ITECX Journal International

International Journal of Technology, Education and Science

Vol. 1, No. 1 (2025) | pp. 19–35 | DOI: 10.55001/itecxj.2025.1.1.002 | OPEN ACCESS

RESEARCH ARTICLE

Deep Learning Approaches for Early Detection of Learning Disabilities in Primary School Students

Euripides Hatzikraniotis, Dimitrios Papadopoulos, Elena Konstantinidou

Aristotle University of Thessaloniki, Greece

Received: January 20, 2025 | Accepted: April 5, 2025 | Published: June 1, 2025

DOI: 10.55001/itecxj.2025.1.1.002

ABSTRACT

Early identification of learning disabilities is critical for effective intervention and long-term educational success. This paper proposes a novel multimodal deep learning framework that analyzes heterogeneous student interaction data—including handwriting patterns, reading behaviors, eye-tracking metrics, and response latency sequences—to detect early signs of dyslexia, dyscalculia, and attention deficit disorders in primary school children. The proposed CNN-LSTM hybrid model was trained on a carefully curated dataset of 2,400 primary school students aged 6–9 across three countries (Greece, Germany, Turkey) and achieved a classification accuracy of 94.3% for dyslexia detection (sensitivity 92.1%, specificity 95.8%) and 91.7% for dyscalculia identification (sensitivity 89.4%, specificity 93.2%). Interpretable attention maps reveal that the most discriminative features include irregular saccadic eye movement patterns during reading and letter reversal sequences in handwriting. The system provides classroom teachers with clear, actionable risk assessments and personalized intervention recommendations. A pilot deployment in 12 schools across two countries demonstrated that AI-assisted early detection led to 40% faster specialist referral compared to traditional screening protocols, with no statistically significant increase in false positive rates.

Keywords: deep learning, learning disabilities, early detection, dyslexia, dyscalculia, educational AI, CNN-LSTM, eye-tracking, multimodal learning

1. Introduction

Learning disabilities affect approximately 5–15% of school-age children worldwide, with dyslexia and dyscalculia representing the most prevalent conditions, estimated to affect 5–10% and 3–6% of the population respectively. Early identification is crucial because the critical developmental window for effective intervention—typically between ages 5 and 8—offers the greatest neuroplasticity and consequently the highest potential for lasting remediation. Children who do not receive timely identification and support are at elevated risk for cascading negative outcomes including chronic academic underachievement, reduced self-efficacy, increased school dropout rates, and long-term socioeconomic disadvantage.

Traditional diagnostic approaches rely heavily on individual clinical assessments conducted by specialist psychologists and educational diagnosticians. These assessments—typically including standardized cognitive batteries such as the WISC-V, reading assessments such as the GORT-5, and behavioral rating scales—are often expensive, time-consuming, and subject to significant delays due to limited specialist availability. In many countries, waiting times between initial teacher referral and formal diagnosis commonly exceed six to twelve months, representing a critical loss of the optimal intervention window.

Recent advances in deep learning, computer vision, and multimodal sensor fusion have opened promising new possibilities for automated screening systems that can analyze behavioral patterns indicative of learning disabilities as they naturally manifest during everyday educational activities. Unlike clinical assessments, which require removing children from the classroom for structured testing, AI-based screening can operate transparently within normal digital learning activities, reducing stigma and sampling bias while enabling continuous monitoring over time.

1.1 Related Work

Prior computational approaches to learning disability detection have primarily relied on single-modality data. Handwriting analysis using image processing techniques has demonstrated diagnostic accuracy of 75–85% for dyslexia in controlled laboratory settings. Eye-tracking studies have identified characteristic saccadic patterns in dyslexic readers, including shorter forward saccades, more frequent regressions, and longer fixation durations. These single-modality approaches, while informative, fail to capture the multimodal and temporally complex nature of learning disabilities, which manifest as characteristic patterns across multiple behavioral channels simultaneously.

The application of deep learning to multimodal educational data represents a relatively nascent but rapidly growing research direction. Several recent studies have demonstrated the feasibility of using recurrent neural networks to model temporal sequences in reading behavior, and convolutional architectures to extract spatial features from handwriting and drawing tasks. However, no prior study has combined these approaches into a unified multimodal framework evaluated in real classroom settings across multiple countries with a sample size sufficient for robust statistical conclusions.

2. Methodology

The study employed a prospective cross-sectional design with longitudinal follow-up validation. Participants were recruited from 18 primary schools across Greece (n = 900), Germany (n = 820), and Turkey (n = 680), totaling 2,400 students aged 6–9 (mean age 7.4 years, SD = 1.1). Ethical approval was obtained from the institutional review boards of all three participating universities, and written informed consent was obtained from parents or legal guardians of all participating students.

2.1 Data Collection Protocol

Data were collected over a 12-month period using a custom digital tablet application that presented age-appropriate reading and mathematical tasks within a gamified interface designed to maintain engagement. Each session lasted approximately 20 minutes and was administered by trained research assistants during regular school hours. Four data modalities were simultaneously captured: (1) handwriting samples via digitizing tablet with 2,048 levels of pressure sensitivity; (2) eye-tracking data via embedded infrared eye-tracking cameras (sampling rate: 120 Hz); (3) touch and stylus interaction logs including timing, pressure, and trajectory; and (4) audio recordings of oral reading passages for phonological analysis.

Ground truth labels were established through comprehensive clinical assessment conducted by licensed educational psychologists blind to the AI model predictions. Assessments included standardized cognitive batteries appropriate to each national context, administered at baseline and at 6-month follow-up. Students were classified as: typically developing (n = 1,848, 77%), dyslexic (n = 312, 13%), dyscalculic (n = 168, 7%), and attention deficit disorder (n = 72, 3%).

2.2 CNN-LSTM Hybrid Architecture

The proposed model employs a dual-stream architecture to jointly process spatial and temporal information. The spatial stream consists of a modified ResNet-34 convolutional backbone that extracts high-level feature representations from handwriting images and eye-tracking heatmaps, with architecture modifications including group normalization layers (replacing batch normalization for improved performance with small batch sizes) and spatial attention mechanisms that allow the model to focus on diagnostically relevant regions.

The temporal stream employs a bidirectional LSTM network with 256 hidden units per direction, processing sequential data including saccade sequences, fixation duration series, pen pressure time series, and reading error sequences. The bidirectional design allows the model to capture both forward and backward temporal dependencies in reading and writing behaviors. Features from both streams are fused via a cross-modal attention mechanism that learns to weight the contribution of each modality based on the current input context, followed by fully connected classification layers with dropout regularization (rate = 0.4).

3. Results

Model performance was evaluated using 5-fold stratified cross-validation, maintaining proportional representation of all diagnostic categories across folds. Performance metrics including accuracy, sensitivity (recall), specificity, precision, and area under the ROC curve (AUC-ROC) were computed for each diagnostic category.

Condition	Accuracy	Sensitivity	Specificity	AUC-ROC
Dyslexia	94.3%	92.1%	95.8%	0.971
Dyscalculia	91.7%	89.4%	93.2%	0.948
ADHD indicators	88.2%	85.6%	89.7%	0.921

Table 1. Classification performance metrics for each diagnostic category (5-fold cross-validation).

3.1 Attention Map Analysis

Gradient-weighted class activation maps (Grad-CAM) applied to the spatial stream revealed interpretable patterns in the model's decision process. For dyslexia detection, the model consistently assigned highest attention weights to letter reversal instances (b/d, p/q confusions), irregular letter spacing, and unusual pen lifts within individual letter formations. For dyscalculia, attention was concentrated on numeral formation irregularities and spatial arrangement errors in multi-digit number writing. These patterns align closely with established clinical indicators, providing important validation of the model's mechanistic interpretability.

3.2 Pilot Deployment Results

The pilot deployment in 12 schools (6 in Greece, 6 in Germany) over a 6-month period demonstrated significant real-world impact. The AI-assisted screening pathway resulted in a 40% reduction in time from initial flag to specialist referral (mean: 8.3 weeks vs. 13.9 weeks for traditional pathways, $p < 0.001$). Importantly, the positive predictive value in the real-world deployment setting was 87.4%, indicating that the vast majority of AI-flagged students who underwent formal assessment received a confirmed diagnosis, avoiding unnecessary burden on specialist services.

4. Discussion

The results demonstrate the strong feasibility and practical utility of multimodal deep learning for early detection of learning disabilities in primary school settings. The classification accuracies achieved by the CNN-LSTM framework—particularly for dyslexia (94.3%) and dyscalculia (91.7%)—substantially exceed those reported for single-modality approaches in the literature and approach the reliability of expert clinical assessment in controlled conditions.

The cross-modal attention mechanism proved particularly valuable, allowing the model to dynamically weight the contribution of different data sources depending on the specific diagnostic question. This flexibility is critical in real-world deployment where data quality varies across students and conditions. The model's ability to maintain high performance even when one data modality is of lower quality—due to technical issues or student non-compliance—demonstrates robustness essential for practical classroom deployment.

4.1 Ethical Considerations

The deployment of AI systems for screening and diagnosis in child populations raises profound ethical questions that must be addressed rigorously before widespread adoption. Privacy considerations are paramount: the behavioral data collected during screening sessions is highly sensitive and must be protected through strong encryption, strict access controls, and clear data retention and deletion policies. We recommend that screening data should not be retained beyond the period necessary for clinical decision-making, and that students should have the right to have their data deleted upon leaving the school system.

The risk of algorithmic bias is also a critical concern. Our cross-country evaluation revealed modest but statistically significant performance differences between national samples ($F(2,84) = 4.73$, $p = 0.011$), likely reflecting genuine differences in educational practices, writing instruction methods, and assessment norms rather than model failures per se. Deploying the system in new national or cultural contexts should therefore involve local validation and, if necessary, fine-tuning on locally representative data.

5. Conclusion

This study presents a validated multimodal deep learning framework for early detection of learning disabilities in primary school children. The CNN-LSTM hybrid architecture achieves clinically meaningful performance across three diagnostic categories, with interpretable outputs that support rather than replace professional clinical judgment. The 40% reduction in specialist referral time observed in pilot deployment underscores the real-world potential of the system to accelerate access to support services during the critical early intervention window.

Future work should focus on longitudinal validation to confirm that early AI-assisted identification translates to improved long-term academic outcomes, development of personalized intervention recommendation engines that integrate with AI detection outputs, and cross-cultural validation studies to ensure equitable performance across diverse student populations globally.

References

[1] Anderson, J. R., & Lebiere, C. (2023). AI in Education: Foundations and Frontiers. Cambridge University Press.

[2] Baker, R. S. (2024). Challenges for AI in education: Review and prospects. Learning and Instruction, 89, 101–115.

[3] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8, 75264–75278.

[4] Dillenbourg, P. (2023). Adaptive Educational Technologies: Trends and Perspectives. Springer International.

[5] European Commission. (2024). AI in Education and Training: Policy Recommendations. Brussels: EC Publications.

[6] Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial Intelligence in Education (2nd ed.). UNESCO/OECD.

[7] Luckin, R. (2024). Machine Learning and Human Intelligence. UCL IOE Press.

[8] Rayner, K., et al. (2013). Eye movements and dyslexia. Developmental Psychobiology, 55(5), 591–610.

[9] Shaywitz, S. E., & Shaywitz, B. A. (2008). Paying attention to reading: The neurobiology of reading and dyslexia. Development and Psychopathology, 20(4), 1329–1349.

[10] Zawacki-Richter, O., et al. (2019). Systematic review of research on AI in higher education. IRRODL, 20(1), 1–27.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. CVPR 2016, 770–778.

[12] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.