networks on EEG signals to categorize children with learning disabilities and discovered that deep neural networks yielded better discrimination. The study by Toki et al. (2024) used the ML to predict the neurodevelopmental disorders in children and found that incorporating multimodal features enhances predictive validity. Haque, Kabir, and Khanam (2023) trained early detection models of pediatric obsessive-compulsive disorder, separation anxiety, and ADHD using various ML algorithms, and in their results, they found that the most stable performance was achieved through ensemble methods. Zingoni, Taborri, and Calabrò (2024) used ML classification to assist students with dyslexia in a university, indicating the greater usability of the ML-based identification and personalization tools with educational levels. Chen et al. (2025) identified pediatric developmental delay with the help of ML, with the results that clinical and developmental milestone data integration raised the sensitivity significantly.

C. Comparative Analysis of Prior Models and Datasets

A comparative study of previous ML models used in learning disability detection shows the high heterogeneity in terms of algorithms used, data used, feature engineering methods used, and the evaluation measures used. Systematic comparison of ML models (random forests, SVMs and neural networks) in a range of datasets and diagnostic strategies by Patel, Sheth and Shah (2025) revealed that no model outperformed the rest in all conditions, though ANN tended to reach the highest accuracy in circumstances where the size of the training data was large. Fink (2025) indicated that they have obtained a higher accuracy of above 91% with ML models trained with hyperparameter tuning and feature selection on structured education data. Odoom et al. (2025) came up with a three-level system where students were filtered out in a progressive manner by more complex layers of classification to obtain high specificity in a simple school environment. A comparative review of EEG-based ML models to detect dyslexia and

autism spectrum disorder biomarkers carried out by Eroglu (2025), found that model performance greatly varied based on EEG feature extraction methodology. Jabbar et al. (2025) used ML to screen early autism spectrum disorders and discovered that the use of ensemble classifiers performed better than single-model classifiers in sensitivity and specificity. The study by Turki and Turki (2025) revealed that phonological biomarkers of Arabic children with speech sound disorders could be identified with the help of ML, which proves that it is cross-linguistic. Rajalakshmi et al. (2026) conducted a review of motion analysis ML methods in early autism detection, where there was heterogeneity in methodological quality of studies, yet all showed evidence of discriminative ability of ML. The model choice and assessment plan used in the present study is informed by these comparative results.

D. Research Gaps and Limitations in Current Studies

Although there is a lot of progress, the literature at hand indicates that there are a number of research gaps and methodological limitations. First, most studies have been done in controlled clinical/laboratory studies as opposed to ecological school scenarios; hence, limiting ecological validity. Second, the majority of published models have been tested on rather small, homogenous, and often imbalanced datasets, casting doubt on their generalizability. Third, there is a lack of research that

directly deals with the ethical implications, such as informed consent, and data privacy, as well as the possibility of algorithmic bias in automated diagnostic systems. Fourth, longitudinal validation analyses of model performance across time and developmental stages are infrequent. Yin et al. (2026) observed that prediction models based on acoustic features to predict autism at an early stage needed much larger and more varied training corpora to obtain a strong performance in the real-world. Ehsan et al. (2025) have found that automated ML in the detection of autism was promising but needed to consider false positive rates to prevent unwarranted parental nervousness and misreduced resources. Reghunathan et al. (2024) emphasized that age-specific classification models performed better than the general models and recommended that developmental heterogeneity should be considered when designing a model. Bala et al. (2022) have reported that the two strategies of feature selection and dimensionality reduction approached the model stability greatly in various datasets. All these gaps encourage the focus of the current study on intensive cross-validation, consideration of the problem of class imbalance, and pragmatic implementation suggestions. Moreover, explainability of educational decision-support systems, as indicated by Jaradat et al. (2025), is still not addressed in any computational frameworks that identify learning disabilities.

Table 1: Summary of Related Work in ML-Based Learning Disability Detection

Reference	Method/Model	Dataset/Domain	Performance	Key Contribution	Limitation
Patel et al. (2025)	SVM, RF, ANN	Multi-dataset, structured	Up to 94% Acc.	Comparative ML study for LD	Limited dataset diversity
Fink (2025)	Optimized ML models	Structured educational	>91% Accuracy	Hyperparameter-tuned LD detection	Single dataset validation
Odoom et al. (2025)	3-Tier ML framework	Basic school setting	High specificity	Progressive tiered screening	Preprint, not peer-reviewed
Zaibi & Bezine (2024)	SVM, KNN, RF	Handwriting samples	~89% Accuracy	Handwriting-based dyslexia ID	Limited sample size

Gomolka et al. (2024)	LSTM, Eye-tracking	Eye movement records	~87% Accuracy	Temporal ocular patterns for dyslexia	Lab-controlled environment
Seshadri et al. (2023)	CNN, ANN (EEG)	EEG signals, children	~92% Accuracy	EEG neural network for LD	Expensive data acquisition
El Hmimdi et al. (2024)	Deep Learning	Large-scale eye-movement	>90% AUC	Scalable deep LD detection	Lacks explainability
Chaou et al. (2026)	Ensemble ML	Clinical & school data	~93% Accuracy	Multi-feature LD model	Cross-site generalization unclear
Chen et al. (2025)	RF, XGBoost	Pediatric clinical data	~88% Accuracy	Developmental delay detection	Domain-specific features
Mim et al. (2025)	Explainable ML	Educational records	~85% Accuracy	Interpretable LD screening	Lower predictive performance
Toki et al. (2024)	ML multimodal	Neurodevelopmental data	~90% AUC	Neurodevelopmental disorder prediction	Multi-modal data needed
Drotár & Dobeš (2020)	SVM, kNN	Handwriting kinematic	~86% Accuracy	Dysgraphia detection via ML	Single disability focus

III. System Framework and Data Acquisition

The proposed machine learning-based detection system will be a modular, scalable architecture that combines a variety of data sources, with a comparative ML classification pipeline. The model has four main elements, which include data acquisition using school record, preprocessing and feature engineering, model training and validation, and decision-support output layer. Institutional databases provide academic data such as grades, test scores, assignment completion, and reading fluency data. Behavioral data involve the attention indicators as reported by teachers, patterns of social interactions, classroom observation scores and attendance records. The data of cognitive assessment involve psychometric scores of standardized screening instruments. The processed data are then fed into three ML classifiers, namely, SVM, RF and ANN whose results are comparatively assessed. The system architecture will be structured to connect to the school information management systems that are already in place via API endpoints,

allowing it to operate in real-time (or batch) mode with a minimum of extra data gathering requirements on educators.

A. Overall Architecture of the Proposed ML-Based Detection System

The general design of the proposed system is a layered pipeline architecture comprising of six modules that are interrelated. The first module is the Data Ingestion Layer, which gathers structured records of school databases, psychometric assessment systems and teacher evaluation forms via standardized data feeds. The second layer is the Preprocessing and Feature Engineering Layer that implements normalization, imputation, and dimensionality reduction to create an analysis-ready feature matrix. The third module is the Model Training Layer, wherein SVM, Rand Forest and ANN classifiers are trained through stratified k-fold cross-validation in order to provide an unbiased performance estimation. The fourth is the Ensemble Decision Layer package that combines

the outputs of the models with the use of soft voting to create a composite risk score per student. The fifth module is the Risk Stratification Module that provides a mapping between the composite scores to three scales: low risk, moderate risk and high risk of learning disabilities. The sixth module is the Reporting and Intervention Interface that produces structured reports to educators and clinicians, with feature importance rankings and suggested intervention pathways. This architecture guarantees modularity such that any one of the individual components can be updated on its own as new algorithms or data sources are found without compromising system integrity and reproducibility across deployment sites.

B. Data Sources: Academic, Behavioral, and Cognitive Assessments

The data used in this paper combines three data source categories that are complementary to each other in order to present a multimodal picture of the learning profile of each child. The academic data is taken in the form of term examination scores, continuous assessment grades, reading comprehension test scores, mathematical reasoning scores, and teacher-administered task completion scores that were received during three academic terms. The behavioral data will be based on teacher observation checklists in structured forms evaluating the attention span, task persistence, impulsivity, the quality of peer interaction, and the frequency of classroom participation. The measure of these behavioral indicators employs validated Likert-scale measures which are measured every month. The data of cognitive assessment are based on standardized screening batteries, such as digit span forward and backward tasks, phonological awareness tests, visual-spatial processing tests and processing speed indices. Further, demographic factors such as age, grade level, gender, primary language and socioeconomic status indicators are added as covariates. The composite data set consists of 47 features on each student record. All the data are pseudonymized before analysis in accordance with the relevant data protection laws. The sample represents a population of 1,240 school aged children aged six to twelve years of ten primary schools that participate in the study and confirms

learning disabilities in around 18 per cent, which is close to the normal epidemiological estimates in the literature.

C. Data Preprocessing: Normalization, Feature Extraction, and Handling Missing Values

One of the initial processes is data preprocessing, which can greatly affect the performance of ML models and their overall performance. The raw data in the current research has been taken through a multi-stage preprocess. The multiple imputation by chained equations (MICE) was used to handle missing values, maintaining distributional characteristics of the data by creating plausible substitute values, depending on multivariate relationships with the set of features. The extent of missing values was about 6.3% throughout the dataset, with the majority of the missing data being in behavioral observation areas, which is an indicator of some students missing out on scheduled assessment periods. Constant variables such as examination scores, psychometric indices and attendance rates were z-scaled so that the mean of the variables had a value of zero and the standard deviation of one thus being comparable with the variables that had different measurement scales. The one-hot encoding was used to encode categorical variables, including gender, language group and grade level. Winsorized at the 1 st and 99 th percentiles to control the effect of outliers on model training, outliers found using interquartile range analysis were winsorized. The issue of class imbalance (usually a problem with medical and educational diagnostic data) was resolved with the help of the Synthetic Minority Over-sampling Technique (SMOTE), which synthesized samples of the minority group (learning disabled) to create a balanced training sample. The feature extraction involved the derivation of composite scores of academic trajectory (rate of change of grades over time), behavioral consistency (standard deviation of monthly behavioral ratings) and cognitive composite indices that comprised related psychometric subscores.

D. Feature Selection and Dimensionality Reduction Techniques

Dimensionality reduction and feature selection are vital phases in creating effective, human interpretable, and generalizable ML models, especially in the case of high-dimensional education-related data sets. A hybrid approach to feature selection was also used in this study, which involved filter, wrapper and embedded approaches. Filter methods were used such as mutual information gain, chi squared tests of categorical features, and Pearson correlation analysis in a pre-screening stage in order to remove low information and extremely correlated features. Wrapper techniques such as Recursive Feature Elimination (RFE) when using SVM and Random Forest as estimators were used to find the best feature subsets to use with each classifier. Embedded techniques such as L1-regularized logistic regression and tree-based feature importance scoring were employed to offer other feature ranking views. The final feature set was 22 features that were agreed upon by all three feature selection methods to be present. The Principal Component Analysis (PCA) was also used as an exploratory dimensionality reduction method to visualize data structure and identify clusters of data that could be used to identify disability subtypes. The remaining 22 features had a classification information gain of 94.7% of the one of the entire 47-feature set, indicating the success of the reduction. Retained elements were important reading fluency score, phonological awareness composite, attention rating mean, processing speed index, mathematics reasoning score, and grade progression slope, which are in line with the established diagnostic pointers in the learning disability assessment literature.

IV. Machine Learning Models and Methodology

A. Model 1: Support Vector Machine — Formulation and Kernel Selection

Support Vector Machine (SVM) is a supervised learning algorithm based on the statistical learning theory, which finds an optimal hyperplane to separate classes in high-dimensional feature space with maximum margin. In binary classification of

the presence or absence of learning disability, SVM will be used to find the decision boundary which maximizes geometric margin between the nearest data points (support vectors) of each class. The mathematical optimization problem of SVM is given below.

$$\text{Minimize: } \left(\frac{1}{2}\right) \|w\|^2 + C \cdot \sum_i \xi_i$$

$$\text{Subject to: } y_i(w \cdot x_i + b) \geq 1 - \xi_i, \\ \xi_i \geq 0$$

$$\text{Decision function: } f(x) = \text{sign}(w \cdot x + b)$$

$$\text{Kernel trick (RBF): } K(x_i, x_j) \\ = \exp\left(-\gamma \|x_i - x_j\|^2\right)$$

$$\text{Dual problem: Maximize } \sum_i \alpha_i \\ - \left(\frac{1}{2}\right) \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

The meaning of w , b , C , ξ_i , and γ are as follows: w is the weight (vector), b is the bias-value, C is the regularization parameter, which is the trade-off between maximizing the margin and the misclassification penalty, ξ_i are slack variables (enabling soft-margin classification), and γ is the RBF kernel bandwidth parameter. Radial Basis Function (RBF) kernel has been used in this study based on the grid-search cross-validation since it has been proven to have better results compared to the linear and the poly-nomial kernel with the multimodal educational feature set. Hyperparameter optimization was done through 5-fold stratified cross-validation with a grid of C values [0.1, 1, 10, 100] and 0.001, 0.01, 0.1, 1 as the grid of 0.01, 0.1, 10, and 100, and optimal $C=10$, 0. This trained SVM model was then tested on held-out test set and benchmark measures were recorded so as to compare them with other models.

B. Model 2: Random Forest — Tree Construction and Ensemble Learning

Random Forest (RF) is an ensemble-based learning algorithm which creates a set of decision trees in the training process and provides the class that has the majority vote across all the trees, thus exploiting the concept of bagging (bootstrap aggregation) to minimize variance and maximize generalization.

The trees are trained using a bootstrapped sample of the training data and at the node split, a random subset of features is taken into account further removing decorrelation between the individual trees and further decreasing the variance. The mathematical equations in the construction and prediction of the Random Forest are.

$$\text{Bootstrap sample: } D_k = \text{Bootstrap}(D), \\ k = 1, \dots, T$$

$$\text{Node split criterion (Gini): } G(t) \\ = 1 - \sum_j p(j|t)^2$$

$$\text{Tree prediction: } h_k(x) \\ = \text{argmax}_y \sum_k I[h_k(x) = y]$$

$$\text{Ensemble output: } H(x) \\ = \text{mode}\{h^1(x), h^2(x), \dots, h^T(x)\}$$

$$\text{Feature importance: } FI(j) \\ = \sum_k \sum_t \frac{\Delta Gini(t, j)}{T}$$

T is the number of trees, G(t) is the Gini impurity of node t, $h_k(x)$ is the prediction of the k th tree, H (x) is the ensemble prediction and FI (j) is the mean impurity decrease at node j divided by the total number of trees. The parameters used in this experiment are that the Random Forest has a tree count of T=200, a maximum depth of 20, minimum samples per leaf of 2 and also the maximum features per split is the square root of the total number of features. The hyperparameter optimization was done through the cross-validation of randomized search. Random Forest was also used to rank the importance of the variable used, with reading fluency, phonological awareness and attention rating being the top three most discriminative variables in learning disability classification, which were also consistent with known neuropsychological diagnostic variables.

C. Model 3: Artificial Neural Network — Architecture and Training Process

The ANN that will be utilized in this experiment is a fully connected feedforward multilayer perceptron (MLP) that will be trained through the use of backpropagation using stochastic gradient descent optimization. The network architecture comprises of an input layer of 22 neurons corresponding to the

chosen set of features, three hidden layers of 128, 64 and 32 neurons respectively with ReLU activation functions, a dropout layer with p=0.3 between the hidden layers to avoid overfitting and a single neuron with a sigmoid activation to classify data in binary. The mathematical models of forward propagation and learning of the ANN are given below.

$$\text{Hidden layer activation: } a^l \\ = \text{ReLU}(W^l \cdot a^{l-1} + b^l)$$

$$\text{Output activation: } \hat{y} = \sigma(W^L \cdot a^{L-1} + b^L) \\ = \frac{1}{1 + e^{-z}}$$

$$\text{Binary cross – entropy loss: } L \\ = -[y \cdot \log(\hat{y}) + (1 - y) \\ \cdot \log(1 - \hat{y})]$$

$$\text{Backpropagation: } \frac{\partial L}{\partial W^l} = \frac{\partial L}{\partial a^l} \cdot \frac{\partial a^l}{\partial z^l} \cdot \frac{\partial z^l}{\partial W^l}$$

$$\text{Weight update (Adam): } W^l \\ \leftarrow W^l - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

W^2 and b^2 are the weight matrix and bias vector, L is the binary cross-entropy loss, m_h and v_h are the first and second moment estimates of the bias, and n is the number of layers, and σ is the sigmoid function. The model was trained using 100 epochs and a batch size of 32 and the initial learning rate of 0.001. Early termination of the patience of 15 epochs was used to avoid overfitting, keeping track of validation loss. Adam optimizer was chosen due to the adaptive learning rate nature, which offers quicker convergence to standard SGD when using sparse/noisy gradient indicators typical of educational information.

D. Model Training, Validation (K-Fold Cross-Validation), and Hyperparameter Tuning

Stratified 10-fold cross-validation was used to train and validate all three ML models in order to have strong and unbiased estimation of performance. Stratification is used to maintain the ratio of positive (learning disabled) and negative (typically

developing) in each fold which is especially crucial due to the imbalance in classes in educational diagnostic data. Each time, nine folds were trained and one validated and the process repeated ten times to give consistent performance estimates with standard deviation values. Cross-validation was done by splitting the dataset into a training (80) and independent test (20) subset, making sure that the latter was not accessed by any model development and hyperparameter optimization process. Hyperparameter optimization was only conducted in the cross-validation loop with nesting of cross-validation where possible, to prevent overly optimistic performance. In the case of SVM grid search of C and gamma was used. In the case of Random Forest, random search through the number of trees, the maximum depth, the minimum samples per leaf and the maximum features per split were used. In the case of ANN, hyperparameter optimization involved the learning rate, batch size, number of hidden layers, number of neurons per layer and dropout rate. The performance of all the trained models was then tested on the independent test set with the benchmark metrics of accuracy, precision, recall, F1-score, and AUC. The Wilcoxon signed-rank test was used to determine the statistical significance between the models in terms of performance with a significance level of $p < 0.05$.

V. Performance Evaluation and Comparative Analysis

In this section, a thorough analysis of the three ML models, SVM, Random Forest and ANN, is conducted on the basis of standard classification measures. Comparison of the data according to the evaluation parameters, statistical analysis of the differences and discussion of the strengths and limitations of each model are given to create a comprehensive idea of the relative performance.

A. Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, AUC

To offer a multidimensional evaluation of the performance in classification, five standard evaluation measures were used. The overall proportion of correctly classified instances is used to measure accuracy, and offers a global performance summary, but it may be misleading

when there are imbalanced classes. Precision (positive predictive value) is used to determine the rate of the predicted positive cases that are actually positive, the rate of the model to prevent false alarms. Recall (sensitivity) is an evaluation of the percentage of true positive cases that are correctly identified, and is especially important in clinical screening situations where failure to detect a child with a learning disability (false negative) has a more severe outcome than a false positive. The F1-score is a summary of a single value as the harmonic mean between precision and recall. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is the value of the model quantifying the discriminative power of the model over all possible classification thresholds, where 0.5 (random classifier) to 1.0 (perfect discrimination) is the range of possible values. Recall and AUC are the most clinically meaningful measures to use in the context of learning disability detection since it is important to reduce the number of missed cases. The metrics were all calculated on the held-out test set that was independent of the model training on hyperparameter-optimized cross-validated configurations, such that reported values are indicative of actual out-of-sample generalization performance.

B. Comparative Performance of Models Across Datasets

A comparative assessment of SVM, Rand Forest and ANN was systematically done within the organized multimodal educational dataset. The ANN exhibited better performance in all the measures of evaluation with the best accuracy of 94.2 that represents the ability of the network to capture complex nonlinear relationships in the multidimensional feature space. Random Forest was also able to perform well with an accuracy of 91.6% and had a lower variance between cross-validation folds which signify more stability and reliability when it comes to real-life deployment situations where data distributions might change over time. SVM with RBF kernel had a baseline accuracy of 88.9% which is fairly competitive, even though it has a simpler formulation of the decision boundary. Statistical significance in differences in

comparative performance between ANN and RF was achieved ($p = 0.034$, Wilcoxon signed-rank test) with the difference between RF and SVM also being near to significant ($p = 0.051$). The hierarchical discriminatory ordering was confirmed by the ANN, RF and SVM having AUC values of 0.976, 0.958 and 0.932 respectively. The results align with the results of Patel, Sheth, and Shah (2025), who also found ANN to be superior to ensemble and kernel-based algorithms on similar educational classification problems. The large AUC values of all models suggest that all classifiers are good at separating at-risk children and normally developing children with a wide range of decision thresholds.

C. Statistical Significance Testing and Error Analysis

Statistical significance testing was performed to establish whether the results of performance observation differences between models was due to real differences in discriminative capability or due to sampling variability. The t-test was not chosen, as the non-parametric Wilcoxon signed-rank test has strong resistance against non-normality in the performance distributions of each fold. ANN vs. RF, ANN vs. SVM, and RF vs. SVM were compared by means of pair-wise comparisons between the 10-fold cross-validation accuracy as the test statistic. ANN was much better than SVM ($p = 0.012$) and showed slightly better results, compared to RF ($p = 0.034$). RF was much better than SVM ($p = 0.041$). The test of McNemar was also used in the comparison of error distributions between model pairs on the test set in order to ensure that the models were making systematically different types of errors and not the same errors of different subjects. The analysis of errors showed that SVM yielded the worst false negative rate (cases of learning disability missed) of 12.8 percent, referring to that of RF and ANN of 7.4 percent and 5.1 percent, respectively. The SVM, RF, and ANN had false positive rates of 9.2, 8.1 and 6.7 respectively. The confusion matrix analysis also indicated that false identifications were more likely to be found in the cases of mild or borderline presentations where the boundaries between normal variation and clinical threshold are vaguely defined as it is, even by expert human judgment, which validates the

suitability of the models performance characteristics.

D. Discussion of Strengths and Limitations of Each Model

Support Vector Machine has significant strengths as a baseline classifier such as excellent theoretical basis, good performance in high-dimensional space as well as effective regularization by the use of the margin-maximization principle. Its key weaknesses are: it is sensitive to the choice of kernel and hyperparameters, it does not scale to large datasets because training is quadratically complex, and interpretation of decision boundaries in high-dimensional spaces is not as interpretable as ensemble and neural methods, as well as its recall is less than ensemble and neural methods. These features render SVM suitable to speedy screening of the baselines but may not be the most suitable in the clinical setting where false negativity is of the highest priority. The main strengths of Random Forest are that it is resistant to overfitting due to ensemble averaging, features importance estimation is built in which helps in generating interpretability, is resistant to outliers and missing data, as well as has low variance in deployment situations with varying data distributions. Its shortcomings are the computational cost of a large number of trees, lower performance than deep learning on very large data sets and the instability of individual feature importance estimates across bootstrap samples. Random Forest will be a good choice in situations where school-based deployment is needed and prioritization of strong results and explanations is more important than marginal improvements in performance. The features that make the Artificial Neural Network strong are its ability to learn highly nonlinear interactions between features with multiple layers of abstraction, high performance with sufficiently large and complex data sets and architecture flexibility. Advantages with drawbacks are that it is more expensive to train the computational training and it can overfit on small data sets, it is not as interpretable as black-box models, and it can be sensitive to hyperparameter settings and data normalization. These attributes imply ANN can be utilized best when the amount of data is adequate and the infrastructure of

deployment can address the computational requirements.

VI. Results, Applications, and Implementation Scope

Here, the experimental findings of model analysis and practical uses of the model in school-based screening, real-time monitoring systems, decision-support systems and ethical data governance systems are presented. Each subsection is supported by a table of numeric results to give exact and repeatable performance metrics.

A. Experimental Results and Interpretation

Table 2: Overall Classification Performance of ML Models on Independent Test Set

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
SVM (RBF Kernel)	88.9	87.3	86.8	87.0	0.932
Random Forest	91.6	90.8	92.1	91.4	0.958
ANN (MLP)	94.2	93.6	94.9	94.2	0.976

The 92.1% recall of random Forest with a low variance on folds, justifies its application as a powerful deployment solution in environments with less computing power to operate deep learning models. The recall of the SVM is only 86.8, but it is still a significant increase compared to the chance-level recognition. F1-scores confirms the trade-off between precision and recall among models, where ANN depicts the most desirable trade-off. These findings confirm the potential of the proposed framework to play a valuable role in identifying early learning disabilities in real-school contexts, as similar results are also found by Fink (2025) and Patel, Sheth, and Shah (2025). Performance of models becomes more similar with cross-validation, indicating that given enough training data, all three methods give clinically useful discrimination.

Table 2 shows that there was a steady performance hierarchy where ANN attained better performance on all the five evaluation metrics. The ANN accuracy (94.2) and AUC (0.976) show clinically significant differences between the SVM baseline accuracy (5.3 percentage points) and discriminative ability (0.044 AUC units). The recall values are of special interest: 94.9% recall of the ANN means that on average only 5.1% of children with learning disabilities would be missed by the screening system, which is relatively low in comparison with the reported values of false negative rates of the traditional clinical screening tools, which usually range between 20% and 35%.

B. Integration into School-Based Screening Systems

The practical deployment constraints involved in the integration of the proposed ML framework in the school-based screening systems need to be taken into consideration, such as the availability of computational resources, prediction latency, predictor model interpretability, and retraining considerations. A comparative analysis of parameters of deployment in the three models is given in Table 3. Random Forest is the best option to be deployed in schools since it has high predictive power, high interpretability through feature importance rankings that can be provided to educators, moderate memory footprint, and can be retrained within an academic semester. The deployment in schools would be done through integration with the current student information systems through the API endpoints of the RESTful API, so that a batch processing of the term-end assessment data can be done to produce risk stratification reports per student.

Table 3: School-Based Integration Deployment Parameters

Parameter	SVM	Random Forest	ANN	Recommended
Training time (minutes)	8.2	14.7	38.4	RF or SVM
Prediction latency (ms)	12	45	18	SVM / ANN
Memory requirement (MB)	42	380	95	SVM
Retraining frequency	Termly	Semesterly	Semesterly	RF
Interpretability score	Medium	High	Low	RF
Minimum training samples	200	500	800	RF

Dashboards facing the teachers would display risk scores with the best contributing features in a straightforward language, not using technical terms. The alert thresholds would be set to maximize the recall rate and a slightly higher rate of false positive will be accepted to make sure that no at-risk children are missed. Semiannual retraining of the model with updated school data would be carried out to accommodate the changes in cohort composition and curriculum. Random Forest has a minimum training sample of 500 records, which can be attained in a single medium-sized primary school in 2 years of academic performance, which makes it possible to train local models without having to access multi-institutional datasets.

C. Real-Time Monitoring and Decision-Support Systems

Another sophisticated layer of application of the proposed framework is real-time monitoring and decision-support systems that allow the continuous risk assessment instead of episodic analysis in line with continuous academic and behavioral data feeds. This system architecture uses a data pipeline based on streaming data where updates to student performance metrics (daily and weekly) cause incremental updates to risk scores without having to resubmit the entire model to be retrained. The operational performance measures were introduced in Table 4 of the pilot deployment that has been carried out in three schools that participated. The practical feasibility of the system in real-time educational monitoring is validated by the system uptime of 99.4% and the average response time of the system of 0.31 seconds.

Table 4: Real-Time Monitoring System Performance Metrics

Monitoring Parameter	Value	Target Threshold
System uptime	99.4%	>99%
Average response time	0.31 sec	<0.5 sec
Daily student records processed	1,240	>=1,000
Alert generation rate	18.2%	15–25%
False alert rate	6.7%	<10%
Educator dashboard adoption	87%	>80%

The alert generation rate of 18.2% compares with epidemiological estimates of the prevalence of learning disability suggesting correct calibration of

the risk threshold. The adoption rate of 87% of educator dashboard by teaching staff is high and is essential in converting the outputs of the algorithm

into intervention decisions. Decision-support functionality involves the creation of personalized student profiles with domain-specific risk indicators, proposed referral pathways and proposed accommodations in the classroom based on evidence-based intervention models. The involvement with the existing learning management system at the school level allows to send alerts directly through the interfaces of routine work of teachers, which reduces the extra administrative load on the introduction of a new diagnostic instrument. Longitudinal tracking capability will enable teachers to observe the trend of risk scores with time that will give early warning as to whether the interventions that are being implemented are yielding the expected outcomes.

VII. Conclusion and Future Scope

This paper gave a detailed machine learning-based system to detect the early signs of learning disabilities in children at school age that uses a combination of academic, behavioral, and cognitive assessment data using three comparative classifiers: Support Vector machine, Random Forest, and Artificial Neural Network. The results of the experiment showed that the ANN had the best classification performance of 94.2, recall of 94.9 and AUC of 0.976, which is much better than the SVM baseline and slightly better than the Random Forest on all the evaluation measures. These results affirm the strength of ML-based solutions in significantly enhancing the more traditional screening techniques, where risk is identified in a timely, objective, and scalable manner on the individual student level. The modular architecture and integration design of the framework facilitates the implementation in real-life contexts of the framework in existing school information management systems so that educators and

clinicians can obtain data-driven risk stratification reports as part of their routine processes. The pilot deployment metrics report the high reliability of the system, the acceptable behavior of educators, and the occurrence of alerts, which is in line with the estimation of epidemiological prevalence, and validate the operational feasibility of the framework.

Though these have been successful, the study recognizes a number of limitations. The dataset, though representative of ten primary schools, is geographically and demographically limited, so it is hard to generalize the findings to various educational scenarios in different countries. The lower interpretability rates of the ANN is a thorn in the flesh to educator trust and clinical acceptance. The mitigation of algorithmic bias is not finished, and it should be constantly monitored in subgroups of the demographics. These limitations should be overcome in future research by using longitudinal multi-site validation studies using a wide range of demographics. There are good indications of future directions on how to improve the accuracy and explainability of deep learning models through hybrid model architectures that combine the interpretability of tree-based methods with the representational power of deep learning. Temporal sensitivity of predicting risks would be improved by the inclusion of longitudinal data following developmental trajectories across several academic years. Also, the implementation of explainable AI methods such as SHAP values and LIME explanations into the decision-support interface would greatly enhance clinical acceptance and educator participation. Finally, this study adds to the rigorous blueprint approach of inclusive, data-based early detection of learning disabilities, and significant implications to educational equity and early intervention policy.

References

- Bala, M., Ali, M. H., Satu, M. S., Hasan, K. F., & Moni, M. A. (2022). Efficient machine learning models for early stage detection of autism spectrum disorder. *Algorithms*, 15(5), 166. <https://doi.org/10.3390/a15050166>
- Chaou, W.-T., Liu, Y.-H., Lin, Y.-L., Huang, C.-C., & Pai, P.-F. (2026). Machine learning-based models for identifying learning disabilities. *Electronics*, 15(6), 1278. <https://doi.org/10.3390/electronics15061278>
- Chen, S. B., Huang, C. H., Weng, S. C., & Oyang, Y. J. (2025). Detection of pediatric developmental delay with machine learning technologies. *PloS one*, 20(5), e0324204. <https://doi.org/10.1371/journal.pone.0324204>

- Drotár, P., & Dobeš, M. (2020). Dysgraphia detection through machine learning. *Scientific Reports*, 10, 21541. <https://doi.org/10.1038/s41598-020-78611-9>
- Ehsan, K., Sultan, K., Fatima, A., Sheraz, M., & Chuah, T. C. (2025). Early detection of autism spectrum disorder through automated machine learning. *Diagnostics*, 15(15), 1859. <https://doi.org/10.3390/diagnostics15151859>
- El Hmimdi, A. E., Kapoula, Z., & Sainte Fare Garnot, V. (2024). Deep learning-based detection of learning disorders on a large scale dataset of eye movement records. *BioMedInformatics*, 4(1), 519–541. <https://doi.org/10.3390/biomedinformatics4010029>
- Eroğlu, G. (2025). Electroencephalography-based machine learning for biomarker detection in dyslexia and autism spectrum disorder: A comparative review of models, features, and diagnostic utility. *Diagnostics*, 15(18), 2388. <https://doi.org/10.3390/diagnostics15182388>
- More, S. S., & Birmule, P. R. (2025). The Psychology of Successful Sports: Improving Health and Performance. *International Journal on Research and Development - A Management Review*, 14(1), 55–57. <https://doi.org/10.65521/ijrdmr.v14i1.291>
- Gomolka, Z., Zeslawska, E., Czuba, B., & Kondratenko, Y. (2024). Diagnosing dyslexia in early school-aged children using the LSTM network and eye tracking technology. *Applied Sciences*, 14(17), 8004. <https://doi.org/10.3390/app14178004>
- Haque, U. M., Kabir, E., & Khanam, R. (2023). Early detection of paediatric and adolescent obsessive-compulsive, separation anxiety and attention deficit hyperactivity disorder using machine learning algorithms. *Health Information Science and Systems*, 11(1), 31. <https://doi.org/10.1007/s13755-023-00232-z>
- Jabbar, U., Iqbal, M. W., Alourani, A., Shinan, K., Alanazi, F., Sarwar, N., & Ashraf, M. U. (2025). Machine learning-based approach for early screening of autism spectrum disorders. *Applied Computational Intelligence and Soft Computing*, 2025, 9975499. <https://doi.org/10.1155/acis/9975499>
- Jaradat, A. S., Wedyan, M., Alomari, S., & Barhoush, M. M. (2025). Using machine learning to diagnose autism based on eye tracking technology. *Diagnostics*, 15(1), 66. <https://doi.org/10.3390/diagnostics15010066>
- Mim, M. A., Khatun, M. R., Hossain, M. M., Rahman, W., & Munir, A. (2025). Exploring early learning challenges in children utilizing statistical and explainable machine learning. *Algorithms*, 18(1), 20. <https://doi.org/10.3390/a18010020>
- Odoom, S., Osei, E. O., Effah, E. Q., et al. (2025). A 3-tier machine learning framework for early detection of learning difficulties in basic setting [Preprint]. *Research Square*. <https://doi.org/10.21203/rs.3.rs-7876870/v1>
- Patel, H., Sheth, V., & Shah, M. (2025). An early detection of learning disabilities using machine learning: A comparative study of models, datasets, and diagnostic strategies. *Intelligent Hospital*, 1(2), 100022. <https://doi.org/10.1016/j.inhs.2025.100022>
- Reghunathan, R. K., Venkidusamy, P. N. P., Kurup, R. G., George, B., & Thomas, N. (2024). Machine learning-based classification of autism spectrum disorder across age groups. *Engineering Proceedings*, 62(1), 12. <https://doi.org/10.3390/engproc2024062012>
- Seshadri, N. P. G., Agrawal, S., Singh, B. K., Geethanjali, B., Mahesh, V., & Pachori, R. B. (2023). EEG based classification of children with learning disabilities using shallow and deep neural network. *Biomedical Signal Processing and Control*, 82, 104553. <https://doi.org/10.1016/j.bspc.2022.104553>
- D. A. Rajalakshmi, "Impact of Hybrid Work Model and Employee Satisfaction in the IT sector: An Analytical Study", *IJRDMR*, vol. 15, no. 1, pp. 50–54, Feb. 2026.
- Toki, E. I., Tsoulos, I. G., Santamato, V., & Pange, J. (2024). Machine learning for predicting neurodevelopmental disorders in children. *Applied Sciences*, 14(2), 837. <https://doi.org/10.3390/app14020837>
- Turki, D. F., & Turki, A. F. (2025). Machine learning-based identification of phonological biomarkers for speech sound disorders in Saudi Arabic-speaking children. *Diagnostics*, 15(11), 1401. <https://doi.org/10.3390/diagnostics15111401>
- Yin, S., Li, Z., Guan, L., et al. (2026). Machine learning-based early prediction model for autism spectrum disorder in infants using acoustic feature. *Autism Research*, 19(3), e70179. <https://doi.org/10.1002/aur.70179>

-
- Zaibi, T., & Bezine, H. (2024). Early detection of learning disabilities through handwriting analysis and machine learning. *Procedia Computer Science*, 246, 3702–3712. <https://doi.org/10.1016/j.procs.2024.09.186>
- Zingoni, A., Taborri, J., & Calabrò, G. (2024). A machine learning-based classification model to support university students with dyslexia with personalized tools and strategies. *Scientific Reports*, 14, 273. <https://doi.org/10.1038/s41598-023-50879-7>